EVELYN SOUZA CHAGAS OLIVEIRA

Modelos lineares e aditivos generalizados para posição, escala e forma na análise de estoques pesqueiros de lagostas (*Panulirus spp*) e de espadarte (*Xiphias gladius*)

RECIFE-PE

FEVEREIRO/2018

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO

PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO

PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

Modelos lineares e aditivos generalizados para posição, escala e forma na análise de estoques pesqueiros de lagostas (*Panulirus spp*) e de espadarte (*Xiphias gladius*)

Tese julgada adequada para obtenção do título de Doutor em Biometria e Estatística Aplicada, defendida e aprovada por unanimidade em 19/02/2018 pela comissão examinadora.

Área de Concentração: Biometria e Estatística Aplicada.

Orientador: Profº. Dr. Humber Agrelli Andrade

RECIFE-PE

FEVEREIRO/2018

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema Integrado de Bibliotecas da UFRPE
Biblioteca Central, Recife-PE, Brasil

O48m Oliveira, Evelyn Souza Chagas.
Modelos lineares e aditivos generalizados para posição, escala e forma na análise de estoques pesqueiros de lagostas (Panulirus spp) e de espadarte (Xiphias gladius) / Evelyn Souza Chagas Oliveira. – Recife, 2018.
141 f.: il.

Orientador(a): Humber Agrelli Andrade.
Tese (Doutorado em Biometria e Estatística Aplicada) – Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, Recife, 2018.
Inclui referências e apêndices.

1. Defeso 2. Taxa de captura 3. GAMLSS 4. GLMM I. Andrade, Humber Agrelli, orient. II. Título

CDD 310

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO

PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

Modelos lineares e aditivos generalizados para posição, escala e forma na análise de estoques pesqueiros de lagostas (*Panulirus spp*) e de espadarte (*Xiphias gladius*)

EVELYN SOUZA CHAGAS OLIVEIRA

Tese julgada adequada para obtenção do título de Doutor em Biometria e Estatística Aplicada, defendida e aprovada por unanimidade em 19/02/2018 pela comissão examinadora.

Orientador:

_____
Profº. Dr. Humber Agrelli Andrade
DEPAq - UFRPE

Banca Examinadora:

_____
Profº. Dr. Paulo José Duarte Neto
DEINFO - UFRPE

_____
Profª. Dra. Tatijana Stosic
DEINFO - UFRPE

_____
Profº. Dr. José Rodrigo Santos Silva
DECAT - UFS

_____
Profº. Dr. Lázaro de Souto Araújo
DCFS - UFPB

Dedico este trabalho aos meus pais Edna e Adelson, ao meu esposo Orlando, e aos meus irmãos Richard e Wendell, vocês foram o alicerce para esta conquista.

“Se cheguei até aqui foi porque me apoiei no ombro dos gigantes” (Isaac Newton)

AGRADECIMENTOS

Ao meu Deus, razão da minha existência e das bênçãos em minha vida, meu louvor, minha adoração e minha gratidão. Sua constante presença ao meu lado como fonte de conhecimento, saber e amor me conduziu, me deu sabedoria, me fortaleceu, dando-me vitórias sobre vitórias em toda minha trajetória acadêmica, percebo a cada dia que Tu És tão imprescindível quanto o ar que respiro. TE ADORO... TE AMO... MEU ETERNO DEUS.

Aos meus pais, Adelson e Edna, que no decorrer da minha vida, ensinaram-me que para alcançar o sucesso é necessário ter dignidade, perseverança e superação. A estes que nos momentos mais tenebrosos e obscuros dessa caminhada me fizeram entender que por mais árdua que seja a batalha, segui-la com Deus e tendo como escudo a fé, alcançaremos os sonhos almejados, Deus os abençoe!

Ao meu esposo Orlando, pelo incentivo dispensado em forma de compreensão, estímulo e amor, muitas vezes tendo que se ausentar para que eu pudesse me dedicar aos estudos, me animando nos momentos difíceis, me fazendo acreditar que posso mais do que imagino. Obrigada por acreditar em mim e ter feito do meu sonho, o nosso sonho!

Aos manos Richard e Wendell, meus colaboradores, obrigada pelo apoio e vibração por cada etapa conquistada, foram ingredientes motivadores para que eu chegasse até o final dessa jornada, meu muito obrigada!

Aos meus cunhados Ogenivaldo, Priscila, Talita e Karla por todo apoio, carinho e incentivo e, também, não poderia esquecer da minha sobrinha July, que, com seu jeitinho doce sempre alegrou os meus dias de tensão, me dando forças para concluir mais uma etapa, obrigada querida!

Aos meus queridos avós, tios e primos, sogros, obrigada pelo apoio transmitido, seja através de um simples olhar expressivo, um sorriso, uma admiração, um abraço e orações. Vocês fizeram valer essa conquista!

Ao meu orientador Prof. Dr. Humber Agrelli, que não mediu esforços no compartilhar conhecimentos e pela dedicação, muitas vezes organizando sua agenda para poder me atender, acreditando e me motivando a alcançar a conclusão dessa etapa: minha gratidão constante e eterna admiração.

A todos os professores do Programa de Pós-Graduação em Biometria e Estatística Aplicada pelos conhecimentos transmitidos, atenção e amizade, como também, aos funcionários desse programa, em particular o secretário Marco, a dedicação e atenção nos momentos solicitados, meus sinceros agradecimentos!

A Universidade Federal Rural de Pernambuco pela oportunidade de adquirir e ampliar meus conhecimentos, sempre me orgulhará.

As minhas grandes amigas Lêda Valéria e Hérica que continuaram comigo desde a faculdade, sempre presentes em cada momento, afirmo: o companheirismo, o carinho e amizade foram importantes para vencer essa jornada. Chegou o momento de seguir viagem sozinha, mas vocês sempre estarão presentes em minha memória e nessa grande conquista.

Aos amigos Samuel, Rodrigo e Alessandro, sempre prestativos, nosso suporte ao chegarmos a Recife, grandes professores e incentivadores nos apoiando nos bons e maus momentos e nos fazendo sonhar alto, minha sincera gratidão. Essa minha conquista também pertence a vocês!

A todos os colegas, da biometria e da pesca pelos momentos vividos juntos e conhecimentos compartilhados e aqueles que de tão próximos, se tornaram uma família: Lêda, Hérica, Kerolly, Nathiely, Alessandro, Rodrigo, Diego, Fábio, Samuel, Rivelino, Djalma, Érika, Renata e Vivian, meu reconhecimento, carinho e gratidão.

Ao meu pastor Luiz Antônio, agradeço pelas orações e sábias palavras de incentivo proferidas nas suas pregações, as quais, fizeram-me enxergar que durante a minha trajetória acadêmica eu enfrentaria obstáculos, a batalha árdua seria, mas munida de fé e perseverança eu iria galgar degraus que me levariam ao alcance dos meus sonhos, dos meus objetivos.

A todo ministério infantil da igreja, da qual faço parte, agradeço o carinho, apoio, força e compreensão em ter que me afastar das minhas atividades, por um período, para a conclusão do doutorado.

Enfim, a todos que contribuíram de forma direta ou indireta para a realização dessa conquista, deste sonho, a minha gratidão permanente e que Deus os conduza sempre com paz, saúde e prosperidade.

RESUMO

Entre os principais recursos pesqueiros explorados no Brasil destacam-se o espadarte do Atlântico Sul e as lagostas espinhosas. No país, uma das regulamentações mais tradicionais para garantir a sustentabilidade da pesca de lagosta é o defeso. Porém as análises que levaram às decisões quanto aos períodos de defeso são fragmentadas, e as análises realizadas anteriormente são de caráter descritivo. Neste trabalho foi construída uma série temporal completa com a recompilação de trabalhos nos quais a gestão da pesca tem se baseado para a tomada de decisões. Para elucidar padrões nas variações das proporções de lagostas em reprodução foram utilizados os modelos lineares generalizados (GLM) e modelos aditivos generalizados para posição, escala e forma (GAMLSS). Os resultados obtidos sugerem uma reavaliação do período de defeso, uma vez que, foram registrados meses que não fazem parte da paralisação das atividades pesqueiras que foram encontradas fêmeas em intensa atividade reprodutiva e, meses que fazem parte do defeso que foram encontradas pequenas proporções de fêmeas maturas. No caso do espadarte, as decisões de gestão são na maioria dos casos baseadas em modelos de avaliações de estoques que demandam estimativas das taxas de Captura por Unidade de Esforço (CPUE) padronizadas. Com frequência a padronização da CPUE é realizada utilizando três abordagens: A) ano incluído no modelo apenas como efeito fixo principal; B) ano incluído no modelo como efeito fixo principal, mas também em interações de efeito fixo e, C) o ano incluído como efeito fixo principal e também nas interações de efeito aleatório. Este trabalho se propõe a calcular as CPUEs padronizadas utilizando as três abordagens e realizar uma comparação dos resultados com base no teste não-paramétrico de Kruskal-Wallis ao nível de 5% de significância. Foram utilizados os modelos lineares generalizados (GLM) e os modelos lineares generalizados mistos (GLMM) para obter estimativas das CPUEs padronizadas do espadarte para cada abordagem. Os resultados obtidos indicaram que a escolha da abordagem pode conduzir a resultados equivocados sobre a situação do estoque quando analisado apenas os anos isoladamente.

Palavras-chave: Defeso, Taxa de captura, GAMLSS e GLMM.

ABSTRACT

Among the principal fishing resources exploited in Brazil are the South Atlantic swordfish and spiny lobsters. In the country, one of the most traditional regulations to ensure the sustainability of lobster fishing is the closed season. However, the analyzes that led to the decisions about the closed seasons are fragmented, and the analyzes fulfilled previously are descriptive character. In this work a complete time series was built with the recompilation of works in which the fishery management has been based for the decisions making. In order to elucidate patterns in the variations of lobsters proportions in reproduction, generalized linear models (GLM) and generalized additive models for location, scale and shape (GAMLSS) were used. The results obtained suggest a reassessment of the closed period, a time that, were recorded months that are not part of the stoppage of fishing activities that females in intense reproductive activity were found, and months that are part of the closed season which were found small proportions of mature females. In the case of swordfish, management decisions are in most cases based on models of stock assessments that require estimates of the rates of Catch per Unit Effort (CPUE) standardized. Often the standardization of the CPUE is performed using three approaches: (a) year included in the model only like fixed effect main; (B) year included in the model like fixed effect, but also in interactions of fixed effect, and c) the year included like fixed effect main and also in the interactions of random effect. This work proposes to calculate the CPUEs standardized using the three approaches and perform a comparison of results based on the nonparametric Kruskal-Wallis test at 5% level of significance. The generalized linear models (GLM) and the generalized linear mixed models (GLMM) were used to obtain estimates of the CPUEs standardized of swordfish for each approach. The results obtained indicated that the choice of approach may lead to erroneous results on the situation of the stock when analyzed only those years alone.

Keywords: closed season, capture rate, GAMLSS and GLMM.

LISTA DE FIGURAS

Figura 1 – Algas calcárias das classes *Rhodophyceae* e *Chlorophyceae*. 8
Figura 2 – Ciclo reprodutivo da lagosta *Panulirus argus*. (a) – óvulos aderidos ao abdômen; (b) massa espermatofórica no esterno e massa ovígera aderida aos pleópodos; (c) restos de massa espermatofórica. 10
Figura 3 – Variações anuais (a) e mensais (b) das proporções de fêmeas capturadas para as espécies *Panulirus argus* e *Panulirus laevicauda*. 41
Figura 4 – Variação anual das proporções de fêmeas capturadas com ovos (PO), com espermoteca (PE), com resto de espermoteca (PRE), com ovos, espermoteca ou resto de espermoteca (PREP) para as espécies *Panulirus argus* (a) e *Panulirus laevicauda* (b). 42
Figura 5 – Relação entre as proporções de lagostas capturadas com ovos (PO), com espermoteca (PE), com resto de espermoteca (PRE), com ovos, espermoteca ou resto de espermoteca (PREP) e os comprimentos dos exemplares de *Panulirus argus* (a) e *Panulirus laevicauda* (b) capturados. 44
Figura 6 – Proporção de lagostas capturadas com ovos (PO), com espermoteca (PE), com resto de espermoteca (PRE), com ovos, espermoteca ou resto de espermoteca (PREP) ao longo dos meses para as espécies *Panulirus argus* (a) e *Panulirus laevicauda* (b). 44
Figura 7 – Diagnóstico do ajuste do modelo para a espécie *Panulirus argus*. 51
Figura 8 – Diagnóstico do ajuste do modelo para a espécie *Panulirus laevicauda*. ..52
Figura 9 – Diagnóstico do modelo ajustado para a espécie *Panulirus argus*. 58
Figura 10 – Diagnóstico do modelo ajustado para a espécie *Panulirus laevicauda*. .58
Figura 11 – Worm-plot dos modelos ajustados com distribuição beta binomial para as espécies *Panulirus argus* (a) e *Panulirus laevicauda* (b). Diferenças entre os z-escores dos valores empíricos padronizados e os z-escores dos valores esperados sob a distribuição teórica contra os z-escores dos valores teóricos.....59
Figura 12 – Predições da esperança calculadas com o modelo ajustado para a espécie *Panulirus argus*, fixando um comprimento total da lagosta (PM) no momento da captura de 12,55cm (a), 15,55cm (b), 18,55cm (c), 21,55cm (d), 24,55cm (e), 27,55cm (f), 30,55cm (g) e 33,55cm (h). 60

Figura 13 – Predição da esperança calculadas com o modelo ajustado para a espécie *Panulirus laevicauda*, fixando um comprimento total da lagosta (PM) no momento da captura de 12,55cm (a), 15,55cm (b), 18,55cm (c), 21,55cm (d), 24,55cm (e) e 27,55cm (f). 62
Figura 14 – Espadarte *Xiphias gladius Linnaeus* 1958. 70
Figura 15 – Estrutura do espinhel de superfície de deriva: (A) samburá (composição entre boias subsequentes) de um espinhel pelágico de superfície; (B) sistema de fixação das linhas secundárias; (C) unidade básica de todo espinhel. 72
Figura 16 – Distribuição das capturas de *Xiphias gladius* por área durante o período de 1990 a 2012. Áreas de pesca: N – Norte; C – Centro e S – Sul.....82
Figura 17 – Captura por unidade de esforço (CPUE) por ano, trimestre e área (N - norte; C - centro; S - sul). 83
Figura 18 – Diagnóstico dos resíduos (abordagem A). 89
Figura 19 – Diagnóstico dos resíduos (abordagem B). 90
Figura 20 – Diagnóstico dos resíduos (abordagem C). 91
Figura 22 – Taxa de captura por unidade de esforço (CPUE) padronizada calculada com as abordagens A, B e C. As estimativas foram escalonadas calculando a diferença entre o valor original e a média da série, dividido pelo desvio padrão da série. 99
Figura 22 – Diferença entre as taxas de captura por unidade de esforço (CPUE) padronizada calculada com as abordagens A e B (a), com as abordagens A e C (b) e com as abordagens B e C(c).....100

LISTA DE TABELAS

Tabela 1 – Distribuições discretas implementadas no pacote gamlss (com função de ligação padrão).....29
Tabela 2 – Quantidade de dias de pesca, período de realização da pesca e número de indivíduos amostrados das espécies *Panulirus argus* e *Panulirus laevicauda* durante os anos de 1962 a 1969.....34
Tabela 3 – Estatísticas descritivas para as espécies *Panulirus argus* e *Panulirus laevicauda* capturadas entre 1962 e 1969. PO – proporção de lagostas capturadas com ovos; PE – proporção de lagostas capturadas com espermoteca; PRE – proporção de lagostas capturadas com resto de espermoteca; PREP – proporção de lagostas capturadas com ovos ou espermoteca ou resto de espermoteca; PM – comprimento das lagostas capturadas; PF – proporção de fêmeas capturadas em relação à soma de fêmeas e machos; CV – coeficiente de variação em porcentagem; p-valor – nível descritivo do teste de Kolmogorov-Smirnov para verificar normalidade das variáveis. 40
Tabela 4 – Estrutura dos cinco melhores modelos ajustados com distribuição Binomial, Poisson e Binomial Negativa. REP – variável resposta, representando o número (Poisson e Binomial Negativa) ou proporção (Binomial) de lagostas capturadas com ovos ou espermoteca ou resto de espermoteca; ANO – ano em que as lagostas foram capturadas; MES – mês que as lagostas foram capturadas; PM – comprimento das lagostas capturadas. Interações particulares são representadas por dois pontos. A letra “F” na frente da variável indica que a variável foi tratada como fator no modelo, caso contrário a variável foi tratada como covariável. Interações particulares são representadas por dois pontos. 46
Tabela 5 – Comparativo dos modelos ajustados para análise das lagostas em fase reprodutiva. Modelos de 1 a 5 foram ajustados com distribuição binomial; modelos de 6 a 10 foram ajustados com distribuição Poisson; modelos de 11 a 15 foram ajustados com distribuição binomial negativa. Deviância – deviância do modelo nulo (composto somente pelo intercepto); GL – graus de liberdade do modelo nulo; Dev. Residual – deviância residual; GL Residual – graus de liberdade residual; LogLik – estatística de log-verossimilhança; BIC – Critério de Informação Bayesiano; AIC – Critério de

Informação de Akaike; Delta AIC – diferença entre o AIC do modelo que está sendo observado e o modelo com menor AIC. 47
Tabela 6 – Análise da deviância para os modelos binomiais selecionados para a espécie *Panulirus argus* e *Panulirus laevicauda*. GL – graus de liberdade; Dev – deviância; ANO – ano em que as lagostas foram capturadas; MES – mês que as lagostas foram capturadas; PM – comprimento das lagostas capturadas. Interações particulares são representadas por dois pontos. 49
Tabela 7 – Comparativo dos modelos ajustados para análise das lagostas em fase reprodutiva. GL – representa graus de liberdade; AIC – Critério de Informação de Akaike; BIC – Critério de Informação Bayesiano.....53
Tabela 8 – Modelos iniciais ajustados para análise das lagostas em fase reprodutiva. GL – representa graus de liberdade; AIC – o Critério de Informação de Akaike; BIC – o Critério de Informação Bayesiano. 54
Tabela 9 – Seleção de termos aditivos para os parâmetros da distribuição beta binomial utilizando a estratégia A. AIC – Critério de Informação de Akaike; REP – proporção de lagostas vermelhas capturadas com ovos ou espermoteca ou resto de espermoteca; ANO – ano em que as lagostas vermelhas foram capturadas; MES – mês que as lagostas vermelhas foram capturadas; PM - comprimento das lagostas vermelhas capturadas. Interações particulares são representadas por dois pontos. 55
Tabela 10 – Seleção de termos aditivos para os parâmetros da distribuição beta binomial utilizando a estratégia A. AIC – Critério de Informação de Akaike; REP – proporção de lagostas verdes capturadas com ovos ou espermoteca ou resto de espermoteca; ANO – ano em que as lagostas verdes foram capturadas; MES – mês que as lagostas verdes foram capturadas; PM - comprimento das lagostas verdes capturadas. Interações particulares são representadas por dois pontos. 56
Tabela 11 – Número de registros de lances de pesca por bandeira retidos para análise do espadarte capturado no Atlântico Sul no período de 1990 a 2012.....79
Tabela 12 – Estrutura dos cinco melhores modelos ajustados com distribuição Poisson e Binomial Negativa. S – número de espadartes capturados; F - bandeira do barco; Y – ano; A - área de pesca; T - trimestre; H – número de samburás (segmento entre uma boia a outra) no espinhel; N – número de anzóis por samburá. Interações particulares são representadas por dois pontos. Interações de primeira ordem entre cada combinação pareada de duas variáveis são representadas por parênteses e elevado a segunda potência ( )^2. Interações particulares são representadas por dois pontos. O

sinal “menos (-)” na frente das interações indica que as mesmas foram eliminadas do modelo. 84
Tabela 13 – Comparativo dos modelos ajustados para padronização da captura por unidade de esforço do *Xiphias gladius* capturado pela frota brasileira de 1990 a 2012. Modelos de 1 a 5 foram ajustados com distribuição Poisson; modelos de 6 a 10 foram ajustados com distribuição Binomial negativa. GL Residual – graus de liberdade residual; LogLik – estatística de log-verossimilhança; BIC – critério de informação Bayesiano; AIC – critério de informação de Akaike; Delta AIC – diferença entre o AIC do modelo que está sendo observado e o modelo com menor AIC. 86
Tabela 14 – Teste da razão de verossimilhança (LR) para o modelo 6 com distribuição binomial negativa e abordagem B. Os fatores são: F – bandeira; Y – ano; A – área; T – trimestre; H – número de samburás (segmento entre uma boia a outra) no espinhel; N – número de anzóis por samburá. GL – graus de liberdade. Interações particulares são representadas por dois pontos. 87
Tabela 15 – Análise da deviância dos modelos ajustados. Testes da razão de verossimilhança (LR) foram utilizados para as abordagens A e B, e teste Wald foi utilizado para a abordagem C. Os fatores são: F – bandeira; Y – ano; A – área; T – trimestre; H – número de samburás (segmento entre uma boia a outra) no espinhel; N – número de anzóis por samburá. GL – graus de liberdade. Interações particulares são representadas por dois pontos. 88
Tabela 16 – Estimativa dos parâmetros (abordagem A). F – bandeira; Y – ano; A – área; T – trimestre; H – número de samburás (segmento entre uma boia a outra) no espinhel; N – número de anzóis por samburá; Significâncias: (***) - 0,001, (**) - 0,01 e (*) - 0,05 92
Tabela 17 – Estimativa dos parâmetros (abordagem B). F – bandeira; Y – ano; A – área; T – trimestre; H – número de samburás (segmento entre uma boia a outra) no espinhel; N – número de anzóis por samburá. Interações particulares são representadas por dois pontos. Significâncias: (***) - 0,001, (**) - 0,01, (*) - 0,05 e (.) - 0,1. 93
Tabela 18 – Estimativa dos parâmetros (abordagem C). F – bandeira; Y – ano; A – área; T – trimestre; H – número de samburás (segmento entre uma boia a outra) no espinhel; N – número de anzóis por samburá. Interações particulares são representadas por dois pontos. Significâncias: (***) - 0,001, (**) - 0,01, (*) - 0,05 e (.) - 0,1. 97

Tabela 19 – Estimativas dos componentes de variâncias para os efeitos aleatórios (Abordagem C). F – bandeira; Y – ano; A – área; T – trimestre; H – número de samburás (segmento entre uma boia a outra) no espinhel; N – número de anzóis por samburá. Interações particulares são representadas por dois pontos. 98
Tabela 20 – Teste de Kruskal-Wallis para comparação anual entre as abordagens A, B e C. GL – graus de liberdade. 101

SUMÁRIO

1 Introdução 1

2 Análise do ciclo reprodutivo das espécies *Panulirus argus* e *Panulirus laevicauda* com modelos lineares generalizados e modelos aditivos generalizados para posição, escala e forma 5

2.1 Introdução 5

2.2 Revisão de literatura 8

2.2.1 Habitat e área de captura das lagostas 8

2.2.2 Ciclo reprodutivo e maturidade sexual das lagostas 9

2.2.3 Ciclo migratório das lagostas 11

2.2.4 Dieta alimentar, crescimento e longevidade das lagostas 12

2.2.5 Modelos de regressão estatísticos 12

2.2.6 Modelos lineares generalizados 14

2.2.6.1 Estimação dos parâmetros 16

2.2.6.2 Teste de hipótese 20

2.2.6.2.1 Teste da razão de verossimilhanças 21

2.2.6.2.2 Teste de Wald 21

2.2.6.3 Análise de deviância e seleção de modelos 21

2.2.6.4 Análise dos resíduos 24

2.2.6.4.1 Teste de Breusch-Pagan 25

2.2.7 Modelos aditivos generalizados para posição, escala e forma (GAMLSS) 25

2.2.7.1 Definição 25

2.2.7.2 Estimação dos parâmetros do modelo 28

2.2.7.3 Distribuições 28

2.2.7.4 Seleção de modelos e diagnósticos 30

2.2.8 Testes estatísticos não-paramétricos 31

2.2.8.1 Teste de Kolmogorov-Smirnov 31

2.2.8.2 Teste de Kruskal-Wallis 32

2.3 Dados e Metodologia 34
2.3.1 Dados 34
2.3.2 Modelos de regressão e variáveis utilizadas 35
2.3.2.1 Variável resposta 35
2.3.2.2 Variáveis explicativas 36
2.3.2.3 Seleção dos modelos lineares generalizados e diagnósticos. 36
2.3.2.4 Seleção dos modelos aditivos generalizados para posição, escala e forma, e diagnósticos 37
2.4 Resultados e Discussões 39
2.4.1 Análise Exploratória 39
2.4.2 Modelos lineares generalizados (GLM) 45
2.4.3 Modelos aditivos generalizados para posição, escala e forma (GAMLSS) 52
2.5 Conclusão 66
3 Padronização da CPUE do *Xiphias gladius* usando efeito fixo e aleatório 67
3.1 Introdução 67
3.2 Revisão de Literatura 69
3.2.1 Biologia, características externas e população do espadarte 69
3.2.2 Ciclo migratório e alimentação do espadarte 71
3.2.3 Pesca com espinhel pelágico 72
3.2.4 Modelos lineares generalizados mistos 74
3.2.4.1 Definição do modelo 74
3.2.4.2 Estimação por máxima verossimilhança 75
3.2.4.2.1 Equações de verossimilhança para parâmetros de efeitos fixos 76
3.2.4.2.2 Equações de verossimilhança para parâmetros de efeitos aleatórios 77
3.2.4.3 Testes de hipóteses 78
3.3 Dados e Metodologia 79
3.3.1 Dados de esforço de pesca e captura de espadarte no Atlântico Sul 79

3.3.2 Modelos de regressão e variáveis utilizadas 80
3.3.3 Seleção de modelos e diagnósticos 80

3.4 Resultados e Discussões 81
3.4.1 Análise Exploratória.....81
3.4.2 Modelos lineares generalizados e modelos lineares generalizados mistos 83

3.5 Conclusão 103

4 Considerações Finais 105

Referências 107

APÊNDICE A.....127

1 Introdução

A pesca é uma importante fonte de alimentação, nutrição, renda e meios de subsistência para milhares de pessoas em todo o mundo. A produção mundial de pescados atingiu 167,2 milhões de toneladas em 2014, das quais 93,4 milhões de toneladas foram provenientes da pesca extrativista, e 73,8 milhões de toneladas da aquicultura. Essa produção deve aumentar em 17%, chegando a 196 milhões de toneladas até 2025, ainda que com crescimento anual mais lento que na década anterior (FAO, 2016). O comércio mundial de peixes e produtos pesqueiros também cresceu significativamente, com as exportações aumentando de US$ 8 bilhões em 1976 para US$ 148 bilhões em 2014 (FAO, 2016).

A captura de pescados do Brasil foi de 767.026 toneladas no ano de 2014, representando 0,82% da produção mundial. Em adição, o Brasil tem tido destaque na aquicultura, com produção de 561.803 toneladas em 2014 e estimativa de crescimento para 2025 de 104,4%. No tocante ao comércio internacional de produtos pesqueiros, o país gastou US$ 1.558.567 com importações em função do déficit de alimentos, enquanto o valor em exportações foi de US$ 207.567. Dentre os recursos pesqueiros mais valiosos comercialmente para o país estão os atuns e afins, e as lagostas (FAO, 2016a).

Em virtude da importância da pesca no mundo, um dos grandes desafios da gestão pesqueira é assegurar o desenvolvimento sustentável em termos econômicos, sociais e ambientais. Como estratégia para cumprir essas metas algumas medidas são adotadas. No caso das lagostas, uma das regulamentações mais tradicionais é o defeso que em princípio tem a finalidade de proteger o estoque no período reprodutivo (IBAMA, 2008). No entanto, a grande maioria dos trabalhos que deram e dão suporte à escolha de períodos de defeso foram baseados na análise de registros feitos em determinados anos isoladamente, não levando em consideração o acúmulo de dados anteriores a cada um dos anos avaliados. Uma análise integrada permite que seja verificada o quão é adequada a decisão de se manter um período de defeso fixo em determinada janela temporal, o que não seria razoável se houvesse variações interanuais importantes quantos aos meses em que ocorrem os picos de desova.

Uma alternativa para a análise da série temporal das informações sobre a atividade reprodutiva é o uso de modelos lineares generalizados (GLM) (NELDER e WEDDERBURN, 1972), que apesar de não ser frequente em estudos com lagostas, tem sido usado para análise de questões relativas a outras espécies aquáticas (*e.g.* SHINOZAKI-MENDES *et al.*, 2014; LIMA, 2016; GALLARDO, 2017). Tais modelos permitiram integrar diferentes variáveis explicativas (*e.g.* mês, ano e sexo do exemplar) em uma formulação única, além de ampliar o leque de possibilidades para a distribuição de probabilidade da variável resposta, uma vez que a mesma pode pertencer à família exponencial de distribuições (*e.g.* distribuições normal, gama e normal inversa para dados contínuos; binomial para proporções; Poisson e binomial negativa para contagens). No entanto, mesmo os GLMs permitindo o uso de distribuições assimétricas, muitas das vezes, os dados apresentam ainda uma variabilidade distinta para diferentes unidades amostrais, de forma não proporcional a uma função da média, levando à necessidade de modelagem do parâmetro de escala do modelo. Este tipo de abordagem pode ser realizado por meio dos modelos aditivos generalizados para posição, escala e forma (GAMLSS) (RIGBY e STASINOPOULOS, 2005).

A classe de modelos aditivos generalizados para posição, escala e forma foi apresentada inicialmente por Rigby e Stasinopoulos (2005) como uma abordagem para superar algumas limitações dos modelos lineares generalizados (GLM) e dos modelos aditivos generalizados (GAM). Trata-se de uma técnica de modelagem estatística que permite o ajuste de uma ampla família de distribuições contínuas e discretas para a variável resposta e possibilita a modelagem explícita, com funções paramétricas e/ou não-paramétricas, de quatro parâmetros da distribuição da variável resposta (*e.g.* média, variância, assimetria e curtose) em relação às variáveis explanatórias. Um estudo do ciclo reprodutivo utilizando o GAMLSS foi desenvolvido por Postuma e Gasalla (2014) com o objetivo de identificar os padrões populacionais explorados de *Doryteuthis plei* em torno da ilha de São Sebastião. Para isto, os autores exploraram o processo de desenvolvimento de machos e órgãos reprodutivos femininos, variáveis biométricas que traduzem a atividade reprodutiva, o tamanho na maturidade e os fatores espaciais e temporais que influenciam a reprodução das espécies localmente.

Para os atuns e afins, as recomendações de gestão para as pescarias são feitas através de estimativas da dimensão e da capacidade produtiva dos estoques

pesqueiros (“avaliações de estoques”). Para tanto, frequentemente é necessário dispor de dados da taxa de captura por unidade de esforço (CPUE), os quais são utilizados para calcular índices relativos de abundância dos estoques. As análises de CPUE destinadas à estimativa de séries temporais de índices relativos de abundância são normalmente chamadas de "padronização" (MAUNDER e PUNT, 2004). Com frequência a padronização da CPUE é baseada em estimativas dos parâmetros para o fator "ano", utilizado como variável explicativa em GLMs ou Modelos Lineares Generalizados Mistos (GLMM) (BRESLOW e CLAYTON, 1993; MCCULLOCH e SEARLE, 2001). De maneira geral podem ser utilizadas três abordagens: A) ano incluído no modelo apenas como efeito fixo principal (*e.g.* BERNASCONI *et al.*, 2015); B) ano incluído no modelo como efeito fixo principal, mas também em interações de efeito fixo (*e.g.* MOURATO *et al.*,2016) e, C) o ano incluído como efeito fixo principal e também nas interações de efeito aleatório (WANG *et al.*, 2014). A vantagem dos GLMMs é que estes modelos admitem a inclusão de efeitos aleatórios no preditor linear e permitem modelar não somente a média da variável resposta, mas também a variância, com a inclusão de termos adicionais na relação entre a resposta média e as variáveis explicativas, que sejam quantidades aleatórias ao invés de parâmetros de efeito fixo somente. No entanto, há pouco entendimento das consequências da escolha de abordagem nas padronizações de CPUEs e nas avaliações de estoque.

Diante dessas duas problemáticas, neste trabalho se propõe: 1) O uso do GAMLSS, além do GLM, para analisar os dados disponíveis sobre lagostas a fim de elucidar os padrões nas variações temporais das proporções de fêmeas em reprodução, o que deve contribuir para avaliação da adequacidade dos períodos de defeso; e 2) A comparação entre as três abordagens para padronizar as CPUEs com aplicação para o caso espadarte (*Xiphias gladius*) capturado na pesca de atuns e afins no Atlântico Sul.

A expectativa é de que as informações geradas possam colaborar com o gerenciamento dos recursos pesqueiros, quanto a diminuição de incertezas na gestão da lagosta, bem como sobre a escolha de abordagem nas estimativas dos índices de abundância para a avaliação do estoque de espadarte no Atlântico Sul.

Esta tese está organizada em quatro capítulos. No capítulo 1 é abordada a importância da pesca para o mundo e as duas problemáticas estudadas. O capítulo 2 que trata somente da primeira problemática, contêm uma introdução mais detalhada sobre o estabelecimento dos períodos de defeso para lagostas e os métodos que

foram empregados para o estudo do ciclo reprodutivo das fêmeas. É ainda apresentada uma revisão de literatura dos aspectos biológicos das espécies, dos modelos lineares generalizados, e dos modelos aditivos generalizados para posição, escala e forma. Se explica também como os dados foram coletados, e os procedimentos utilizados para seleção e avaliação dos modelos ajustados. Por fim são apresentados os resultados, a discussão, e as conclusões. No capítulo 3 é abordada a segunda problemática, a qual refere-se a estimativa das taxas de captura por unidade de esforço (CPUE) padronizadas como índices relativos de abundância do espadarte do Atlântico Sul. Este capítulo foi estruturado de forma semelhante ao capítulo 2, contendo introdução, revisão de literatura, descrição dos dados, metodologia, resultados, discussões, e conclusões. No capítulo 4, são apresentadas as conclusões gerais sobre os resultados obtidos nos capítulos 2 e 3, e sugestões para a resolução dos problemas abordados.

2 Análise do ciclo reprodutivo das espécies *Panulirus argus* e *Panulirus laevicauda* com modelos lineares generalizados e modelos aditivos generalizados para posição, escala e forma

2.1 Introdução

A exploração lagosteira no Brasil abrange do Amapá ao Espírito Santo, sendo que a maioria das capturas são das lagostas vermelha (*Panulirus argus*) (70%) e verde (*Panulirus laevicauda*) (29%), enquanto que outras espécies têm participação pouco expressiva (FONTELES-FILHO, 2000). Na região litorânea a pesca de lagostas destaca-se pela grande importância social e econômica. Milhares de empregos são gerados, e estima-se que aproximadamente 100 mil pessoas dependam dessa pesca, direta ou indiretamente (CEPENE, 2013). A atividade é voltada, basicamente, para o mercado externo, gerando um ingresso anual de divisas de 50 a 90 milhões de dólares (ANON, 2013).

A pescaria de lagostas em escala comercial teve início em meados da década de 1950 no Ceará, que é o principal produtor (ANON, 2013). Com o aumento do esforço de pesca nas décadas subsequentes e expansão da atividade, ocorreram quedas nas taxas de captura. Desde a década de 1960, diversas medidas vêm sendo adotadas com objetivo de se evitar um cenário de sobrepesca (IBAMA, 2008). Tradicionalmente tem se optado por medidas pautadas no ciclo reprodutivo como limites mínimos de tamanhos para a captura, e o defeso (1° de dezembro a 31 de maio) no período de reprodução (IVO e PEREIRA, 1996). A motivação para a adoção de um tamanho mínimo seria o de aumentar a probabilidade de sobrevivência dos recrutas até a fase matura. Já o defeso reprodutivo é destinado à proteção dos exemplares no período de desova, com aumento da taxa de sobrevivência, o que resultaria na geração de maior quantidade de ovos e larvas, que eventualmente poderiam ter sucesso e atingir a idade de recrutamento.

A proteção de indivíduos em reprodução e o limite de tamanho para a captura foram sugestões iniciais de Coelho (1962). Posteriormente, com a criação da Superintendência do Desenvolvimento da Pesca – SUDEPE, instituição federal vinculada ao Ministério da Agricultura, criada pela Lei Delegada Nº10, de 11 de outubro de 1962, responsável pelo ordenamento pesqueiro no Brasil, foram introduzidas as primeiras medidas de proteção ao estoque pesqueiro entre os anos de 1967 e 1969. Medidas estas fundamentadas no estudo de Coelho (1962) e em vários outros (*e.g.* Paiva e Silva, 1962; Paiva e Costa, 1964; Paiva, 1966; Paiva e Costa, 1969) conduzidos na Estação de Biologia Marinha da Universidade Federal do Ceará, hoje, Instituto de Ciências do Mar – LABOMAR. Apesar de datarem de algumas décadas atrás, estes estudos continuaram sendo parte importante da base de informações para as decisões no gerenciamento dos recursos pesqueiros, com destaque para as diretrizes do "Plano de Gestão para uso Sustentável de Lagostas no Brasil" (IBAMA, 2008).

Na grande maioria dos casos, os trabalhos que deram e dão suporte à escolha dos períodos de defeso, foram baseados na análise de registros feitos em determinados anos isoladamente e não levam explicitamente em consideração o acúmulo de dados anteriores a cada um dos anos avaliados. Até então, a única análise em que há o uso dos dados de maneira menos fragmentada é a de Mesquita (1973), que realizou um estudo cronológico da reprodução da lagosta *P. argus* em águas costeiras do Ceará no período de 1962 a 1970, com decomposição simples da série temporal. Os resultados indicaram que a época de maior intensidade reprodutiva corresponde ao período de março a junho e que houve queda no número de lagostas em reprodução no decorrer do período analisado. Apesar de não ser frequente com lagostas, há de maneira mais geral aplicações de modelagem de séries temporais para outros crustáceos, incluindo estudos sobre ciclos reprodutivos (*e.g.* PÉREZ-GONZÁLEZ *et al.*, 2012; CASTIGLIONI *et al.*, 2013; LIMA *et al.*, 2013; SUNDELÖF *et al.*, 2013; MAXWELL *et al.*, 2013; BERTELSEN, 2013; MILTON *et al.*, 2014).

Os modelos lineares generalizados, propostos por Nelder e Wedderburn (1972), são uma alternativa para a análise da série temporal das informações sobre a atividade reprodutiva, pois permitem integrar diferentes variáveis explicativas (*e.g.* mês, ano e sexo do exemplar) em uma formulação única, além de permitir que a distribuição de probabilidade da variável resposta pertença à família exponencial de distribuições. Os modelos lineares generalizados têm sido amplamente utilizados em

diversas áreas do conhecimento, como por exemplo na zootecnia com o intuito de estudar a prolificidade de caprinos (SARMENTO *et al.*, 2010), na ciência animal para o estudo da relação entre comprimento e peso de peixes (ROSSI *et al.*, 2007), na entomologia para estudar o efeito de fatores ambientais sobre o desenvolvimento e sobrevivência de pupas de *Ophyra albuquerquei* Lopes (KRUGER *et al.*, 2011), na economia para estudar o nível de evidenciação mínima das companhias brasileiras abertas durante o período inicial de harmonização das normas internacionais de contabilidade no Brasil (MAIA *et al.*, 2012).

Os modelos aditivos generalizados para posição, escala e forma (GAMLSS), apresentados por Rigby e Stasinopoulos (2005), são muito mais flexíveis e podem proporcionar um excelente ajuste na presença de distribuições altamente assimétricas e curtóticas. O GAMLSS permite o uso de uma grande variedade de distribuições para a variável resposta, e de várias funções de ligação para os efeitos das variáveis preditoras sobre a variável resposta. Existem cerca de 80 tipos diferentes de famílias de distribuições disponíveis no GAMLSS, com um, dois, três ou quatro parâmetros. Os dois primeiros parâmetros μ e σ caracterizam posição e escala, enquanto os dois parâmetros restantes ν e τ, se existirem, caracterizam forma, como por exemplo, assimetria e curtose. Todos os quatro parâmetros podem ser modelados como funções paramétricas lineares/não-lineares e/ou funções não-paramétricas suavizadas de variáveis explicativas (STASINOPOULOS e RIGBY, 2007).

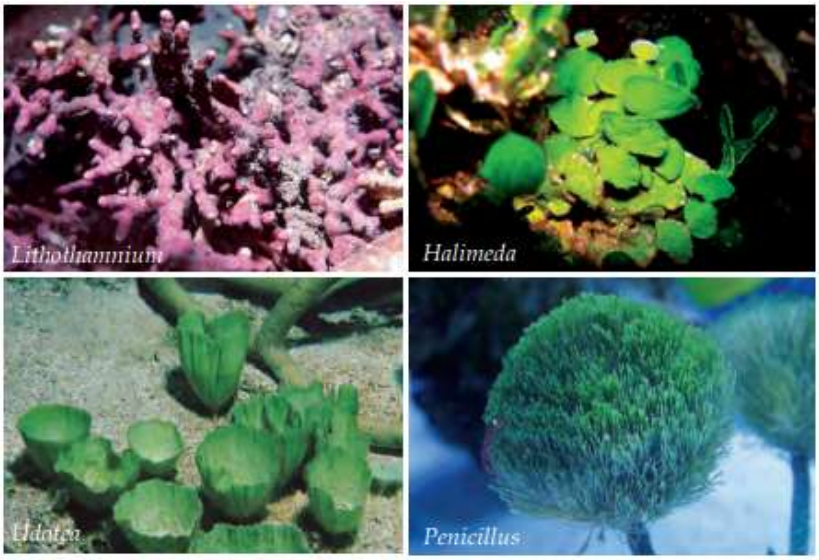
No presente trabalho propõe-se o uso do GAMLSS, além do GLM, para analisar os dados disponíveis sobre lagostas a fim de elucidar os padrões nas variações temporais das proporções de fêmeas em reprodução. Os resultados são comparados com os obtidos anteriormente (*e.g.* PAIVA e COSTA, 1970; MESQUITA, 1973; NASCIMENTO, 1984; SOARES e CAVALCANTE, 1985; SOARES, 1990; SOARES e FONTELES-FILHO, 2000) para obtenção de maior entendimento do ciclo reprodutivo e da adequacidade do período de defeso utilizado para a gestão dos estoques pesqueiros.

2.2 Revisão de literatura

2.2.1 Habitat e área de captura das lagostas

O habitat das lagostas é formado por substrato de algas calcárias bentônicas (Figura 1), sendo esses sedimentos compostos por algas vermelhas da família *Rhodophyceae*, principalmente do gênero *Lithothamnium*, com teores de 75-95% de carbonato de cálcio e 4-10% de carbonato de magnésio e, por algas verdes da família *Chlorophyceae*, principalmente dos gêneros *Halimeda, Udotea* e *Penicillus*. A riqueza de carbonato de cálcio, neste ambiente, é de extrema importância para a formação do exoesqueleto no ciclo de vida da lagosta (FONTELES-FILHO, 2000, 2007).

Figura 1 – Algas calcárias das classes *Rhodophyceae* e *Chlorophyceae*.



Fonte: IVO *et al.*, 2012.

A área de captura das lagostas compreende a plataforma continental das regiões Norte e Nordeste, entre os estados do Amapá e Espírito Santo, abrangendo uma superfície de 74.607km^2 (FONTELES-FILHO, 2007). IVO *et al.* (2012) acrescentam que, ao longo do tempo, os limites da área de captura foram sendo gradativamente estendidos a partir do seu núcleo central, correspondente à plataforma dos estados do Ceará e Pernambuco, até atingir a área total de 110.785km^2. Santana *et al*. (2016), fazendo uma avaliação das lagostas de espinhos,

destacam a ocorrência da espécie *Panulirus argus* desde o Amapá até São Paulo, além do Atol das Rocas e arquipélagos de Fernando de Noronha e São Pedro e São Paulo, já para a espécie *Panulirus laevicauda* foram registradas ocorrências desde o Amapá até Santa Catarina, além do Atol das Rocas e arquipélago de Fernando de Noronha.

As lagostas capturadas nas regiões são na grande maioria do gênero *Panulirus*, em particular de três espécies: *P. argus* - lagosta vermelha, *P. laevicauda* - lagosta verde e *P. echinatus* – lagosta pintada, sendo que a última é pouco expressiva nas pescarias. Estudos mostram que as duas principais espécies de lagostas capturadas (vermelha e verde) se distribuem e se deslocam perpendicularmente a costa, sendo que a lagosta vermelha é encontrada da faixa costeira até uma profundidade máxima que varia de 41 a 50m, enquanto que a lagosta verde é encontrada em profundidades máximas que variam de 31 a 40 metros (FONTELES-FILHO, 2000; DIAS-NETO, 2010).

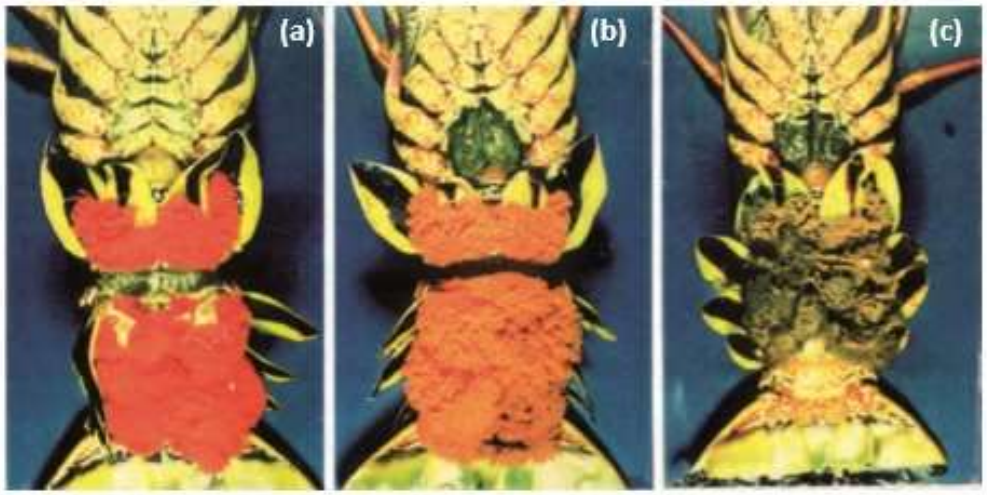
A área de captura é dividida em três subáreas de acordo com as características oceanográficas e o padrão de distribuição das lagostas: Norte (SN), Nordeste Setentrional (SNS) e Nordeste Oriental (SNO). A subárea mais propícia para as lagostas é o Nordeste Setentrional, devido à extensa plataforma continental e a existência de ótimas condições para formação e manutenção do substrato de algas calcárias. A lagosta vermelha, dominante em número e de maior porte, é encontrada com maior facilidade nas subáreas do Norte e Nordeste Setentrional, enquanto que a lagosta verde habita principalmente o Nordeste Setentrional, sendo menos abundante e ocupando menor espaço territorial (IVO *et al.*, 2012). Estudos recentes desenvolvidos por Araújo e Amaral (2016) na plataforma interna do Rio Grande do Norte confirmam a ocorrência de diversas feições geomorfológicas, tais como algas calcárias e bioclásticas, arrecifes de arenitos e bancos de macroalgas nessa região, que favorecem a ocorrência das lagostas.

2.2.2 Ciclo reprodutivo e maturidade sexual das lagostas

Paiva e Costa (1968a) dividem o ciclo de reprodução das lagostas em três fases: (1) acasalamento, cuja característica de ocorrência é a presença de massa espermatofórica na parte ventral do abdômen das fêmeas, seguindo-se a etapa da fecundação; (2) desenvolvimento embrionário, identificado pela presença de ovos

aderidos aos pleópodos das fêmeas e (3) pós-desova, identificada pela presença de restos de massa espermatofórica e ausência de ovos nas fêmeas. Já para Ivo *et al.* (2012) o ciclo de reprodução da lagosta inicia-se desde a maturação das gônadas (ovário) no interior do cefalotórax. As etapas seguintes compreendem: o acasalamento emparelhado do macho com a fêmea, em contato direto das respectivas regiões ventrais; a deposição da massa espermatofórica pelo macho sobre o esterno da fêmea; a liberação e fertilização dos óvulos maduros após rompimento da massa espermatofórica; a aglutinação dos óvulos e fertilização numa massa ovígera que fica aderida ao abdômen da fêmea por meio dos pleópodos e; pós-desova, identificada pela presença de restos de massa espermatofórica e ausência de ovos no abdômen. Na Figura 2 estão representadas algumas dessas fases.

Figura 2 – Ciclo reprodutivo da lagosta *Panulirus argus*. (a) – óvulos aderidos ao abdômen; (b) massa espermatofórica no esterno e massa ovígera aderida aos pleópodos; (c) restos de massa espermatofórica.



Fonte: PORTO, 1992.

As lagostas são espécies ovulíparas, com fecundação interna e desenvolvimento embrionário externo. O período de incubação dura de 4 a 6 semanas, quando ocorre a eclosão do filosoma, larva transparente de hábitos pelágicos e fototropismo positivo, que passa por até 11 estágios durante 10 meses (IVO *et al.*, 2012). Por meio de metamorfose a filosoma transforma-se em puerulus, pós-larva já com forma definitiva, aos 12 meses. Nessa ocasião apresenta hábitos pelágicos e passa a pós-puerulus, com hábitos bentônicos, após o endurecimento da

carapaça aos 18 meses. Já como juvenis (24-36 meses de idade), as lagostas adquirem a coloração típica da espécie e definem as características sexuais. Após atingirem a maturidade sexual, tornam-se adultos (36 meses) e reprodutores (48 meses) capazes de fechar o ciclo de formação de uma coorte (indivíduos de uma população na mesma faixa etária ou de tamanho), no período de uma geração que dura, em média, 4 anos (IVO *et al.*, 2012).

Os palinurídeos têm desovas parceladas, em termos individuais e populacionais, motivo pelo qual são encontrados exemplares em reprodução durante todos os meses do ano (DIAS-NETO e DIAS, 2015). Durante o período de desova, ocorre uma migração para zonas mais profundas (40 a 50 metros da costa) com velocidade média de 133m/dia (FONTELES-FILHO, 2007). Os valores médios de produção de ovos são 294.175 ovos e 630 ovos/g (*P. argus*), e 166.036 e 597 ovos/g (*P. laevicauda*) (IVO *et al.*, 2012).

2.2.3 Ciclo migratório das lagostas

Quando as larvas filosoma são levadas por correntes para a zona costeira inicia-se o ciclo migratório, que prossegue com a descida das pós-larvas puerulus para a zona bentônica, que após o endurecimento da carapaça, atingem o estado juvenil nas zonas de criação, e se dispersam para as zonas de alimentação. Quando se tornam maduros, migram para zonas ainda mais afastadas da costa para realizar a desova, dando início a um novo ciclo com a liberação de ovos para o meio ambiente (IVO *et al.*, 2012).

De acordo com Fonteles-Filho e Ivo (1980), as lagostas realizam migração durante o primeiro e segundo trimestre para as áreas de desova que se localizam em regiões mais afastadas da costa, com profundidades entre 40 e 50 metros, através de migração direcional, perpendicular à costa, em que predomina a baixa dispersão aleatória. No terceiro e quarto trimestres, as lagostas se dispersam em sentido paralelo à costa, quando realizam migração trófica, em que predominam movimentos aleatórios relacionados à procura de alimentos. Santana *et al*. (2016), fazendo uma avaliação das lagostas de espinhos, destacam que, em função do ciclo biológico e padrões de dispersão, a distribuição temporal não é uniforme ao longo do ano.

2.2.4 Dieta alimentar, crescimento e longevidade das lagostas

As lagostas são animais omnívoros (preferencialmente carnívoros) e de hábitos noturnos. Sua dieta alimentar consiste de moluscos gastrópodes e crustáceos (alimentos essenciais), equinodermos, algas, cnidários e briozoários (alimentos secundários) e esponjas (alimentos ocasionais) (SANTANA *et al*., 2016).

As lagostas se encontram no quarto nível trófico da cadeia alimentar, são espécies de ciclo longo, tendo como característica uma taxa de crescimento apenas mediana. A lagosta-vermelha tem crescimento mais lento, mas atinge maiores comprimentos que a lagosta-verde. As taxas anuais de crescimento em termos absoluto e relativo são de 2,6 cm/ano e 24,4 %/ano (vermelha), e 2,4 cm/ano e 25,7 %/ano (verde). A longevidade na parcela do estoque sujeita à captura e no ciclo de vida geral são estimados, respectivamente, em 13,9 anos e 37,3 anos para *P. argus*, e 12,5 anos e 34,7 anos para *P. laevicauda* (FONTELES-FILHO, 2007; IVO *et al.*, 2012).

2.2.5 Modelos de regressão estatísticos

A análise de dados é uma ferramenta fundamental para o conhecimento do atual estado dos estoques pesqueiros e é crucial para o planejamento e ordenamento da atividade. Costa *et al*. (2013) destacam que estudos relacionados a análises da dinâmica temporal da produção de pescado se tornam importantes à medida que podem auxiliar políticas de fomento e direcionar programas de governos na busca do avanço econômico local e/ou regional. Sendo assim, a análise estatística de dados temporais técnico-científicos e avaliações de estoque para subsidiar o adequado ordenamento da pesca são imprescindíveis.

Modelos estatísticos são instrumentos extremamente úteis para auxiliar na compreensão de aspectos essenciais de um conjunto de dados (LOBATO JUNIOR, 2010). Um dos métodos estatísticos mais utilizados para ajustar um modelo matemático para explicar o fenômeno observado e proporcionar predições dentro e, se aplicável, fora dos limites investigados é a técnica de análise de regressão, que tem como principal objetivo, verificar a existência de uma relação funcional entre uma variável resposta com uma ou mais variáveis explicativas. (AMADOR *et al.*, 2011).

De acordo com Paula (2013), durante muitos anos os modelos de regressão normais lineares foram utilizados na tentativa de descrever a maioria dos fenômenos aleatórios. Mesmo quando o fenômeno sob estudo não apresentava uma resposta para a qual fosse razoável a suposição de normalidade, algum tipo de transformação era sugerida (*e.g.* BOX e COX, 1964). Com o desenvolvimento computacional ocorrido na década de 70, Nelder e Wedderburn (1972) propuseram os modelos lineares generalizados (GLM), como uma extensão dos modelos lineares clássicos. A ideia básica consiste em abrir o leque de opções para a distribuição da variável resposta, permitindo que a mesma pertença à família exponencial de distribuições, bem como dar maior flexibilidade para a relação funcional entre a média da variável resposta e o preditor linear.

Como nem sempre a relação da variável resposta com as variáveis explicativas é de natureza linear, o uso de termos aditivos não lineares pode ser necessário (HASTIE e TIBSHIRANI, 1990), o que levou então à proposta de Modelos Aditivos Generalizados (GAM). Os GAMs correspondem a generalizações de modelos de regressão não-paramétricos, sua estrutura geral é formada pela variável resposta, uma função suavizadora (*smoother function*) como *loess* ou *cubic spline*, que incorpora o efeito não linear das variáveis explicativas sobre a variável resposta e o erro do modelo. Tanto os GLMs quanto os GAMs restringem a variável resposta a distribuições de probabilidade pertencentes à família exponencial e não admitem variáveis explicativas para modelar outros parâmetros que não seja a média.

Posteriormente surgiram os modelos lineares generalizados mistos (GLMM) (BRESLOW e CLAYTON, 1993; MCCULLOCH e SEARLE, 2001), utilizados para descrever um conjunto de dados cuja estrutura de tratamentos envolve alguns fatores que são fixos e outros que são aleatórios. Neste caso, pode-se modelar a variância dos dados, além da média. Com a inclusão de efeitos fixos e aleatórios aos fatores, o modelo misto pode acomodar covariâncias entre as observações, sendo assim uma opção para a análise de dados longitudinais. Uma de suas grandes vantagens é a possibilidade de se incorporar na análise uma matriz de variância-covariância com um menor número de parâmetros que a de modelos multivariados gerais.

Os modelos aditivos generalizados para posição, escala e forma (GAMLSS), propostos por Rigby e Stasinopoulos (2005), é uma classe de modelos bastante flexível, pois permite que a variável resposta pertença a uma família mais ampla que a família exponencial. Além de considerar uma estrutura sistemática para o parâmetro

de posição μ, são também considerados parâmetros de escala (σ) e de forma (ν e τ), que podem ser associados a grupos de variáveis explicativas, a partir de um preditor linear η. O preditor η, é composto por duas componentes, uma paramétrica e outra não-paramétrica. A segunda componente possibilita o uso de funções suavizadoras, e a inclusão de termos de efeitos aleatórios.

Diversas classes de modelos estatísticos existentes podem ser usadas para lidar com diferentes problemáticas na área da pesca, de forma a auxiliar na tomada de decisões para o manejo e sustentabilidade dos estoques pesqueiros. Como exemplo cita-se o controle do número de leões marinhos (*Otaria flavescens*) da América do Sul (GONZÁLEZ *et al.*, 2015), a interferência dos fatores ambientais na Captura por Unidade de Esforço (CPUE) do agulhão negro (*Makaira nigricans*) (DE OLIVEIRA *et al.*, 2015), a avaliação de mudanças interanuais no tamanho da maturidade de uma sardinha (*Strangomera bentincki*) na região centro-sul do Chile (BUSTOS e CUBILLOS, 2016), a estratificação de uma área de estudo para padronizar dados de captura por unidade de esforço (CPUE) (ONO *et al.*, 2015). Cabe mencionar ainda o uso de modelos estatístico no monitoramento e gerenciamento de ameaças emergentes e contínuas de doenças marinhas (TANAKA *et al.*, 2017).

2.2.6 Modelos lineares generalizados

Os modelos lineares generalizados (GLM), introduzidos por Nelder e Wedderburn (1972), apresentam estrutura de regressão linear e possuem em comum o fato da variável resposta seguir uma distribuição pertencente à família exponencial. Os GLMs são constituídos por três componentes (MCCULLAGH e NELDER, 1989):

a) aleatório – $\boldsymbol{y}$ um vetor de observações de comprimento n, denotando uma realização de um vetor de variáveis aleatórias Y, independentemente distribuídas com média μ. A distribuição de Y pode ser qualquer uma da família exponencial, logo a função densidade de probabilidade (f.d.p) de Y_i é dada por:

$$f(y_i; \theta_i; \emptyset) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a_i(\emptyset)} + c(y_i, \emptyset)\right\}, \qquad (1)$$

em que $b(\cdot)$ e $c(\cdot)$ são funções conhecidas, θ_i é o parâmetro natural ou canônico e, em geral, $a_i(\emptyset) = \frac{\emptyset}{w_i}$, sendo w_i pesos a priori e $\emptyset > 0$, conhecido, denominado de parâmetro de dispersão ou escala. Dessa forma:

$$\mathrm{E}(Y_i) = \mu_i = b'(\theta_i) \tag{2}$$

$$\mathrm{Var}(Y_i) = a_i(\emptyset)b''(\theta_i) = a_i(\emptyset)V(\mu_i) = a_i(\emptyset)V_i \tag{3}$$

em que $V_i = \frac{d\mu_i}{d\theta_i}$ é chamada de função de variância, e como depende unicamente da média tem-se que o parâmetro natural pode ser expresso como:

$$\theta_i = \int V_i^{-1}\, d\mu_i = q(\mu_i) \tag{4}$$

para $q(\mu_i)$ uma função conhecida de μ_i.

b) preditor linear – a parte sistemática do modelo, na qual as variáveis explicativas entram na forma de uma soma linear dos seus efeitos

$$\eta_\mathrm{i} = \sum_{j=1}^{p} x_{ij}\, \beta_j = \mathbf{x}_i^T \boldsymbol{\beta} \quad \text{ou} \quad \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}, \tag{5}$$

em que $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^T$ é a matriz do modelo, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$ é um vetor de parâmetros desconhecidos a serem estimados e $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_n)^T$ é o preditor linear.

c) função de ligação (g) – faz a conexão entre a média das observações (componente aleatório) e a parte sistemática, na forma

$$g(\mu_\mathrm{i}) = \eta_\mathrm{i} = \mathbf{x}_i^T \boldsymbol{\beta}, \tag{6}$$

sendo $g(.)$ uma função monótona e diferenciável.

Segundo Dobson (2008) para a formulação de um GLM deve-se efetuar uma análise exploratória dos dados considerando cada variável isoladamente a fim de

verificar a qualidade, distribuição e características gerais. Em seguida, devem ser elaborados modelos que envolvem a variável resposta com n realizações independentes e respectivas realizações das potenciais variáveis explicativas.

Os GLM têm sido amplamente utilizados em ciências pesqueiras. A aplicação mais comum é a padronização de capturas por unidade de esforço (CPUE) para estimar índices relativos de abundância com base em dados comerciais. Maunder e Langley (2004) aplicaram este método ao estoque de *Pseudocaranx dentex* na costa oeste da Nova Zelândia para verificar se o modelo de avaliação de estoque explica toda a variação anual nos dados de CPUE e avaliar como os dados de captura por idade são ponderados. Hazin *et al.* (2007) analisaram dados de 62007 lances de espinhel entre os anos de 1978 a 2005 para padronizar a série de espadartes capturados por espinhéis brasileiros no Oceano Atlântico. Andrade (2007) estimou taxas de captura para o bonito-listrado (*Katsuwonus pelamis*) no sudoeste do oceano Atlântico Sul com o uso de modelos lineares generalizados. Mourato *et al.* (2016) também utilizaram GLM para gerar uma série de captura por unidade de esforço para o agulhão-vela capturado pela pesca esportiva baseada nos estados de São Paulo, Rio de Janeiro, Espírito Santo e Bahia de 1996 a 2014.

2.2.6.1 Estimação dos parâmetros

Após a seleção de um modelo particular é necessário fazer a estimação dos parâmetros. Conforme Cordeiro e Demétrio (2008), o algoritmo de ajuste de um modelo linear generalizado é o dos mínimos quadrados iterados ponderados, mediante o qual são obtidos os estimadores de máxima verossimilhança dos parâmetros $\boldsymbol{\beta}$ do preditor linear.

O logaritmo da função de verossimilhança para um conjunto de observações independentes $y_1, y_2, \dots, y_n$ é dado pela soma das contribuições individuais, isto é,

$$l = l(\boldsymbol{\theta}; \mathbf{y}) = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} l(\theta_i; y_i) = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \left\{ \frac{1}{a_i(\emptyset)} [y_i \theta_i - b(\theta_i)] + c(y_i; \emptyset) \right\}, \qquad (7)$$

sendo que $\mathrm{E}(Y_i) = \mu_i$, $\eta_i = g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$ e $\theta_i = \int V_i^{-1}\, d\mu_i = q(\mu_i)$.

Uma propriedade da família exponencial de distribuições é que seus elementos satisfazem condições de regularidade suficientes para assegurar que o máximo global do logaritmo da função de verossimilhança $l(\boldsymbol{\theta};\mathbf{y})$ é dado unicamente pela solução do sistema de equações $\mathbf{U}_{\boldsymbol{\theta}} = \frac{dl}{d\boldsymbol{\theta}} = 0$ ou equivalentemente, $\mathbf{U}_{\boldsymbol{\beta}} = \frac{dl}{d\boldsymbol{\beta}} = 0$. Tem-se, então que a função escore é

$$U_j = \sum_{i=1}^{n} \frac{\partial l(\theta_i; y_i; \emptyset)}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial l_i}{\partial \beta_j} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \frac{1}{a_i(\emptyset)} (y_i - \mu_i) \frac{1}{V(\mu_i)} \frac{d\mu_i}{d\eta_i} \mathrm{x}_{ij}. \quad (8)$$

Logo, o estimador de máxima verossimilhança $\hat{\beta}$ é a solução da equação $U_j = 0$, $j = 1,2, \dots, p$. Em geral, os resultados obtidos dessa igualdade não são lineares e, dessa maneira, deve-se obter uma solução numérica utilizando processos iterativos do tipo Newton-Raphson. O método iterativo de Newton-Raphson para a solução de uma equação $f(x) = 0$ é baseado na aproximação de Taylor para a função $f(x)$ nas vizinhanças do ponto x_0, ou seja,

$$f(x) = f(x_0) + (x - x_0) f'(x_0) = 0, \quad (9)$$

obtendo-se

$$x = x_0 - \frac{f(x_0)}{f'(x_0)} \quad (10)$$

ou, de uma forma mais geral,

$$x^{(m+1)} = x^{(m)} - \frac{f\left(x^{(m)}\right)}{f'(x^{(m)})}, \quad (11)$$

em que $x^{(m+1)}$ o valor de x no passo $(m+1)$, $x^{(m)}$ o valor de x no passo m, $f(x^{(m)})$ a função $f(x)$ avaliada em $x^{(m)}$ e $f'(x^{(m)})$ a derivada da função $f(x)$ avaliada em $x^{(m)}$.

Utilizando-se a versão multivariada do método de Newton-Raphson para solucionar o sistema de equações $\mathbf{U}_{\boldsymbol{\beta}} = \frac{dl}{d\boldsymbol{\beta}} = 0$, tem-se:

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{m} + (\mathbf{I}_0^{-1})^{(m)} \mathbf{U}^{(m)}, \quad (12)$$

em que $\boldsymbol{\beta}^m$ e $\boldsymbol{\beta}^{m+1}$ são os vetores de parâmetros estimados nos passos m e $(m+1)$, $\mathbf{U}^{(m)}$ o vetor escore, que corresponde ao vetor de derivadas parciais de 1ª ordem de $f(x)$, com elementos $\frac{\partial l}{\partial \beta_j}$, avaliado no passo m e $(\mathbf{I}_0^{-1})^{(m)}$ a inversa da negativa da matriz de derivadas parciais de 2ª ordem de $f(x)$, com elementos $\frac{-\partial^2 l}{\partial \beta_j \, \partial \beta_k}$, avaliada no passo m.

O método de Newton-Raphson é bastante útil quando as derivadas de 2ª ordem são obtidas facilmente. Quando isso não ocorre, utiliza-se o método escore de Fisher que envolve a substituição da matriz de informação observada $\mathbf{I}_0$, pela matriz de informação esperada de Fisher, $\mathfrak{I}$. Logo,

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^m + (\mathfrak{I}^{-1})^{(m)}\mathbf{U}^{(m)} \tag{13}$$

em que $\mathfrak{I}$ tem elementos dados por $\mathfrak{I}_{jk} = E\left[\frac{-\partial^2 l}{\partial \beta_j \, \partial \beta_k}\right] = E\left[\frac{\partial l}{\partial \beta_j}\frac{\partial l}{\partial \beta_k}\right]$, que é a matriz de covariâncias dos $U_j's$.

Multiplicando-se ambos os lados de (13) por $\mathfrak{I}^{(m)}$ tem-se

$$\mathfrak{I}^{(\mathrm{m})}\boldsymbol{\beta}^{(\mathrm{m+1})} = \mathfrak{I}^{(\mathrm{m})}\boldsymbol{\beta}^{\mathrm{m}} + \mathbf{U}^{(\mathrm{m})}. \tag{14}$$

Substituindo-se em (8), obtém-se

$$\mathfrak{I}_{jk} = E\left(U_j U_k\right) = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \frac{1}{a_i(\emptyset)} \frac{1}{V(\mu_i)} \left(\frac{d\mu_i}{d\eta_i}\right)^2 x_{ij} x_{ik} \tag{15}$$

e fazendo-se $a_i(\emptyset) = \frac{\emptyset}{w_i}$, com $\emptyset > 0$, constante, w_i peso a priori e $W_i = \frac{w_i}{V(\mu_i)}\left(\frac{d\mu_i}{d\eta_i}\right)^2$, tem-se que:

$$\mathfrak{I} = \frac{1}{\emptyset}\mathbf{X}^T\mathbf{W}\mathbf{X} \tag{16}$$

com elementos $\mathfrak{I}_{jk} = \sum_{i=1}^{n} \frac{1}{\emptyset} x_{ij} \, W_i x_{ik}$, $\mathbf{X}$, a matriz do modelo e $\mathbf{W} = \mathrm{diag}(W_1, W_2, \dots, W_n)$.

Em particular, para as funções de ligação canônicas $W_i = w_i V(\mu_i)$, pois $\frac{d\theta_i}{d\mu_i} = \frac{d\eta_i}{d\mu_i} = V^{-1}(\mu_i)$.

Além disso, rearranjando-se os termos de U_j tem-se

$$U_j = \sum_{i=1}^{n} \frac{w_i}{\emptyset} \frac{(y_i - \mu_i)}{V(\mu_i)} \frac{d\mu_i}{d\eta_i} x_{ij} = \sum_{i=1}^{n} \frac{1}{\emptyset} x_{ij} W_i \frac{d\eta_i}{d\mu_i} (y_i - \mu_i) \tag{17}$$

e, portanto o vetor escore $\mathbf{U}$ fica

$$\mathrm{U} = \frac{1}{\emptyset} \mathrm{X}^T \mathrm{W} \Delta (\mathrm{y} - \mu) \tag{18}$$

com $\Delta = \mathrm{diag}\left\{\frac{d\eta_1}{d\mu_1}, \frac{d\eta_2}{d\mu_2}, \dots, \frac{d\eta_n}{d\mu_n}\right\} = \mathrm{diag}\{g'(\mu_1), g'(\mu_2), \dots, g'(\mu_n)\}$.

Logo, substituindo em $\mathfrak{I}$ e $\mathbf{U}$ em (14)

$$\frac{1}{\emptyset} \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X} \boldsymbol{\beta}^{(m+1)} = \frac{1}{\emptyset} \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X} \boldsymbol{\beta}^{(m)} + \frac{1}{\emptyset} \mathbf{X}^T \mathbf{W}^{(m)} \Delta^{(m)} (\mathbf{y} - \boldsymbol{\mu})^{(m)} \tag{19}$$

$$\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X} \boldsymbol{\beta}^{(m+1)} = \mathbf{X}^T \mathbf{W}^{(m)} \left[\mathbf{X} \boldsymbol{\beta}^{(m)} + \Delta^{(m)} (\mathbf{y} - \boldsymbol{\mu})^{(m)}\right] \tag{20}$$

e fazendo-se $\mathbf{z}^{(m)} = \mathbf{X} \boldsymbol{\beta}^{(m)} + \Delta^{(m)} (\mathbf{y} - \boldsymbol{\mu})^{(m)} = \eta^{(m)} + \Delta^{(m)} (\mathbf{y} - \boldsymbol{\mu})^{(m)}$, denominada variável dependente ajustada, tem-se

$$\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X} \boldsymbol{\beta}^{(m+1)} = \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{z}^{(m)} \tag{21}$$

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{z}^{(m)} \tag{22}$$

que tem a forma da solução das equações normais para o modelo linear obtido pelo método dos mínimos quadrados ponderados, exceto que nesse caso a solução $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}^{(m+1)}$ deve ser resolvida iterativamente. Vale ressaltar que a expressão (22) independe de $\emptyset$.

Usualmente, o processo iterativo começa com a especificação de uma estimativa inicial $\boldsymbol{\beta}^{(0)}$ e sucede com a sua alteração até que a convergência seja obtida e, portanto, $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}^{(m+1)}$. Ressalta-se que cada observação pode ser considerada como uma estimativa do seu valor médio, isto é, $\hat{\mu} = y_i$ e, portanto,

$$\hat{\eta}_i = g(\hat{\mu}_i) = g(y_i). \tag{23}$$

Usando-se $\hat{\boldsymbol{\eta}}$ como a variável dependente e $\mathbf{X}$, a matriz do modelo, obtém-se o vetor $\boldsymbol{\beta}^{(0)}$.

Os métodos mais usados para a estimação de $\emptyset$ são o método da máxima verossimilhança, o método dos momentos e perfil de verossimilhança. No entanto, em algumas situações o método da verossimilhança pode-se tornar computacionalmente intratável quando não existe solução explícita. A estimativa de máxima verossimilhança para $\emptyset$, interpretando o logaritmo da função de verossimilhança $l(\boldsymbol{\beta}, \emptyset, \mathbf{y})$ como função de β e de $\emptyset$, dado $\mathbf{y}$, é obtida fazendo-se

$$\frac{\partial l(\boldsymbol{\beta}, \emptyset)}{\partial \emptyset} = 0. \tag{24}$$

É importante destacar que, para a distribuição binomial e Poisson, $\emptyset = 1$.

Um estimador consistente para $\emptyset$ (de momentos) que não envolve processo iterativo é baseado na estatística de Pearson, sendo dado por (PAULA, 2013):

$$\widehat{\emptyset} = \frac{n-p}{\sum_{i=1}^{n} \left\{ \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} \right\}} \tag{25}$$

2.2.6.2 Teste de hipótese

Na teoria de decisão estatística, os testes de hipóteses assumem uma importância fundamental, visto que permitem fazer inferências sobre uma população com base numa amostra. Além disso, os testes possibilitam avaliar se os parâmetros são diferentes de zero e comparar a qualidade de ajuste de vários modelos. De acordo com a teoria de máxima verossimilhança, para avaliar hipóteses relativas aos parâmetros $\beta's$ pode ser utilizado o teste da razão de verossimilhanças e o teste de Wald (CORDEIRO e DEMÉTRIO, 2008). Para as descrições dos mesmos, considere as hipóteses

$$H_0: \boldsymbol{\beta}_1 = \boldsymbol{\beta}_{1,0}$$
$$H_a: \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_{1,0}$$

sendo $\boldsymbol{\beta}_{1,0}$ um valor especificado para $\boldsymbol{\beta}_1$. Seja $\widehat{\boldsymbol{\beta}} = \left[\widehat{\boldsymbol{\beta}}_1^T \ \widehat{\boldsymbol{\beta}}_2^T\right]^T$ o estimador de máxima verossimilhança para $\boldsymbol{\beta}$ sem restrição e $\widehat{\boldsymbol{\beta}}_0 = \left[\boldsymbol{\beta}_{1,0}^T \ \boldsymbol{\beta}_{2,0}^T\right]^T$ em que $\widehat{\boldsymbol{\beta}}_{2,0}$ é o estimador de máxima verossimilhança para $\boldsymbol{\beta}_2$ sob H_0.

2.2.6.2.1 Teste da razão de verossimilhanças

O teste da razão de verossimilhança (TRV) é preferível quando o interesse é avaliar hipóteses relativas a vários coeficientes $\boldsymbol{\beta}'\mathbf{s}$. Baseia-se no logaritmo da razão entre duas verossimilhanças e tem como finalidade avaliar se um modelo com menos parâmetros se ajusta quase tão bem quanto um modelo saturado com grande quantidade de parâmetros (sem restrição). A estatística do teste é dada por:

$$\Lambda = -2\ln\lambda = 2[l(\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2; \mathbf{y}) - l(\boldsymbol{\beta}_{1,0}, \widehat{\boldsymbol{\beta}}_{2,0}; \mathbf{y})], \tag{26}$$

em que $\boldsymbol{l}(\widehat{\boldsymbol{\beta}}_{\mathbf{1}}, \widehat{\boldsymbol{\beta}}_{\mathbf{2}}; \mathbf{y})$ representa o valor maximizado da log-verossimilhança do modelo saturado e $\boldsymbol{l}(\boldsymbol{\beta}_{\mathbf{1,0}}, \widehat{\boldsymbol{\beta}}_{\mathbf{2,0}}; \mathbf{y})$ o valor maximizado da log-verossimilhança do modelo restrito. Rejeita-se $\boldsymbol{H_0}$, a um nível de significância $\boldsymbol{\alpha}$, se $\boldsymbol{\Lambda} > \boldsymbol{\chi}^2_{\boldsymbol{q},\mathbf{1}-\propto}$ (CORDEIRO e DEMÉTRIO, 2008; DOBSON, 2008).

2.2.6.2.2 Teste de Wald

O teste de Wald é o mais utilizado para hipóteses relativas a um único coeficiente β_j. Baseia-se na distribuição normal assintótica de $\widehat{\boldsymbol{\beta}}$ e é uma generalização da estatística *t* de *Student* (WALD, 1943). A vantagem com relação ao teste de verossimilhança é que não necessita do cálculo do $\widehat{\boldsymbol{\beta}}_{2,0}$. A estatística do teste é dada por:

$$W = (\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,0})^T [\widehat{\mathrm{V}}\mathrm{ar}(\widehat{\boldsymbol{\beta}}_1)]^{-1} (\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,0}), \tag{27}$$

sendo $\widehat{\mathrm{V}}\mathrm{ar}(\widehat{\boldsymbol{\beta}}_1)$ a $\mathrm{Var}(\widehat{\boldsymbol{\beta}}_1)$ avaliada em $\widehat{\boldsymbol{\beta}} = [\widehat{\boldsymbol{\beta}}_1^T\ \widehat{\boldsymbol{\beta}}_2^T]^T$. Para amostras grandes, rejeita-se H_0, a um nível de significância α, se $W > \chi^2_{q,1-\propto}$.

2.2.6.3 Análise de deviância e seleção de modelos

A deviância é utilizada como uma medida de discrepância para verificação do ajuste de modelos (Nelder e Wedderburn, 1972). A análise de deviância é uma generalização da análise da variância para os modelos lineares generalizados, que

visa obter, a partir de uma sequência de modelos, cada um incluindo mais termos do que os anteriores, os efeitos de fatores, covariáveis e suas interações. A deviância deve ser utilizada para comparar modelos encaixados, ou seja, modelos que são obtidos a partir de um outro impondo alguma restrição aos parâmetros.

Sejam os modelos M_p e M_q $(p < q)$ com p e q parâmetros, respectivamente. A estatística $D_p - D_q$ com $(q - p)$ graus de liberdade é interpretada como uma medida de variação dos dados, explicada pelos termos que estão em M_q e não estão em M_p, incluídos os efeitos dos termos em M_p e ignorando quaisquer efeitos dos termos que não estão em M_q. Tem-se, assintoticamente, para $\emptyset$ conhecido, que

$$S_p - S_q = \frac{1}{\emptyset}\left(D_p - D_q\right) \sim \chi^2_{q-p}, \tag{28}$$

que é simplesmente o teste da razão de verossimilhanças definido na sessão 2.2.6.2.1, em que, S_p e S_q são denominados de desvio escalonado sob os modelos M_p e M_q, respectivamente; D_p e D_q são denominados de desvio sob os modelos M_p e M_q, respectivamente. Os desvios D_p e D_q são funções apenas dos dados y e das médias ajustadas $\hat{\mu}_i$, enquanto os desvios escalonados S_p e S_q dependem de D_p e D_q, respectivamente, e do parâmetro de dispersão $\emptyset$.

A deviância é sempre maior ou igual a zero e, à medida que entram as variáveis explicativas (ou covariáveis) no componente sistemático, decresce até se tornar zero para o modelo saturado. Quanto melhor for o ajuste aos dados tanto menor será o valor da deviância. Assim, um modelo bem ajustado aos dados com uma verossimilhança grande tem uma deviância pequena.

Outra medida comumente utilizada para verificar a qualidade do ajuste em modelo linear generalizado é o R^2 de McFadden, nomeado por Dobson (2008) como aumento proporcional na deviância explicada. Sua interpretação é análoga ao coeficiente de determinação (R^2) dos modelos de regressão linear. Essa estatística indica o quanto da variância da variável resposta é explicada pela variância das variáveis explicativas. Seu valor está compreendido entre o intervalo 0 a 1. Quanto maior o valor, mais explicativo é o modelo. O Pseudo R^2 de McFadden é calculado pela razão da log-verossimilhança do modelo ajustado com o modelo composto apenas do intercepto (modelo nulo). Dessa forma, tem-se que:

$$pseudo\ R^2 = 1 - \frac{L_1}{L_0}, \tag{29}$$

em que:

L_0= log-verossimilhança do modelo nulo;

L_1= log-verossimilhança do modelo ajustado.

Na etapa de seleção de modelos, o Critério de Informação de Akaike (AIC) (AKAIKE, 1974) e o Critério de Informação Bayesiano (BIC) (SCHWARZ, 1978) vêm sendo bastante utilizados para auxiliar na escolha do modelo mais adequado para representar um conjunto de dados. Esses critérios são baseados no máximo da função de verossimilhança como medida de ajustamento e são definidos como:

$$AIC = -2l(\widehat{\boldsymbol{\beta}}) + 2p \tag{30}$$

$$BIC = -2l(\widehat{\boldsymbol{\beta}}) + p\log(n), \tag{31}$$

em que $l(\widehat{\boldsymbol{\beta}})$ representa o logaritmo da sua função de verossimilhança avaliada no seu máximo, p o número de parâmetros do modelo investigado e n número de observações. Como regra de decisão, adota-se como modelo mais adequado o que apresenta menor valor na estatística. No AIC se pesa o balanço entre viés e variância. No entanto, segundo Wagenmakers e Farrell (2004), quando o número de observações n cresce muito, a probabilidade de que o AIC selecione o verdadeiro modelo com baixa dimensão entre os diversos avaliados não se aproxima da unidade, sendo mais adequado o uso do BIC.

Para examinar se os modelos diferem do melhor modelo por um parâmetro e têm essencialmente os mesmos valores da log-verossimilhança maximizada, outra medida que pode ser utilizado é o Delta AIC, sugerido por Burnham e Anderson (2002). A estatística Delta AIC é definida pela diferença entre os valores AIC para dois modelos aninhados. Modelos com Delta AIC entre 0 e 2 unidades merecem atenção e devem ser examinados mais detalhadamente. Modelos com Delta AIC entre 4 e 7 tem suporte consideravelmente pequeno. Modelos com Delta AIC>10 não têm essencialmente suporte e podem ser omitidos de uma consideração mais aprofundada, ou pelo menos esses modelos não conseguem explicar uma variação substancial nos dados.

2.2.6.4 Análise dos resíduos

Os resíduos são ferramentas importantes para escolha de um modelo estatístico. De acordo com Paula (2013) a análise de resíduos é utilizada para avaliar a adequação da distribuição proposta para a variável resposta e detectar a presença de pontos aberrantes ou observações influentes que exercem um peso desproporcional nas estimativas dos parâmetros do modelo. Dobson (2008) classifica observações que destoam do conjunto de dados 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).

Os tipo de resíduos mais utilizados para os GLMs são os resíduos ordinários, os resíduos de Pearson Studentizado e o resíduo componente do desvio padronizado. Os resíduos ordinários expressam as discrepâncias entre os valores observados y_i e seus valores ajustados $\hat{\mu}_i$, sendo definidos por: $r_i = y_i - \hat{\mu}_i$. No entanto, os resíduos ordinários não são muito informativos, por não apresentar variância constante. Sendo assim, os resíduos de Pearson Studentizado são comumente utilizados e são dados por:

$$t_{S_i} = \frac{\emptyset^{1/2}(y_i - \hat{\mu}_i)}{\sqrt{\hat{V}_i(1 - h_{ii})}} \tag{32}$$

em que h_{ii} é o i-ésimo elemento da diagonal da matriz de projeção $\boldsymbol{H}$, onde $\boldsymbol{H} = \boldsymbol{W}^{1/2}\boldsymbol{X}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}^{\mathbf{1/2}}$, $\boldsymbol{X}$ é a matriz do modelo e $\boldsymbol{W}$ a matriz de pesos. Os resíduos padronizados definidos a partir dos componentes da função desvio também são muito utilizados nos GLMs e são dados por:

$$t_{D_i} = \frac{\emptyset^{1/2} d(y_i; \hat{\mu}_i)}{\sqrt{(1 - h_{ii})}} \tag{33}$$

em que 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. Os resíduos padronizados podem ser comparados com a distribuição Normal para avaliar a adequação dos pressupostos distributivos e identificar quaisquer valores incomuns (DOBSON, 2008; PAULA, 2013).

2.2.6.4.1 Teste de Breusch-Pagan

O teste de Breusch-Pagan é um teste estatístico, baseado no teste multiplicador de Lagrange, que tem por objetivo testar a hipótese nula de que as variâncias dos erros são iguais (homocedásticas) contra a hipótese alternativa de que as variâncias dos erros são diferentes (heterocedásticas). É indicado para grandes amostras e quando a suposição de normalidade nos erros é assumida. O procedimento para realização do teste é dividido em cinco etapas:

Etapa 1: Ajusta-se o modelo de regressão e encontra-se os resíduos $\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_n$;

Etapa 2: Obtêm-se o estimador de máxima verossimilhança de σ^2, dado por: $\hat{\sigma}^2 = \sum \hat{r}_i^2 / n$;

Etapa 3: Constrói-se variáveis p_i, definidas como $p_i = \hat{r}_i^2 / \hat{\sigma}^2$, que são simplesmente cada resíduo elevado ao quadrado e dividido por $\hat{\sigma}^2$;

Etapa 4: Faz-se uma nova regressão para a variável resposta p construída sobre as variáveis explicativas;

Etapa 5: Obtêm-se a soma dos quadrados explicados pela regressão ($SQReg$) e define-se a estatística do teste como:

$$BP = \frac{SQReg}{2} \tag{34}$$

Pressupondo que r_i se distribui normalmente, se há homocedasticidade e se o tamanho da amostra n aumenta indefinidamente, então $BP \sim \chi^2_{m-1}$. Se a estatística do teste for maior que o valor crítico de χ^2 a um nível de significância prefixado, rejeita-se a hipótese de homocedasticidade (GUJARATI, 2006).

2.2.7 Modelos aditivos generalizados para posição, escala e forma (GAMLSS)

2.2.7.1 Definição

Na estrutura de regressão GAMLSS, os p parâmetros $\boldsymbol{\theta}^T = (\theta_1, \theta_2, \ldots, \theta_P)$ de uma função densidade de probabilidade $f(y|\boldsymbol{\theta})$ são modelados usando termos aditivos. Especificamente o modelo assume que, para $i = 1, 2, .., n$, as observações y_i

são independentes e condicionais a $\boldsymbol{\theta}^i$, com função densidade de probabilidade $f(y_i|\boldsymbol{\theta}^i)$, onde $\boldsymbol{\theta}^{iT} = (\theta_{i1}, \theta_{i2}, \dots, \theta_{iP})$ é um vetor de p parâmetros relacionado às variáveis explanatórias e efeitos aleatórios. Ressalta-se que quando os valores assumidos pelas covariáveis são estocásticos ou as observações y_i dependem de seus valores passados, então $f(y_i|\boldsymbol{\theta}^i)$ é entendida como sendo condicional a estes valores.

Seja $\boldsymbol{y}^T = (y_1, y_2, .., y_n)$ o vetor de observações da variável resposta. Além disso, para $k = 1, 2, .., p$, considere uma função de ligação monótona conhecida $g_k(.)$ relacionando o k-ésimo parâmetro $\boldsymbol{\theta}_k$ às variáveis explanatórias e efeitos aleatórios por meio de um modelo aditivo dado por:

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \boldsymbol{X}_k\boldsymbol{\beta}_k + \sum_{j=1}^{J_k} \boldsymbol{Z}_{jk}\gamma_{jk} \tag{35}$$

em que $\boldsymbol{\theta}_k$ e $\boldsymbol{\eta}_k$ são vetores de tamanho n, por exemplo, $\boldsymbol{\theta}_k^T = (\theta_{1k}, \theta_{2k}, \dots, \theta_{nk})$, $\boldsymbol{\beta}_k^T = (\beta_{1k}, \beta_{2k}, \dots, \beta_{J'_k k})$ é um vetor de parâmetros de tamanho J'_k, $\boldsymbol{X}_k$ é uma matriz de covariáveis conhecidas e de ordem $n \times J'_K$, $\boldsymbol{Z}_{jk}$ é uma matriz de covariáveis fixas conhecidas e de ordem $n \times q_{jk}$ e γ_{jk} é uma variável aleatória q_{jk}-dimensional. O modelo definido em (35) é denominado de GAMLSS (RIGBY e STASINOPOULOS, 2005).

Os vetores γ_{jk}, para $j = 1, 2, .., J_k$, podem ser combinados em um único vetor γ_k com uma única matriz de covariáveis $\boldsymbol{Z}_k$, no entanto, a formulação proposta em (35) é preferida por facilitar o uso dos algoritmos de retroajuste (*backfitting*) e por permitir que combinações de diferentes tipos de termos aditivos e/ou de efeitos aleatórios sejam facilmente incorporadas no modelo.

Se, para $k = 1, 2, .., p$, $J_k = 0$ então o modelo (35) se reduz a um modelo completamente paramétrico dado por:

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \boldsymbol{X}_k\boldsymbol{\beta}_k \tag{36}$$

Se $\boldsymbol{Z}_{jk} = \mathbf{I}_n$, em que $\mathbf{I}_n$ é uma matriz identidade de ordem $n \times n$, e $\gamma_{jk} = \mathbf{h}_{jk} = h_{jk}(\mathbf{x}_{jk})$ para todas as combinações de j e k no modelo (35), tem-se que:

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \boldsymbol{X}_k\boldsymbol{\beta}_k + \sum_{j=1}^{J_k} h_{jk}(\mathbf{x}_{jk}) \tag{37}$$

em que $\mathbf{x}_{jk}$ para $j = 1, 2, .., J_k$ e $k = 1, 2, .., p$ são vetores de tamanho n. A função h_{jk} é uma função desconhecida da variável explanatória X_{jk} e $\mathbf{h}_{jk} = h_{jk}(\mathbf{x}_{jk})$ é um vetor que avalia a função h_{jk} em $\mathbf{x}_{jk}$. Os vetores explanatórios $\mathbf{x}_{jk}$ são assumidos como sendo conhecidos. O modelo definido na equação (37), denominado de GAMLSS semiparamétrico, é um caso especial importante do modelo (35) e pode conter termos paramétricos, não paramétricos e de efeitos aleatórios.

Os dois primeiros parâmetros populacionais $\boldsymbol{\theta}_1$e $\boldsymbol{\theta}_2$ no modelo (35), geralmente são caracterizados como parâmetro de posição e escala, indicados por $\boldsymbol{\mu}$ e $\boldsymbol{\sigma}$, enquanto que os parâmetros restantes, se houver, são caracterizados como parâmetros de forma, embora o modelo pode ser aplicado de forma mais geral aos parâmetros de qualquer distribuição populacional.

Para muitas famílias de distribuições populacionais, um máximo de dois parâmetros de forma $\boldsymbol{\nu}(= \boldsymbol{\theta}_3)$ e $\boldsymbol{\tau}(= \boldsymbol{\theta}_4)$ é utilizado, resultando nos seguintes modelos:

$$\left.\begin{aligned}
g_1(\boldsymbol{\mu}) &= \boldsymbol{\eta}_1 = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \sum_{j=1}^{J_1} \boldsymbol{Z}_{j1}\gamma_{j1}\,, \\
g_2(\boldsymbol{\sigma}) &= \boldsymbol{\eta}_2 = \boldsymbol{X}_2\boldsymbol{\beta}_2 + \sum_{j=1}^{J_2} \boldsymbol{Z}_{j2}\gamma_{j2}\,, \\
g_3(\boldsymbol{\nu}) &= \boldsymbol{\eta}_3 = \boldsymbol{X}_3\boldsymbol{\beta}_3 + \sum_{j=1}^{J_3} \boldsymbol{Z}_{j3}\gamma_{j3}\,, \\
g_4(\boldsymbol{\tau}) &= \boldsymbol{\eta}_4 = \boldsymbol{X}_4\boldsymbol{\beta}_4 + \sum_{j=1}^{J_4} \boldsymbol{Z}_{j4}\gamma_{j4}\,.
\end{aligned}\right\} \tag{38}$$

O modelo GAMLSS (35) é mais geral que o GLM, o GAM, o GLMM ou o GAMM, na medida em que a distribuição da variável dependente não se limita à família exponencial e todos os parâmetros (não apenas a média) são modelados em termos de efeitos fixos e aleatórios (RIGBY e STASINOPOULOS, 2005).

2.2.7.2 Estimação dos parâmetros do modelo

Os vetores paramétricos β_k e os parâmetros de efeitos aleatórios γ_{jk} para $j = 1, 2, .., J_k$ e $k = 1, 2,3,4$ são estimados na estrutura GAMLSS (para valores fixos dos hiperparâmetros suavizados λ_{jk}) por meio da maximização da função de verossimilhança penalizada l_p, dada por:

$$l_p = l - \frac{1}{2}\sum_{k=1}^{4}\sum_{j=1}^{J_k}\gamma_{kj}^T \boldsymbol{G}_{kj}(\lambda_{kj})\gamma_{kj} \tag{39}$$

em que $l = \sum_{i=1}^{n} log\, f(y_i|\boldsymbol{\theta}^i) = \sum_{i=1}^{n} log\, f(y_i|\mu_i, \sigma_i, \nu_i, \tau_i)$ é o log da função de verossimilhança. Para o modelo paramétrico GAMLSS (36), a equação (39) será reduzida para l, e os vetores β_k serão estimados pela maximização desta função. Rigby e Stasinopoulos (2005) sugeriram dois algoritmos para o ajuste dos modelos GAMLSS. O primeiro, algoritmo CG, é uma generalização do algoritmo de Cole e Green (1992) e usa as derivadas de primeira ordem e os valores esperados (exatos ou aproximados) das derivadas de segunda ordem e das derivadas cruzadas da função de verossimilhança em relação aos parâmetros da distribuição. No entanto, para muitas funções densidade de probabilidade, os parâmetros são ortogonais, uma vez que os valores esperados das derivadas cruzadas da função de verossimilhança são zero (*e.g.* modelos de posição e escala e modelos da família de dispersão). Nesse caso, o segundo, o algoritmo RS, que é uma generalização do algoritmo usado por Rigby e Stasinopoulos (1996a, b) no ajuste da média e da dispersão de modelos aditivos, é mais adequado, pois não usa os valores esperados das derivadas cruzadas. Mais detalhes sobre os algoritmos CG e RS podem ser obtidos em Rigby e Stasinopoulos (2005) e Stasinopoulo*s et al*. (2017).

2.2.7.3 Distribuições

A função densidade de probabilidade da variável resposta Y, $f(\mathrm{y}|\boldsymbol{\theta})$, onde $\boldsymbol{\theta} = (\mu, \sigma, \nu, \tau)$, geralmente, é deliberadamente deixada sem distribuição explícita especificada. A implementação do GAMLSS no programa R por meio da função “gamlss” do pacote “gamlss”, tem como restrição a exigência de que a função $f(\mathrm{y}|\boldsymbol{\theta})$ e sua primeira derivada (e opcionalmente o valor esperado das derivadas de segunda

ordem e as derivadas cruzadas) com relação a cada um dos parâmetros de $\boldsymbol{\theta}$ sejam calculáveis. Derivadas explícitas são preferíveis, mas derivações numéricas podem ser usadas, mesmo que resulte em velocidade computacional reduzida. Existem mais de 80 distribuições diferentes implementadas no software R, dentre elas: distribuições discretas, contínuas e mistas, que permitem a modelagem de distribuições altamente assimétricas e curtóticas (RIGBY e STASINOPOULOS, 2007; STASINOPOULOS *et al*., 2017).

A Tabela 1 apresenta algumas famílias de distribuições discretas implementadas no software R e utilizadas neste trabalho.

Tabela 1 – Distribuições discretas implementadas no pacote gamlss (com função de ligação padrão).

Distribuições	μ	σ	ν	τ
Beta Binomial	logit	log	-	-
Binomial	logit	-	-	-
Geométrica	log	-	-	-
Delaporte	log	log	logit	-
Binomial Negativa Tipo I	log	log	-	-
Binomial Negativa Tipo II	log	log	-	-
Poisson	log	-	-	-
Poisson Inversa Gaussiana	log	log	-	-
Sichel (μ é a média)	log	log	identidade	-
Binomial Ajustada em Zero	logit	logit	-	-
Logarítmica Ajustada em Zero	logit	logit	-	-
Binomial Negativa Ajustada em Zero	log	log	logit	-
Poisson Ajustada em Zero	log	logit	-	-
Beta Binomial Inflacionada de Zero	logit	log	logit	-
Binomial Inflacionada de Zero	logit	logit	-	-
Binomial Negativa Inflacionada de Zero	log	log	logit	-
Poisson Inflacionada de Zero	log	logit	-	-
Poisson Inflacionada de Zero (μ é a média)	log	logit	-	-
Poisson Inversa Gaussiana Inflac. de Zero	log	log	logit	-
Família Binomial Negativa	log	log	identidade	-
Beta Binomial Negativa Inflac. de Zero	log	log	log	logit
Sichel Inflacionada de Zero	log	log	identidade	logit
Beta Binomial Negativa	log	log	log	-
Poisson Dupla	log	log	-	-
Zipf Ajustada de Zero	log	logit	-	-
Geométrica (original)	log	-	-	-

Fonte: Adaptada de Stasinopoulos *et al*. (2017).

2.2.7.4 Seleção de modelos e diagnósticos

A seleção do modelo GAMLSS é realizada a partir de comparações de modelos concorrentes em que são utilizadas diferentes combinações dos componentes $\mathcal{M}$ = $\{\mathcal{D}, \mathcal{G}, \mathcal{T}, \boldsymbol{\lambda}\}$, em que $\mathcal{M}$ representa um modelo GAMLSS, $\mathcal{D}$ especifica a distribuição da variável resposta, $\mathcal{G}$ o conjunto das funções de ligação $(g_1, \dots, g_p)$ para os parâmetros $(\theta_1, \dots, \theta_p)$, $\mathcal{T}$ define o conjunto de termos preditores $(t_1, \dots, t_p)$ para os preditores $(\eta_1, \dots, \eta_p)$ e λ explicita o conjunto de hiperparâmetros.

Na estrutura de regressão GAMLSS paramétrica, cada modelo aninhado $\mathcal{M}$ pode ser avaliado a partir de seu desvio global ajustado (GD), dado por GD $= -2l(\hat{\theta})$ onde $l(\hat{\theta}) = \sum_{i=1}^{n}(\hat{\theta}^i)$. Ao comparar modelos GAMLSS não-aninhados, o critério de informação Akaike generalizado (GAIC) (AKAIKE, 1983) pode ser usado para penalizar sobreajustes. Isso é obtido adicionando aos desvios globais ajustados uma penalidade fixa k para cada grau de liberdade efetivo que é usado no modelo, ou seja, GAIC(k) = GD $+ k \cdot df$, onde df é o total de graus de liberdade efetivos utilizados no modelo e GD é o desvio global ajustado. O modelo com o menor valor do critério GAIC(k) é o selecionado. O critério de informação de Akaike (AIC) (AKAIKE, 1974) e o Critério de Informação Bayesiano (BIC) (SCHWARZ, 1978) são casos especiais do critério GAIC(k), e correspondem a k=2 e $k = \log(n)$, respectivamente. Acrescenta-se que os dois critérios possuem fundamentação assintótica (RIGBY e STASINOPOULOS, 2005).

Para avaliar a adequabilidade de cada modelo ajustado $\mathcal{M}$ e, em particular, a distribuição do componente $\mathcal{D}$, utiliza-se os resíduos (dos quantis aleatórios normalizados) de Dunn e Smyth (1996), que é uma versão aleatorizada do resíduo de Cox e Snell (1968). Estes resíduos são dados por $\hat{r}_i = \Phi^{-1}(u_i)$, em que Φ^{-1} é a inversa da função de distribuição acumulada (CDF) de uma normal padrão e $u_i = F(y_i|\hat{\theta}^i)$ se y_i é uma observação de uma resposta contínua. Considera-se ainda u_i um valor aleatório de uma distribuição uniforme no intervalo $\left[F(y_i - 1|\hat{\theta}^i), F(y_i|\hat{\theta}^i)\right]$ se y_i é uma observação de uma resposta inteira discreta, em que $F(y_i|\theta)$ é a função de distribuição de $\mathcal{D}$. Quando o modelo está corretamente especificado os verdadeiros resíduos r_i seguem distribuição normal padrão (RIGBY e STASINOPOULOS, 2005).

2.2.8 Testes estatísticos não-paramétricos

A designação de “teste não paramétrico” deve-se ao fato de não ser necessário fazer suposições quanto a forma específica da distribuição de probabilidade da população da qual a amostra foi extraída. Por não dependerem desse conhecimento são denominados de *testes de distribuição livre*. Os métodos não-paramétricos se aplicam a uma grande variedade de situações, porque não exigem populações distribuídas normalmente. Quando as exigências sobre as distribuições populacionais são satisfeitas, os testes não paramétricos são, em geral, menos eficientes do que seus correspondentes paramétricos, mas a redução da eficiência pode ser compensada por um tamanho maior da amostra (SIEGEL, 2006; TRIOLA, 2014).

2.2.8.1 Teste de Kolmogorov-Smirnov

O teste de Kolmogorov-Smirnov de uma amostra é um teste de aderência, concernente com o grau de concordância entre a distribuição de um conjunto de valores da amostra (escores observados) e alguma distribuição teórica especificada. Tem como finalidade determinar se os escores em uma amostra podem ser pensados razoavelmente como tendo vindo de uma população tendo essa distribuição teórica. O método pode ser descrito da seguinte forma:

Seja $F_0(X)$ uma função completamente especificada de distribuição de frequências relativas acumuladas – a distribuição teórica sob H_0. Isto é, para qualquer valor de X, o valor de $F_0(X)$ é a proporção de casos esperados com escores iguais ou menores que X.

Seja $S_N(X)$ a distribuição de frequências relativas acumuladas de uma amostra aleatória de N observações. Se X_i é um escore qualquer possível, então $S_N(X_i) = F_i/N$, onde F_i é o número de observações menores ou iguais a X_i. $F_0(X_i)$ é a proporção esperada de observações menores ou iguais a X_i.

Agora, sob a hipótese nula de que a amostra tenha sido extraída de uma distribuição teórica especificada, é esperado que para qualquer valor de X_i, $S_N(X_i)$ esteja bastante próximo de $F_0(X_i)$. Isto é, quando H_0 é verdadeira esperaríamos que as diferenças entre $S_N(X_i)$ e $F_0(X_i)$ fossem pequenas e dentro dos limites de erros

aleatórios. O teste de Kolmogorov-Smirnov focaliza sobre o maior dos desvios. O maior valor absoluto de $F_0(X_i) - S_N(X_i)$ é chamado de *desvio máximo* D:

$$D = max|F_0(X_i) - S_N(X_i)|, \qquad i = 1,2,..,N \tag{40}$$

A distribuição amostral de D sob H_0 é conhecida. A região de rejeição consiste de todos os valores de D, que sejam tão grandes que a probabilidade associada com suas ocorrências quando H_0 é verdadeira seja menor ou igual a $\propto$=0,05 (SIEGEL, 2006).

2.2.8.2 Teste de Kruskal-Wallis

Quando são violadas as pressuposições de normalidade e homocedasticidade, os resultados de uma análise de variância tradicional (ANOVA) não são confiáveis, pois a probabilidade de se cometer um erro do Tipo I afasta-se marcadamente de $\propto$. A alternativa não-paramétrica para a ANOVA a um critério de classificação é o teste de Kruskal-Wallis, que é uma generalização do teste de Wilcoxon-Mann-Whitney. A análise de variância de um fator de Kruskal-Wallis por postos é um teste extremamente útil para decidir se k amostras independentes provêm de populações diferentes. Tem como finalidade verificar se as diferenças entre as amostras significam genuínas diferenças entre as populações ou se elas representam meramente o tipo de variações que seriam esperadas entre amostras aleatórias de uma mesma população (SIEGEL, 2006; TRIOLA, 2013).

A técnica de Kruskal-Wallis testa a hipótese nula de que as k amostras provém da mesma população ou de populações idênticas com a mesma mediana. Dessa forma, considerando θ_j a mediana para o j-ésimo grupo ou amostra, tem-se que a hipótese nula de que as medianas são as mesmas, pode ser escrita como: $H_0: \theta_1 = \theta_2 = \ ... = \theta_k$; e a hipótese alternativa como: $H_a: \theta_i \neq \theta_j$ para alguns grupos i e j.

No cálculo do teste de Kruskal-Wallis, cada uma das N observações é substituída por postos, no qual todos os escores de todas as k amostras são organizados através de postos em uma única série. O menor escore é substituído pelo posto 1, o seguinte menor escore é substituído pelo posto 2 e o maior escore é

substituído pelo posto N, onde N é o número total de observações independentes nas amostras.

Em seguida, a soma dos postos em cada amostra é encontrada. Destas somas, pode-se calcular o posto médio para cada amostra ou grupo. Assim sendo, se as amostras são da mesma população ou de populações idênticas, os postos médios devem ser quase os mesmos, enquanto que se as amostras viessem de populações com medianas diferentes, os postos médios deveriam diferir. O teste de Kruskal-Wallis trabalha com as diferenças entre os postos médios para determinar se elas são tão discrepantes que provavelmente não tenham vindo de amostras que tenham sido extraídas de uma mesma população. A estatística do teste é dada por:

$$KW = \left[\frac{12}{N(N+1)}\sum_{j=1}^{k} n_j \bar{R}_j^2\right] - 3(N+1) \tag{41}$$

em que, k é o número de amostras ou grupos, n_j é o número de casos na j-ésima amostra, N é o número de casos na amostra combinada (a soma dos $n_j's$), $\bar{R}_j$ é a média dos postos na j-ésima amostra ou grupo. Para amostras grandes, rejeita-se H_0, a um nível de significância α, se $KW > \chi^2$ com $k-1$ graus de liberdade.

Quando ocorrem empates entre dois ou mais escores (independentemente do grupo) a cada um deles é atribuída a média dos postos empatados.

Como a variância da distribuição amostral de KW é influenciada por empates, sugere-se uma correção na estatística dada na equação (41). Então uma expressão geral para KW corrigida para empates é:

$$KW = \frac{\left[\frac{12}{N(N+1)}\sum_{j=1}^{k} n_j \bar{R}_j^2\right] - 3(N+1)}{1 - \frac{\sum_{i=1}^{g} t_i^3 - t_i}{N^3 - N}} \tag{42}$$

em que, g é o número de agrupamentos de postos diferentes empatados, t_i é o número de postos emprestados no i-ésimo agrupamento e N é o número de observações através de todas as amostras (SIEGEL, 2006).

2.3 Dados e Metodologia

2.3.1 Dados

Foram analisadas informações sobre as duas espécies de lagostas predominantes nas capturas no nordeste brasileiro, *Panulirus argus* (lagosta vermelha) e *Panulirus laevicauda* (lagosta verde). Os dados foram recompilados da série de artigos publicados na década de 1960 (Paiva e Costa, 1963, 1964, 1965, 1966; 1967, 1968, 1969, 1970), os quais fazem parte dos documentos científicos que subsidiam as decisões sobre o defeso desde então. As informações publicadas nestes artigos foram obtidas a partir de amostras coletadas em diferentes pontos de desembarque de pescado em Fortaleza-CE, com destaque para a praia de Mucuripe. Para cada amostra foi tomado um total de 50 indivíduos, retirados ao acaso, dos desembarques de lagostas na praia. Apenas, em duas amostras do ano de 1962, não foi possível a retirada de 50 indivíduos. A Tabela 2 apresenta a quantidade de dias de pescaria, o período e o total de indivíduos amostrados.

Tabela 2 – Quantidade de dias de pesca, período de realização da pesca e número de indivíduos amostrados das espécies *Panulirus argus* e *Panulirus laevicauda* durante os anos de 1962 a 1969.

Ano	Dias	Período	Indivíduos amostrados		
			Panulirus argus	*Panulirus laevicauda*	Total
1962	202	13/02 a 31/12	6.888	3.182	10.070
1963	181	02/01 a 26/12	6.509	2.541	9.050
1964	166	08/01 a 30/12	4.814	3.486	8.300
1965	181	04/01 a 29/12	4.566	4.484	9.050
1966	174	04/01 a 29/12	4.454	4.246	8.700
1967	183	02/01 a 29/12	5.449	3.701	9.150
1968	213	02/01 a 23/12	2.590	8.060	10.650
1969	178	04/01 a 30/12	6.008	2.892	8.900

Fonte: Adaptada de Paiva e Costa (1963, 1964, 1965, 1966; 1967, 1968, 1969, 1970).

As capturas monitoradas foram efetuadas com botes lagosteiros motorizados que operavam com manzuás (tipo de armadilha), isca de peixes marinhos, peixes de água doce ou mocotó de boi. Os exemplares de lagostas capturados foram

identificados com chave sistemática fundamentada nos estudos de Faria e Silva (1937), Smith (1948, 1958), Chace Jr. e Dumont (1949) e Sheard (1949). Para cada indivíduo foi anotado o sexo e o comprimento total, medido como a distância entre a margem anterior do entalhe formado pelos espinhos rostrais e a extremidade posterior do telso, estando o animal completamente estendido sobre uma superfície plana. Também foi registrado se havia presença de características associadas às diferentes fases de reprodução, que se inicia no acasalamento quando há pareamento ventral e deposição de massa espermatofórica (espermoteca) dos machos na parte ventral do abdômen das fêmeas, com posterior fecundação dos ovos. A presença de espermoteca íntegra no externo da fêmea indica que ocorreu acasalamento recente. O desenvolvimento embrionário ocorre nos ovos que ficam aderidos aos pleópodos das fêmeas. A presença de massa de ovos é um indicativo dessa fase embrionária. Após a desova, permanecem resquícios da espermoteca, e há a ausência de ovos (PAIVA e COSTA, 1968; IGARASHI, 2010). Assim a identificação da presença de espermoteca íntegra, de ovos e, finalmente, de resquício de espermoteca na fêmea são características que permitem identificar se o exemplar esteve em processo reprodutivo ou em que fase do ciclo ele está. Os indicadores para os machos não são tão aparentes externamente. Portanto, somente dados de fêmeas são usados para a análise da variação temporal da reprodução. Além da especificação do sexo, comprimento total e da fase reprodutiva, os artigos publicados na década de 1960 também contêm informações de caráter temporal, como o mês e ano em que os exemplares de lagostas foram capturados.

2.3.2 Modelos de regressão e variáveis utilizadas

Nas análises foram utilizados modelos lineares generalizados e modelos aditivos generalizados para posição, escala e forma, os quais assumem uma variável resposta em função de uma ou mais variáveis explicativas em uma estrutura aditiva. Segue abaixo descrições das variáveis consideradas.

2.3.2.1 Variável resposta

A variável resposta utilizada foi o número de lagostas fêmeas capturadas em atividade reprodutiva (com ovos ou espermoteca íntegra, ou resto de espermoteca),

adicionando o logaritmo do número de lagostas capturadas como *offset*. Para a modelagem da variável resposta foram consideradas as distribuições Poisson, Binomial Negativa, Geométrica, Delaporte, Binomial Negativa Tipo I, Binomial Negativa Tipo II, Poisson Inversa Gaussiana, Sichel (μ é a média), Logarítmica Ajustada em Zero, Binomial Negativa Ajustada em Zero, Poisson Ajustada em Zero, Binomial Negativa Inflacionada de Zero, Poisson Inflacionada de Zero, Poisson Inflacionada de Zero (μ é a média), Poisson Inversa Gaussiana Inflacionada de Zero, Família Binomial Negativa, Beta Binomial Negativa Inflacionada de Zero, Sichel Inflacionada de Zero, Beta Binomial Negativa, Poisson Dupla, Zipf Ajustada de Zero e Geométrica - original).

A variável resposta também foi tratata como uma variável binária. O “sucesso” é atribuído à ocorrência de lagosta fêmea em atividade reprodutiva e o “fracasso” é atribuído à ocorrência de fêmea que não está em atividade reprodutiva. Para este caso, foram ajustados modelos com distribuições Binomial, Beta Binomial, Binomial Ajustada em Zero, Beta Binomial Inflacionada de Zero e Binomial Inflacionada de Zero.

2.3.2.2 Variáveis explicativas

- Comprimento da lagosta (PM) – esta variável se refere ao comprimento total da lagosta no momento da captura. Como o registro foi feito em intervalo de classe, considerou-se o ponto médio.
- Mês (MES) – refere-se ao mês em que o exemplar foi capturado;
- Ano (ANO) – refere-se ao ano em que o exemplar foi capturado.

2.3.2.3 Seleção dos modelos lineares generalizados e diagnósticos

A ordem de entrada das variáveis explicativas nos modelos investigados, foi feita ajustando-se um modelo para cada uma das variáveis separadamente, e as deviâncias foram usadas para classificação (NELDER e WEDDERBURN, 1972). A variável explicativa que resultou na menor deviância foi então selecionada para o primeiro posto na estrutura do modelo. 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 menor deviância. O processo

continuou até a obtenção do modelo base com todas as variáveis explicativas (efeitos principais). Por fim, além dos fatores principais, foram incluídas interações de primeira ordem entre as variáveis. A partir de então procedeu-se com uma abordagem "*backward*" para a seleção de variáveis explicativas e interações nos modelos finais, com base no AIC (AKAIKE, 1974), e no BIC (SCHWARZ, 1978). Para examinar se os modelos diferem do melhor modelo por um parâmetro e têm essencialmente os mesmos valores da log-verossimilhança maximizada foi utilizado o Delta AIC (BURNHAM e ANDERSON, 2002). Todas as variáveis explicativas foram testadas como fator e covariável nos modelos investigados. Para se ter uma medida do poder explicativo dos modelos, foi calculado o coeficiente de determinação Pseudo-R^2.

Para avaliar a qualidade do ajuste do modelo selecionado e influências dos registros individuais, foram utilizados gráficos de diagnóstico de resíduos e cálculos baseados na diagonal da matriz de projeção (*e.g.*, distâncias de Cook, alavancagem) (*e.g.* VENABLES e RIPLEY, 2002). Os testes de Kolmogorov-Smirnov e Breusch-Pagan (BREUSCH e PAGAN, 1979) também foram empregados para verificar se os resíduos apresentaram distribuição normal e comportamento homocedástico. Todas as análises foram realizadas com o programa R (R 3.4.1 Core Team, 2017), com o auxílio da função "glm" do pacote "stats".

2.3.2.4 Seleção dos modelos aditivos generalizados para posição, escala e forma, e diagnósticos

As variáveis explicativas utilizadas para modelar o parâmetro $\boldsymbol{\mu}$ nos GLMs selecionados (ver seção anterior), foram mantidas para o ajuste inicial dos modelos com as diversas distribuições disponíveis no pacote gamlss. Não foram adicionadas variáveis no preditor linear para modelar os parâmetros $\boldsymbol{\sigma}$, $\boldsymbol{\nu}$ e $\boldsymbol{\tau}$, sendo os modelos iniciais compostos apenas pelo intercepto.

A comparação entre os ajustes dos modelos investigados com as diversas distribuições foram realizadas utilizando-se o Critério de informação de Akaike Generalizado (GAIC), em que a penalidade adotada foi $p = 2$ e $p = \log(n)$, caracterizando o critério de informação de Akaike (AIC) e o Critério de Informação Bayesiano (BIC). Entre os modelos candidatos aquele que apresentou menor valor do AIC ou BIC foi selecionado.

Ainda na tentativa de melhorar o ajuste dos modelos selecionados foi utilizado a estratégia A, implementada na função stepGAICAll.A(), destacada por Stasinopoulo*s et al*. (2017). Essa estratégia tem como finalidade selecionar termos aditivos usando um GAIC para todos os parâmetros da distribuição, assumindo uma distribuição específica para variável resposta. A Estratégia A pode ser dividida em sete etapas: (1) Utiliza-se um procedimento de seleção GAIC "*forward"* para selecionar um modelo apropriado para $\boldsymbol{\mu}$, com $\boldsymbol{\sigma}$, $\boldsymbol{\nu}$ e $\boldsymbol{\tau}$ configurado como constantes; (2) Dado o modelo para $\boldsymbol{\mu}$ obtido em (1) e para $\boldsymbol{\nu}$ e $\boldsymbol{\tau}$ configurado como constantes, usa-se um procedimento de seleção "*forward"* para selecionar um modelo apropriado para $\boldsymbol{\sigma}$; (3) Dado os modelos para $\boldsymbol{\mu}$ e $\boldsymbol{\sigma}$ obtidos em (1) e (2), respectivamente, e com $\boldsymbol{\tau}$ configurado como constante, usa-se um procedimento de seleção "*forward"* para selecionar um modelo apropriado para $\boldsymbol{\nu}$; (4) Dado os modelos para $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$ e $\boldsymbol{\nu}$ obtidos em (1), (2) e (3), respectivamente, usa-se um procedimento de seleção "*forward"* para selecionar um modelo apropriado para $\boldsymbol{\tau}$; (5) Dado os modelos para $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$ e $\boldsymbol{\tau}$ obtidos em (1), (2) e (4), respectivamente, usa-se um procedimento de seleção "*backward"* para selecionar um modelo apropriado para $\boldsymbol{\nu}$; (6) Dado os modelos para $\boldsymbol{\mu}$, $\boldsymbol{\nu}$ e $\boldsymbol{\tau}$ obtidos em (1), (5) e (4), respectivamente, usa-se um procedimento de seleção "*backward"* para selecionar modelo apropriado para $\boldsymbol{\sigma}$; (7) Dado os modelos para $\boldsymbol{\sigma}$, $\boldsymbol{\nu}$ e $\boldsymbol{\tau}$ obtidos em (6), (5) e (4), respectivamente, usa-se um procedimento de seleção "*backward"* para selecionar um modelo apropriado para $\boldsymbol{\mu}$ e o processo é finalizado. O modelo final conterá um diferente subconjunto de termos (não necessariamente os mesmos termos) para cada $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$, $\boldsymbol{\nu}$ e $\boldsymbol{\tau}$.

Para a análise dos resíduos do GAMLSS foi utilizada a função "plot()". Esta função produz quatro gráficos de resíduos quantílicos normalizados (resíduos contra valores ajustados; resíduos contra variável explanatória; função densidade dos resíduos; qq-plot dos resíduos e um sumário de medidas resumo da distribuição dos resíduos, em que informa a média, variância, coeficiente de assimetria, coeficiente de curtose e correlação de Filliben, permitindo verificar o pressuposto da normalidade dos resíduos. O teste de Filliben (1975) verifica a normalidade de uma variável por meio do cálculo do coeficiente de correlação, utilizando a mediana da distribuição normal padronizada. Se os valores tabelados, para um determinado tamanho de amostra for maior do que a estatística do teste, a hipótese de que os dados seguem uma distribuição normal é rejeitada para um nível de probabilidade prefixado.

Além dos gráficos de resíduos e das medidas resumo, também foram utilizados o teste de Kolmogorov-Smirnov para verificar a normalidade dos resíduos e o teste de Breusch-Pagan (BREUSCH e PAGAN, 1979) para detectar comportamento heterocedástico. Outra ferramenta de diagnóstico que foi utilizada para visualizar o quão bem o modelo estatístico selecionado se ajustou aos dados foi o gráfico denominado de *Worm Plo*t (WP) de VAN BUUREN e FREDRIKS (2001).

2.4 Resultados e Discussões

2.4.1 Análise Exploratória

No período de 1962 a 1969 o número de lagostas vermelhas com características indicativas de atividade reprodutiva (com ovos ou espermoteca ou resto de espermoteca) atingiu um percentual médio anual de 30,3% (Tabela 3). Já para as lagostas verdes a média anual em atividade reprodutiva é de 65,5%. Entre as características que indicam reprodução, ovos, espermoteca íntegra ou resquícios de espermoteca, esta última é a mais frequente nos abdomens das fêmeas de ambas as espécies.

O comprimento (PM) das lagostas vermelhas capturadas variou de 12,55 cm a 36,55 cm, enquanto que para as lagostas verdes o comprimento variou de 12,55 cm a 28,55 cm. Portanto, os maiores comprimentos encontrados no momento da captura ocorreram para as lagostas vermelhas.

Os coeficientes de assimetria positivos, assim como a média maior que a mediana, indicam que as distribuições de frequência das variáveis proporção de fêmeas capturadas com ovos, com espermetoca, com resto de espermoteca, proporção de fêmeas em atividade reprodutiva (para a espécie *P. argus*), proporção de fêmeas e comprimento da lagosta vermelha são assimétricas posititivas (ou distribuição com cauda à direita). Já o coeficiente de curtose informa que as distribuições das variáveis proporção de fêmeas com ovos, com espermoteca e proporção de fêmeas são leptocúrticas, enquanto as variáveis proporção de fêmeas com resto de espermoteca, proporção de fêmeas em atividade reprodutiva e comprimento da lagosta têm distribuição platicúrtica. De acordo com o teste de Kolmogorov-Smirnov, há evidências estatísticas para rejeitar a hipótese de que as

variáveis apresentam distribuição normal, considerando um nível de significância de 5%.

Tabela 3 – Estatísticas descritivas para as espécies *Panulirus argus* e *Panulirus laevicauda* capturadas entre 1962 e 1969. PO – proporção de lagostas capturadas com ovos; PE – proporção de lagostas capturadas com espermoteca; PRE – proporção de lagostas capturadas com resto de espermoteca; PREP – proporção de lagostas capturadas com ovos ou espermoteca ou resto de espermoteca; PM – comprimento das lagostas capturadas; PF – proporção de fêmeas capturadas em relação à soma de fêmeas e machos; CV – coeficiente de variação em porcentagem; p-valor – nível descritivo do teste de Kolmogorov-Smirnov para verificar normalidade das variáveis.

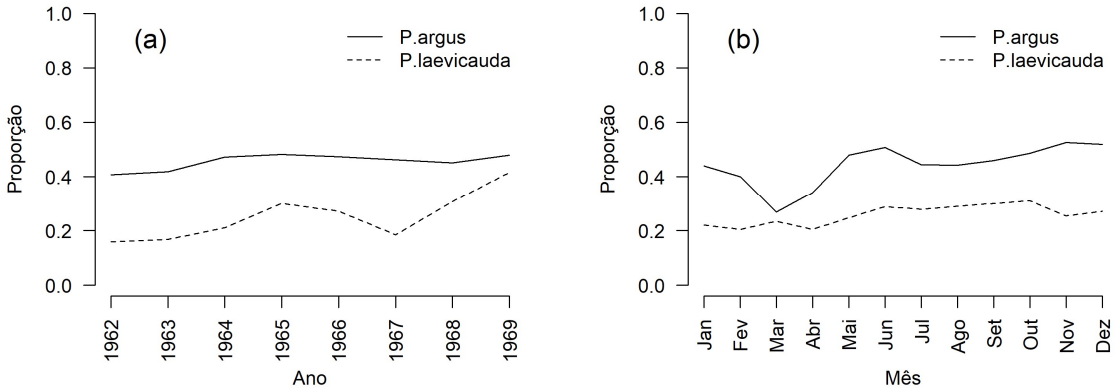
Panulirus argus (lagosta vermelha)						
Estatísticas / Variáveis	PO	PE	PRE	PREP	PM	PF
Média	0,032	0,041	0,231	0,303	22,630	0,456
Desvio padrão	0,030	0,039	0,177	0,210	4,107	0,028
Mínimo	0,002	0,001	0,017	0,030	12,550	0,406
Máximo	0,079	0,103	0,526	0,598	36,550	0,483
Mediana	0,022	0,029	0,194	0,302	22,550	0,467
CV (%)	96,090	95,871	76,725	69,440	18,150	6,283
Assimetria	4,576	3,364	1,027	0,068	0,201	0,621
Curtose	27,881	14,628	2,639	1,351	2,526	3,446
p-valor	< 2,2e-16*	< 2.2e-16*	< 2,2e-16*	< 2,2e-16*	5,5e-09*	< 2,2e-16*
Panulirus laevicauda (lagosta verde)						
Estatísticas / Variáveis	PO	PE	PRE	PREP	PM	PF
Média	0,187	0,108	0,36	0,655	19,42	0,252
Desvio padrão	0,138	0,104	0,135	0,099	2,701	0,089
Mínimo	0,021	0,002	0,218	0,529	12,550	0,159
Máximo	0,438	0,264	0,582	0,799	28,550	0,416
Mediana	0,147	0,067	0,308	0,676	19,550	0,241
CV (%)	73,621	95,986	37,638	15,083	13,909	35,126
Assimetria	1,596	2,658	0,669	-0,587	0,117	1,456
Curtose	4,523	10,019	2,172	1,868	2,644	4,763
p-valor	< 2,2e-16*	< 2,2e-16*	< 2,2e-16*	< 2,2e-16*	6,9e-07*	< 2,2e-16*

Fonte: Elaborada pelo autor.
Nota: Nível de significância: p<0,05 (*).

Apenas 25,2% das lagostas verdes capturadas são fêmeas, enquanto que para as lagostas vermelhas o percentual de fêmeas capturadas atingiu 45,6% (Figura 3). Há, portanto, uma predominância de machos nas amostras para ambas as espécies, ainda que menos acentuada no caso da lagosta vermelha. Na Figura 3 estão registradas as proporções de fêmeas ao longo dos anos e meses. Para a espécie *P. argus,* a proporção de fêmeas capturadas não apresenta grande variabilidade

interanual (Figura 3(a)). No entanto, a proporção de fêmeas aumentou ao longo dos anos no caso da lagosta verde, com destaque para os anos 1965 e 1969.

Figura 3 – Variações anuais (a) e mensais (b) das proporções de fêmeas capturadas para as espécies *Panulirus argus* e *Panulirus laevicauda*.



Fonte: Elaborada pelo autor.

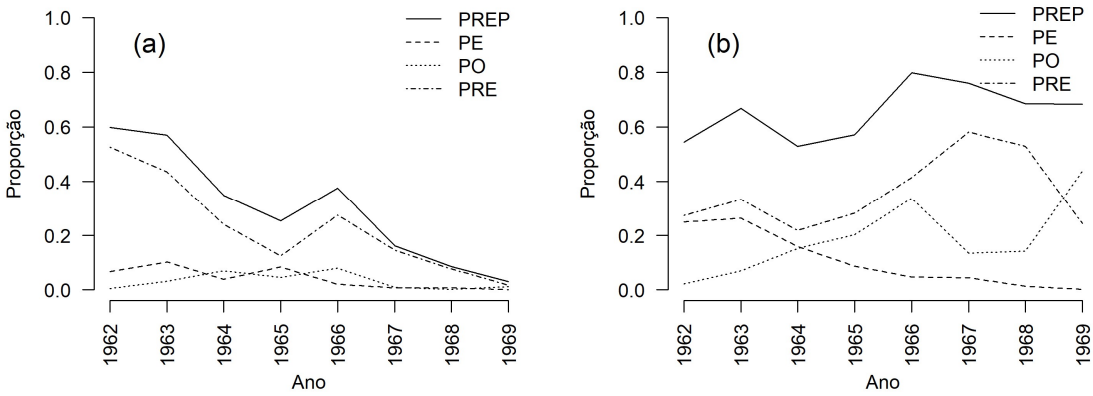
As fêmeas da espécie *P. argus* são relativamente mais frequentes a partir de maio (Figura 3(b)). Nos meses de fevereiro, março e abril as proporções de fêmeas capturadas foram menores para a espécie. Entretanto, para a espécie *P. laevicauda,* ocorreu pequena variabilidade mensal nas proporções de fêmeas. Paiva e Costa (1964) mencionam que a razão sexual varia amplamente para *P. argus*, ora predominando machos e ora fêmeas, dependendo do período do ano. Os autores executaram um teste de homogeneidade e mostraram que as diferenças constatadas entre machos e fêmeas eram significativas. A explicação para as grandes alterações mensais nas proporções sexuais está no comportamento diferente dos sexos da espécie *P. argus* durante o período de reprodução. Quanto maior a intensidade do processo reprodutivo na população, maiores as frequências relativas de machos nas capturas, pois as fêmeas desde o acasalamento até a liberação das larvas apresentam reduzida atividade trófica e permanecem preferencialmente entocadas. Passado o período de reprodução, as fêmeas que estiveram em regime de baixa ingestão de alimentos, procuram suprir a demanda energética ativamente, e acabam por ser atraídas pelas iscas dos aparelhos de pesca, aumentando a sua participação relativa nas capturas. No caso da *P. laevicauda*, o processo é diferente, pois um

número regular de fêmeas da espécie se encontram em reprodução ao longo de todo o ano. Portanto, a proporção das fêmeas varia menos mensalmente e é menor se comparada com a lagosta vermelha. Assim, de maneira geral, se considerado todo o período anual, as fêmeas da espécie *P. argus* são mais suscetíveis à ação de aparelhos de pesca, do que as da espécie *P. laevicauda* (PAIVA e COSTA, 1963).

Estudos desenvolvidos por Lima e Andrade (2017) na costa de Pernambuco confirmam a dominância da lagosta vermelha, com valores compreendidos entre 60% e 80% dos totais capturados entre 1999 e 2006. Enfatizam ainda que as capturas de lagostas são menores durante o período de defeso e atingem valores mais elevados logo nos meses iniciais da temporada de pesca, fato que pode ser consequência do recrutamento de novos indivíduos para a pesca, com comprimento suficiente para serem selecionados nos aparelhos de pesca, ou ao aumento da disponibilidade de fêmeas adultas para a pesca, as quais estariam menos ativas, abrigadas ou em áreas fora de alcance da frota comercial no período de defeso.

Na Figura 4 estão registradas as variações anuais das proporções de fêmeas capturadas em fase reprodutiva para as espécies *P. argus* e *P. laevicauda*.

Figura 4 – Variação anual das proporções de fêmeas capturadas com ovos (PO), com espermoteca (PE), com resto de espermoteca (PRE), com ovos, espermoteca ou resto de espermoteca (PREP) para as espécies *Panulirus argus* (a) e *Panulirus laevicauda* (b).



Fonte: Elaborada pelo autor.

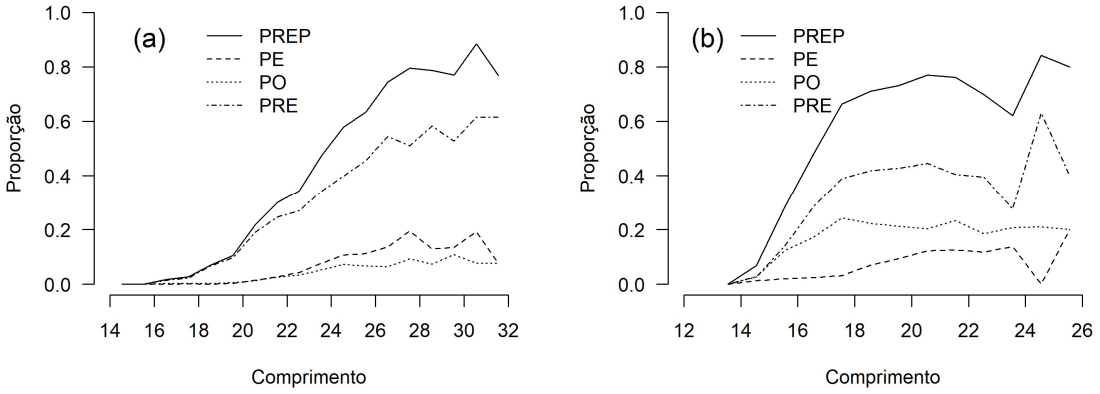
A captura de lagostas vermelhas em atividade reprodutiva (PREP) diminuiu no decorrer do período analisado (Figura 4(a)). Já as lagostas verdes (Figura 4(b)) destacam-se com uma maior proporção de exemplares capturados com indícios de

atividade reprodutiva ao longo de todos os anos. Para ambas as espécies, a fase reprodutiva caracterizada pela presença de resto de espermoteca no externo da fêmea foi predominante frente às demais fases. No entanto, para a lagosta vermelha (Figura 4(a)) verifica-se que a proporção de indivíduos capturados com esta característica no decorrer do período analisado diminuiu. A proporção de lagostas vermelhas capturadas com ovos ou espermoteca oscila em torno de 0,1 até o ano de 1966 e se aproxima de zero nos anos posteriores. Para as lagostas verdes, a proporção de fêmeas capturadas com ovos mostrou uma tendência crescente ao longo dos anos, enquanto a proporção de lagostas verdes capturadas com espermoteca apresentou uma tendência decrescente.

A queda na proporção de lagostas vermelhas capturadas em atividade reprodutiva pode ter sido ocasionada pela Portaria Nº 70, de 12 de abril de 1961, da Diretoria de Caça e Pesca do Departamento Nacional de Produção Animal – DNPA, do Ministério da Agricultura, que estabeleceu o primeiro defeso na plataforma do Ceará a Alagoas com duração de três meses, a partir de 1962 (IGARASHI, 2007; CAVALCANTE *et al*., 2011). Posteriormente, com a criação das primeiras medidas de proteção ao estoque pesqueiro estabelecido pela SUDEPE entre os anos de 1967 e 1969, nota-se que a percentagem de lagostas vermelhas capturadas em reprodução não ultrapassa os 20% e tende a diminuir nos sucessivos anos estudados. Para as lagostas verdes, a redução na proporção de exemplares em reprodução ao longo do período analisado pode não ter ocorrido pelo fato desta espécie se reproduzir durante todo o ano e o impacto do defeso seria diluído (MATSUURA, 1987; SOARES e FONTELES FILHO, 2000).

Independente da espécie capturada, há relação entre a probabilidade de que o exemplar capturado esteja em atividade reprodutiva e o comprimento do mesmo (Figura 5). Os menores exemplares de lagostas vermelhas capturadas em atividade reprodutiva aparecem com aproximadamente 17 cm de comprimento. Já para as lagostas verdes o menor comprimento encontrado em atividade reprodutiva tinha aproximadamente 15 cm. Esse resultado é concordante com o trabalho de Igarashi (2007) que mostra que as atividades reprodutivas da *P. laevicauda* e *P. argus* podem ter início quando elas atingem um comprimento total superior a 15 e 16 cm, respectivamente.

Figura 5 – Relação entre as proporções de lagostas capturadas com ovos (PO), com espermoteca (PE), com resto de espermoteca (PRE), com ovos, espermoteca ou resto de espermoteca (PREP) e os comprimentos dos exemplares de *Panulirus argus* (a) e *Panulirus laevicauda* (b) capturados.

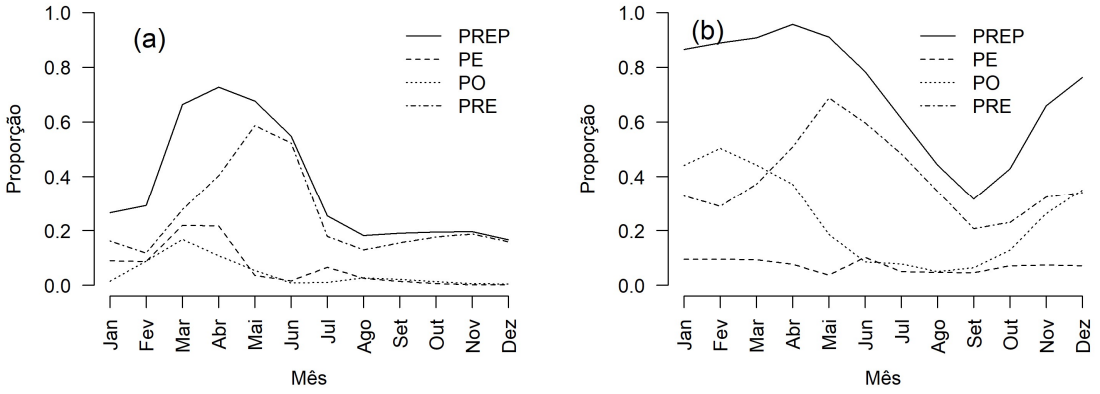


Fonte: Elaborada pelo autor.

De forma geral, observa-se que para comprimentos entre 21 cm e 31 cm há proporção maior de exemplares em atividade reprodutiva entre os capturados. Para as lagostas verdes, a maior probabilidade de encontrar exemplares em fase reprodutiva ocorre entre os comprimentos de 16 cm a 26 cm (Figura 5). De acordo com estudos desenvolvidos por Lima *et al.* (2014), a produção de ovos (fase reprodutiva) está diretamente relacionada com o seu tamanho, lagostas que são maiores em comprimento tem uma maior probabilidade de produzir ovos.

O período de maior atividade reprodutiva para as fêmeas de *P. argus* ocorre no primeiro semestre (Figura 6(a)).

Figura 6 – Proporção de lagostas capturadas com ovos (PO), com espermoteca (PE), com resto de espermoteca (PRE), com ovos, espermoteca ou resto de espermoteca (PREP) ao longo dos meses para as espécies *Panulirus argus* (a) e *Panulirus laevicauda* (b).



Fonte: Elaborada pelo autor.

As lagostas vermelhas capturadas com ovos têm no geral baixa frequência relativa mensal. No entanto, durante o primeiro semestre, nos meses de fevereiro, março e abril, a proporção de fêmeas capturadas com ovos ganha destaque quando comparado com os demais meses. Cabe traçar o paralelo com a informação de que também houve queda na proporção de fêmeas amostradas entre fevereiro e abril (Figura 3(b)), meses em que a proporção de fêmeas com ovos é maior (Figura 6(a)). A proporção de lagostas capturadas com espermoteca no abdômen é relativamente mais frequente nos meses de março e abril (Figura 6(a)). Já as fêmeas capturadas com resto de espermoteca são relativamente mais frequentes durante os meses de fevereiro a abril, porém mesmo após meados do mês junho é possível encontrar lagostas nesta fase reprodutiva (Figura 6(a)).

Para a espécie *P. laevicauda,* uma proporção regular de fêmeas se encontram em reprodução ao longo de todos os meses (Figura 6(b)). No entanto, o período de maior atividade reprodutiva para esta espécie ocorre no primeiro semestre. Um segundo e menos intenso período de reprodução inicia-se em outubro. As fêmeas capturadas com ovos são relativamente mais frequentes durante os meses de janeiro a abril e outubro a dezembro. A proporção de fêmeas capturadas com espermoteca no abdômen é relativamente estável ao longo dos meses. Para a fase do ciclo reprodutivo caracterizada por resto de espermoteca no abdômen, nota-se que as maiores frequências de fêmeas com essa característica ocorreram nos meses de abril a julho.

2.4.2 Modelos lineares generalizados (GLM)

A estrutura dos quinze melhores modelos ajustados com distribuições binomial, poisson e binomial negativa (cinco para cada distribuição) para elucidar os padrões nas variações das proporções ou quantidades de lagostas em reprodução são apresentados na Tabela 4. A seleção desses modelos foi baseado em AIC, BIC e no diagnóstico dos resíduos, conforme descrito na sessão 2.3.2.3.

Tabela 4 – Estrutura dos cinco melhores modelos ajustados com distribuição Binomial, Poisson e Binomial Negativa. REP – variável resposta, representando o número (Poisson e Binomial Negativa) ou proporção (Binomial) de lagostas capturadas com ovos ou espermoteca ou resto de espermoteca; ANO – ano em que as lagostas foram capturadas; MES – mês que as lagostas foram capturadas; PM – comprimento das lagostas capturadas. Interações particulares são representadas por dois pontos. A letra "F" na frente da variável indica que a variável foi tratada como fator no modelo, caso contrário a variável foi tratada como covariável. Interações particulares são representadas por dois pontos.

	Modelos	Distribuição
	Panulirus argus	
1	REP = ANO+F(MES)+PM+ANO:F(MES)+ANO:PM+F(MES):PM	Binomial
2	REP = ANO+F(MES)+PM+ ANO:F(MES)+ANO:PM	Binomial
3	REP = F(ANO)+F(MES)+PM+F(ANO):PM+F(MES):PM	Binomial
4	REP = ANO+F(MES)+PM+ANO:F(MES)+F(MES):PM	Binomial
5	REP = F(ANO)+F(MES)+PM+F(ANO):PM	Binomial
6	REP = ANO+F(MES)+PM+ ANO:F(MES)+ANO:PM+F(MES):PM	Poisson
7	REP = ANO+F(MES)+PM+ANO:F(MES)+ANO:PM	Poisson
8	REP = F(ANO)+F(MES)+PM+F(ANO):PM+F(MES):PM	Poisson
9	REP = ANO+F(MES)+PM+ANO:F(MES)+F(MES):PM	Poisson
10	REP = F(ANO)+F(MES)+PM+F(ANO):PM	Poisson
11	REP = ANO+F(MES)+PM+ ANO:F(MES)+ANO:PM+F(MES):PM	Binomial Negativa
12	REP = ANO+F(MES)+PM+ ANO:F(MES)+ANO:PM	Binomial Negativa
13	REP = F(ANO)+F(MES)+PM+ F(ANO):PM+F(MES):PM	Binomial Negativa
14	REP = ANO+F(MES)+PM+ ANO:F(MES)+F(MES):PM	Binomial Negativa
15	REP = F(ANO)+F(MES)+PM+ F(ANO):PM	Binomial Negativa
	Panulirus laevicauda	
1	REP = F(MES)+PM+ F(ANO)+F(MES):PM+PM:F(ANO)	Binomial
2	REP = F(MES)+PM+ F(ANO)+F(MES):PM	Binomial
3	REP = F(MES)+PM+ F(ANO)+F(ANO):PM	Binomial
4	REP = F(MES)+PM+ F(ANO)	Binomial
5	REP = F(MES)+PM+ ANO+F(MES):ANO+PM:ANO	Binomial
6	REP = F(MES)+PM+ F(ANO)+F(MES):PM+PM:F(ANO)	Poisson
7	REP = F(MES)+PM+ F(ANO)+F(MES):PM	Poisson
8	REP = F(MES)+PM+ F(ANO)+F(ANO):PM	Poisson
9	REP = F(MES)+PM+ F(ANO)	Poisson
10	REP = F(MES)+PM+ ANO+F(MES):ANO+PM:ANO	Poisson
11	REP = F(MES)+PM+ F(ANO)+F(MES):PM+PM:F(ANO)	Binomial Negativa
12	REP = F(MES)+PM+ F(ANO)+F(MES):PM	Binomial Negativa
12	REP = F(MES)+PM+ F(ANO)+F(ANO):PM	Binomial Negativa
14	REP = F(MES)+PM+ F(ANO)	Binomial Negativa
15	REP = F(MES)+PM+ ANO+F(MES):ANO+PM:ANO	Binomial Negativa

Fonte: Elaborada pelo autor.

Dentre os quinze modelos selecionados para cada espécie (cinco para cada distribuição), os modelos com distribuição binomial foram os que apresentaram menores valores nas estatísticas AIC, BIC e log-verossimilhança para ambas as

espécies (Tabela 5). No entanto, ressalta-se as restrições das comparações quando a variável resposta é diferente, sendo assim, a análise dos resíduos foi decisiva para a escolha dos modelos com distribuição binomial.

Tabela 5 – Comparativo dos modelos ajustados para análise das lagostas em fase reprodutiva. Modelos de 1 a 5 foram ajustados com distribuição binomial; modelos de 6 a 10 foram ajustados com distribuição Poisson; modelos de 11 a 15 foram ajustados com distribuição binomial negativa. Deviância – deviância do modelo nulo (composto somente pelo intercepto); GL – graus de liberdade do modelo nulo; Dev. Residual – deviância residual; GL Residual – graus de liberdade residual; LogLik – estatística de log-verossimilhança; BIC – Critério de Informação Bayesiano; AIC – Critério de Informação de Akaike; Delta AIC – diferença entre o AIC do modelo que está sendo observado e o modelo com menor AIC.

Modelos	Deviância	GL	Dev. Residual	GL Residual	LogLik	BIC	AIC	Delta AIC	Theta
				Panulirus argus					
1	12398,3	1135	2264,1	1099	-1830,5	3921,2	3734,9	0,0	-
2	12398,3	1135	2345,3	1110	-1871,0	3925,0	3794,1	59,2	-
3	12398,3	1135	2475,5	1098	-1836,1	4139,6	3948,2	213,3	-
4	12398,3	1135	2551,9	1100	-1974,3	4202,0	4020,7	285,8	-
5	12398,3	1135	2592,4	1109	-1994,6	4179,1	4043,2	308,3	-
6	7849,3	1135	1819,4	1099	-2073,5	4407,3	4220,9	0,0	-
7	7849,3	1135	1988,1	1110	-2157,9	4498,6	4367,7	146,8	-
8	7849,3	1135	2028,7	1098	-2178,1	4623,6	4432,2	211,3	-
9	7849,3	1135	2180,2	1100	-2253,8	4761,0	4579,8	358,9	-
10	7849,3	1135	2219,0	1109	-2273,6	4736,5	4600,6	379,7	-
11	3676,2	1135	1127,6	1099	-1973,8	4214,8	4023,5	0,0	5,8
12	3283,4	1135	1122,0	1110	-2015,6	4221,2	4085,2	61,7	4,4
13	3245,3	1135	1118,2	1098	-2018,4	4311,2	4114,8	91,3	4,3
14	2984,3	1135	1144,4	1100	-2066,1	4392,6	4206,3	182,8	3,6
15	2958,3	1135	1110,9	1109	-2053,1	4303,1	4162,2	138,7	3,5
				Panulirus laevicauda					
1	3848,3	710	1430,5	673	-1182,7	2615,0	2441,5	0,0	-
2	3848,3	710	1472,2	680	-1203,6	2610,8	2469,2	27,7	-
3	3848,3	710	1519,5	684	-1227,2	2631,7	2508,4	66,9	-
4	3848,4	711	1557,5	691	-1246,2	2623,8	2532,5	91,0	-
5	3848,5	712	1662,1	685	-1298,5	2767,8	2649,1	207,6	-
6	1450,5	710	671,9	673	-1347,3	2944,1	2770,6	5,9	-
7	1450,6	710	680,0	680	-1351,4	2906,2	2764,7	0,0	-
8	1450,6	710	737,9	684	-1380,3	2938,0	2814,7	50,0	-
9	1450,6	710	740,4	691	-1381,6	2894,5	2803,2	38,5	-
10	1450,6	710	754,8	685	-1388,8	2948,3	2829,5	64,8	-
11	1394,3	710	655,5	673	-1347,0	2950,1	2772,0	5,9	335,0
12	1393,8	710	663,4	680	-1351,1	2912,2	2766,1	0,0	332,0
13	1323,3	710	696,6	684	-1378,8	2941,5	2813,6	47,5	131,6
14	1325,2	710	699,8	691	-1380,1	2898,1	2802,2	36,1	134,0
15	1329,4	710	716,4	685	1387,8	2952,7	2829,4	63,3	139,7

Fonte: Elaborada pelo autor.

Para a espécie *P. argus*, o modelo mais parcimonioso, com menor número de parâmetros, e que melhor descreve os dados de lagostas capturadas em atividade reprodutiva no nordeste brasileiro foi o modelo 1 com distribuição binomial (Tabela 5), que considera todas as interações de primeira ordem entre as variáveis e cujo tratamento das variáveis explicativas no modelo foi “ANO” e “PM” como covariável e “MES” como fator (Tabela 4). Para a espécie *P. laevicauda,* o critério de informação de Akaike e o critério de informação Bayesiano são conflitantes sob o ponto de vista dos modelos 1 e 2, com a indicação de diferentes modelos binomiais.

Segundo Wagenmakers e Farrell (2004), quando o número de observações n é muito elevado, a probabilidade de que o AIC selecione o verdadeiro modelo com baixa dimensão entre os diversos avaliados não se aproxima da unidade, sendo mais adequado o uso do BIC. Vrieze (2012) destaca que, assintoticamente, o BIC é consistente, pois selecionará o modelo verdadeiro se o mesmo estiver entre os modelos candidatos considerados. O AIC não é consistente nessas circunstâncias.

Quando o modelo verdadeiro não está no conjunto de modelos candidatos, o AIC é eficiente, na medida em que assintoticamente é escolhido um modelo minimizando o erro quadrático médio de predição ou estimativa. O BIC não é eficiente nessas circunstâncias. Em adição, ao contrário do BIC, o AIC também possui uma propriedade minimax, na medida em que minimiza o risco máximo possível em tamanhos de amostra finitos. Medel e Salgado (2013) usando dados simulados, testaram as hipóteses do BIC ser mais parciomonioso e melhor do que o AIC para fins de previsão. A conclusão desse estudo foi que ambos os procedimentos de teste fornecem o mesmo resultado: o BIC mostra uma superioridade dentro e fora da amostra sobre AIC somente em um contexto de amostra de grande tamanho.

Burnham e Anderson (2002) destacam que modelos com Delta AIC entre 0 e 2 unidades devem ser examinados para ver se eles diferem do melhor modelo por um parâmetro e têm essencialmente os mesmos valores da log-verossimilhança maximizada. Modelos com Delta AIC>10 não têm essencialmente suporte e podem ser omitidos de uma consideração mais aprofundada, ou pelo menos esses modelos não conseguem explicar uma variação substancial nos dados. Sendo assim, o modelo binomial 2 ajustado para a espécie *P. laevicauda* apresentou desempenho fraco, já que o Delta AIC calculado está acima do limite de 10 quando comparado ao “melhor” modelo (modelo 1).

Aplicou-se ainda o teste da razão de verossimilhanças entre os modelos binomiais 1 e 2 e conclui-se que a interação entre as variáveis PM e ANO do modelo 1 é importante para explicar a variação na proporção de lagostas em atividade reprodutiva, considerando um nível de significância de 5%. Assim o modelo escolhido, apesar do valor do BIC ser ligeiramente superior, foi o modelo 1 com um maior número de variáveis explicativas.

Estendeu-se o teste de razão de verosimilhanças para os demais modelos encaixados e concluiu-se que existem diferenças estatisticamente significativas entre a maioria dos modelos na qualidade do ajustamento (p-valor < 0,05), com exceção apenas dos modelos de Poisson 6 e 7, 8 e 9 para a espécie *P. laevicauda* e, dos modelos binomiais negativos 11 e 13, 12 e 13, 12 e 14, 12 e 15 para a espécie *P. argus* e, 11 e 12, 13 e 14, 14 e 15 para a espécie *P. laevicauda*.

A análise da deviância apresentada na Tabela 6, refere-se aos modelos binomiais selecionados para elucidar os padrões nas variações das proporções de lagostas em reprodução.

Tabela 6 – Análise da deviância para os modelos binomiais selecionados para a espécie *Panulirus argus* e *Panulirus laevicauda*. GL – graus de liberdade; Dev – deviância; ANO – ano em que as lagostas foram capturadas; MES – mês que as lagostas foram capturadas; PM – comprimento das lagostas capturadas. Interações particulares são representadas por dois pontos.

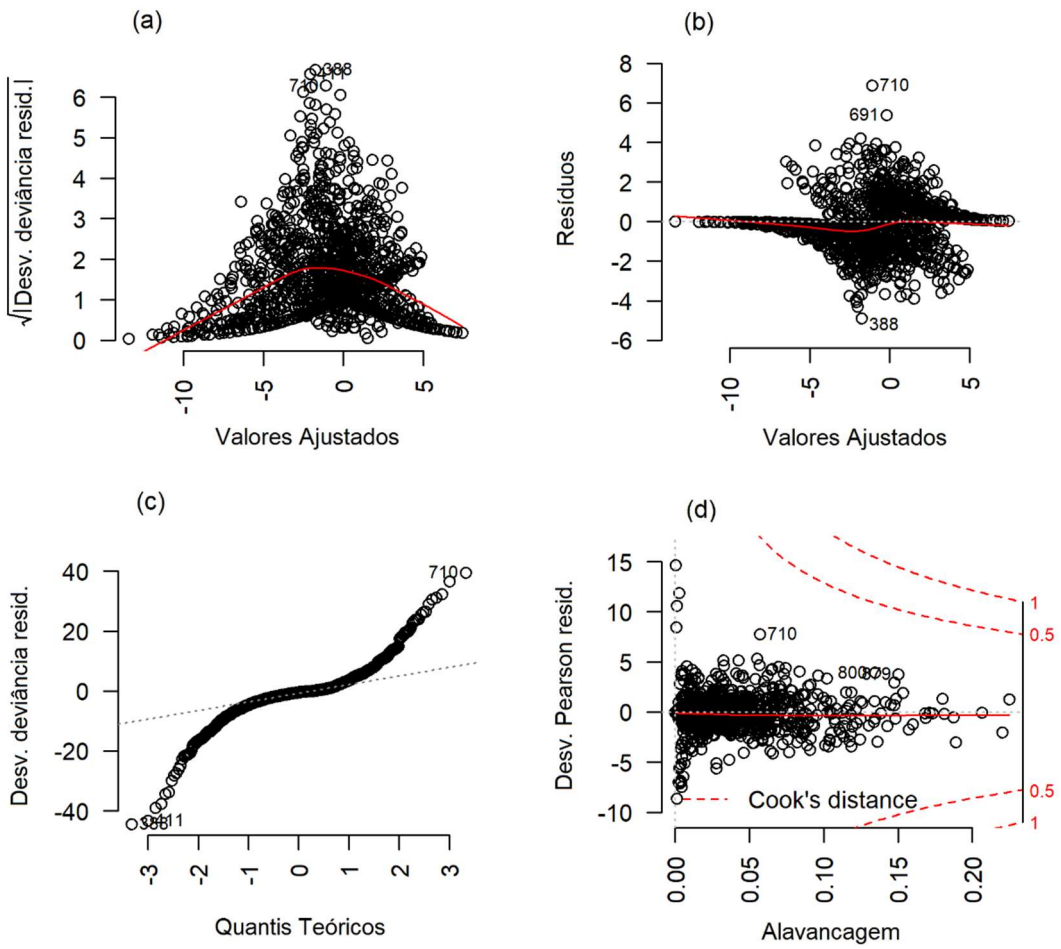
	GL	Deviância	GL Residual	Dev. Residual	Pr (>Chi)
			Panulirus argus		
			1135	12398,3	
ANO	1	3542,5	1134	8855,9	< 2,2e-16*
MES	11	2974,0	1123	5881,9	< 2,2e-16*
PM	1	2375,8	1122	3506,0	< 2,2e-16*
ANO:MES	11	782,2	1111	2723,8	< 2,2e-16*
ANO:PM	1	378,5	1110	2345,3	< 2,2e-16*
MES:PM	11	81,2	1099	2264,1	8,7e-13*
			Panulirus laevicauda		
			710	3848,3	
MES	11	1745,6	699	2102,7	< 2,2e-16*
PM	1	207,6	698	1895,1	< 2,2e-16*
ANO	7	337,6	691	1557,5	< 2,2e-16*
MES:PM	11	85,3	680	1472,2	1,4e-13*
PM:ANO	7	41,7	673	1430,5	5,8e-07*

Fonte: Elaborada pelo autor.
Nota: Nível de significância: p<0,05 (*).

A deviância residual para o modelo nulo (composto somente pelo intercepto) foi de 12.398,3 e 3.848,3 para as espécies *P. argus* e *P. laevicauda*, respectivamente (Tabela 6). A adição sucessiva de cada termo no modelo promoveu uma redução significativa na deviância (α=0,05), confimando a importância de cada variável e interação incluída no modelo. A proporção de deviância explicada pelo modelo ajustado para a espécie *P.argus*, foi de aproximadamente, 81,73%, sendo que o fator ano é responsável por 34,96% desse total, seguido da variável mês, com 29,35% da variabilidade explicada. Para a espécie *P. laevicauda*, a proporção de deviância explicada pelo modelo foi de aproximadamente, 62,83%, sendo que o fator mês explicou 72,2% desse total, seguido da variável ano, com 13,96% da variabilidade explicada.

Na análise dos resíduos, vários problemas foram identificados nos modelos ajustados para ambas as espécies. As distribuições dos resíduos para a *P. argus* apresentam comportamentos heterocedásticos (Figura 7(a)), confirmado pelo teste de Breusch-Pagan ($p<0,05$). Na Figura 7(b), nota-se que há indícios de viés nas estimativas. Os resíduos não possuem uma distribuição normal, com violações nas caudas (Figura 7(c)), confirmado pelo teste de Kolmogorov-Smirnov ($p<0,05$). As observações que correspondem a resíduos relativamente elevados têm pouco poder de alavancagem (Figura 7(d)).

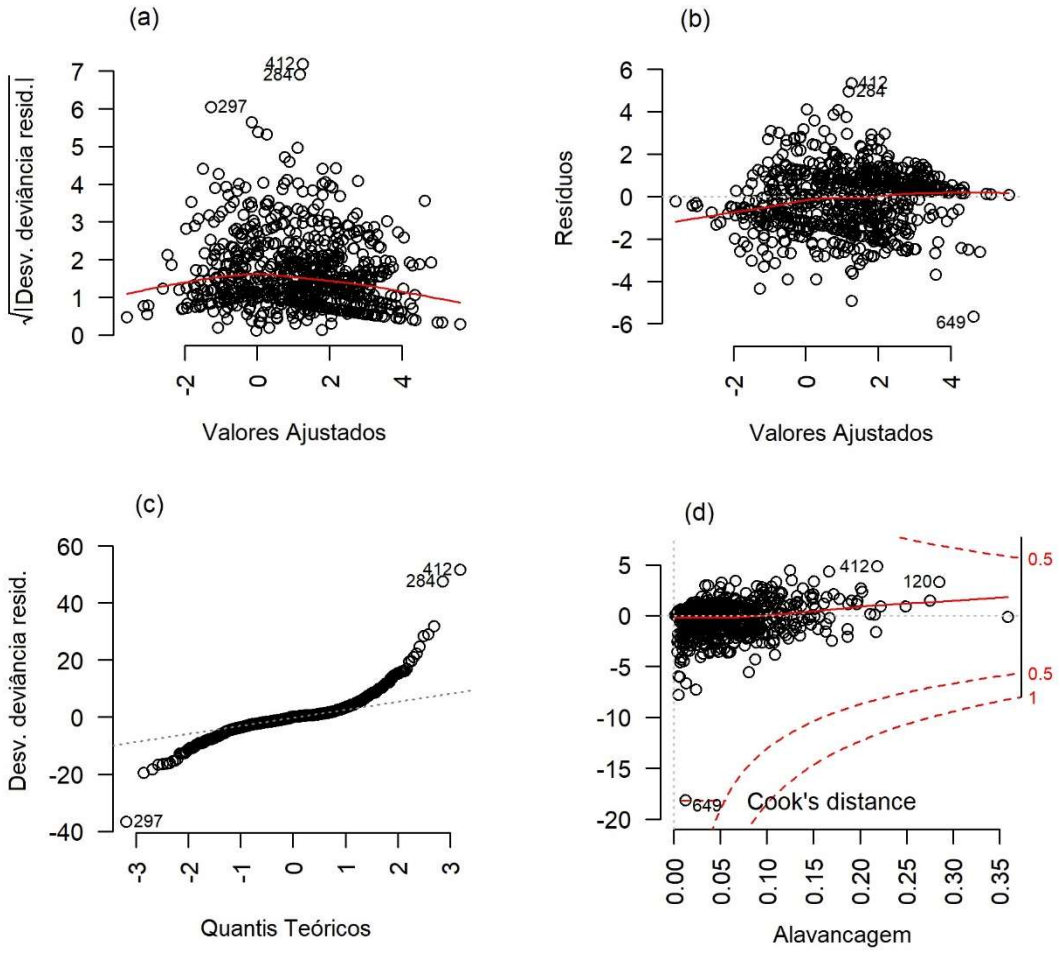
Figura 7 – Diagnóstico do ajuste do modelo para a espécie *Panulirus argus*.



Fonte: Elaborada pelo autor.

Para a espécie *P. laevicauda*, os resíduos apresentam comportamento heterocedástico (Figura 8(a)), confirmado pelo teste de Breusch-Pagan ($p<0,05$), há indícios de viés nas estimativas (Figura 8(b)), a distribuição de probabilidade dos resíduos não é normal como confirmado pelo teste de Kolmogorov-Smirnov ($p<0,05$) (Figura 8(c)), e as observações que correspondem a resíduos relativamente elevados têm pouco poder de alavancagem (Figura 8(d)).

Figura 8 – Diagnóstico do ajuste do modelo para a espécie *Panulirus laevicauda*.



Fonte: Elaborada pelo autor.

Com base na análise de diagnóstico dos resíduos conclui-se que os modelos selecionados não se ajustam bem aos dados, visto que apresentam vários problemas. Diante deste fato, foram utilizados os modelos aditivos generalizados para posição, escala e forma (gamlss) que permitem flexibilidade na modelagem estatística.

2.4.3 Modelos aditivos generalizados para posição, escala e forma (GAMLSS)

As variáveis utilizadas para modelar o parâmetro $\boldsymbol{\mu}$ nos GLMs selecionados, foram mantidas para o ajuste inicial dos modelos com as diversas distribuições disponíveis no pacote gamlss. Nessa etapa inicial não foram adicionadas variáveis no preditor linear para modelar os parâmetros $\boldsymbol{\sigma}$ e $\boldsymbol{\nu}$, sendo os modelos compostos

apenas pelo intercepto. O modelo com distribuição beta binomial foi superior se consideradas as estatísticas AIC e BIC (Tabela 7).

Tabela 7 – Comparativo dos modelos ajustados para análise das lagostas em fase reprodutiva. GL – representa graus de liberdade; AIC – Critério de Informação de Akaike; BIC – Critério de Informação Bayesiano.

Distribuições	GL	AIC	BIC	GL	AIC	BIC
	Panulirus argus			*Panulirus laevicauda*		
BB	38	3203,5	3394,9	39	2192,3	2370,4
ZIBB	39	3205,5	3401,9	40	2192,7	2375,4
NBF	39	3987,1	4183,4	40	2767,9	2950,6
ZIBI	38	3619,3	3810,6	39	2395,2	2573,3
NBII	38	4000,5	4191,8	39	9329,2	9507,3
ZIBNB	40	4003,3	4204,7	41	2639,6	2826,8
ZISICHEL	40	4009,5	4211,0	-	-	-
ZINBI	39	4011,7	4208,1	40	2758,9	2941,6
BNB	39	4012,8	4209,1	40	3204,0	3386,7
DPO	38	4019,9	4211,3	39	2766,7	2944,8
NBI	38	4023,5	4214,8	39	2772,0	2950,1
SICHEL	39	4024,4	4220,8	-	-	-
PIG	38	4025,1	4216,5	39	2772,0	2950,1
ZIP2	38	4117,9	4309,2	39	2756,9	2935,0
ZIP	38	4117,9	4309,2	39	2756,9	2935,0
ZABI	38	4262,3	4453,6	39	2602,1	2780,2
GEOM	37	4292,4	4478,7	38	3451,4	3624,9
ZANBI	39	4561,8	4758,2	40	2940,9	3123,6
ZAP	38	4618,0	4809,3	39	2938,9	3117,0
ZALG	38	5272,4	5463,8	39	3955,2	4133,3
ZAZIPF	38	6952,0	7143,4	39	5502,5	5680,6
GEOMo	37	7413,7	7600,0	38	5908,9	6082,5
ZIPIG	39	5265,6	5462,0	40	2763,4	2946,1
DEL	39	4027,8	4224,2	40	2774,6	2957,3

Distribuições: BB – Beta Binomial; ZIBB – Beta Binomial Inflacionada de Zero; NBF – Família Binomial Negativa; ZIBI – Binomial Inflacionada de Zero; NBII – Binomial Negativa Tipo II; ZIBNB – Beta Binomial Negativa Inflacionada de Zero; ZISICHEL – Sichel Inflacionada de Zero; ZINBI – Binomial Negativa Inflacionada de Zero; BNB – Beta Binomial Negativa; DPO – Poisson Dupla; NBI - Binomial Negativa Tipo I; SICHEL – Sichel (μ é a média); PIG – Poisson Inversa Gaussiana; ZIP2 – Poisson Inflacionada de Zero (μ é a média); ZABI – Binomial Ajustada em Zero; GEOM – Geométrica; ZANBI – Binomial Negativa Ajustada em Zero; ZAP – Poisson Ajustada em Zero; ZALG – Logarítmica Ajustada em Zero; ZAZIPF – Zipf Ajustada de Zero; GEOMo – Geométrica (original); ZIPIG – Poisson Inversa Gaussiana Inflacionada de Zero; DEL – Delaporte. Modelos com distribuição Sichel e modelos com distribuição Sichel Inflacionada de Zero não convergiram para a espécie *Panulirus laevicauda.*

Na Tabela 8 são apresentados os valores das estatísticas de ajustamento do modelo beta binomial selecionado inicialmente para cada espécie.

Tabela 8 – Modelos iniciais ajustados para análise das lagostas em fase reprodutiva. GL – representa graus de liberdade; AIC – o Critério de Informação de Akaike; BIC – o Critério de Informação Bayesiano.

	Panulirus argus	*Panulirus laevicauda*
Distribuição	Beta Binomial	Beta Binomial
Nº observações estimadas	1136	711
GL. Residual	1098	672
Nº Interações	7	7
Deviância global	3127,5	2114,3
AIC	3203,5	2192,3
BIC	3394,9	2370,4

Fonte: Elaborada pelo autor.

Ainda na tentativa de melhorar o ajuste dos modelos selecionados foi utilizado a estratégia A, implementada na função stepGAICAll.A(), destacada por Stasinopoulo*s et al*. (2017) e apresentada na sessão 2.3.2.4. De acordo com os resultados apresentados na Tabela 9, nota-se que a adição sucessiva de termos para modelar o parâmetro $\boldsymbol{\mu}$ e $\boldsymbol{\sigma}$ melhora o ajuste do modelo para a espécie *Panulirus argus*. Sendo assim, as variáveis PM, MES, ANO e todas as interações de primeira ordem ANO:MES, ANO:PM, MES:PM foram utilizadas para modelar os parâmetros de localização e escala da distribuição beta binomial.

Tabela 9 – Seleção de termos aditivos para os parâmetros da distribuição beta binomial utilizando a estratégia A. AIC – Critério de Informação de Akaike; REP – proporção de lagostas vermelhas capturadas com ovos ou espermoteca ou resto de espermoteca; ANO – ano em que as lagostas vermelhas foram capturadas; MES – mês que as lagostas vermelhas foram capturadas; PM - comprimento das lagostas vermelhas capturadas. Interações particulares são representadas por dois pontos.

Modelos	AIC
Panulirus argus	
Parâmetro de distribuição: μ	
REP=INTERCEPTO	4528,3
REP=INTERCEPTO+PM	3991,0
REP=INTERCEPTO+ANO	4296,9
REP=INTERCEPTO+MES	4363,6
REP=INTERCEPTO+PM	3991,0
REP=INTERCEPTO+PM+MES	3753,1
REP=INTERCEPTO+PM+ANO	3788,4
REP=INTERCEPTO+PM+MES	3753,1
REP=INTERCEPTO+PM+MES+ANO	3462,3
REP=INTERCEPTO+PM+MES+MES:PM	3749,5
REP=INTERCEPTO+PM+MES+ANO	3462,3
REP=INTERCEPTO+PM+MES+ANO+ANO:MES	3326,0
REP=INTERCEPTO+PM+MES+ANO+ANO:PM	3339,7
REP=INTERCEPTO+PM+MES+ANO+MES:PM	3429,3
REP=INTERCEPTO+PM+MES+ANO+ANO:MES	3326,0
REP=INTERCEPTO+PM+MES+ANO+ANO:MES+ANO:PM	3217,7
REP=INTERCEPTO+PM+MES+ANO+ANO:MES+MES:PM	3287,7
REP=INTERCEPTO+PM+MES+ANO+ANO:MES+ANO:PM	3217,7
REP=INTERCEPTO+PM+MES+ANO+ANO:MES+ANO:PM+MES:PM	3203,5
Parâmetro de distribuição: σ	
REP=INTERCEPTO	3203,5
REP=INTERCEPTO+PM	3205,4
REP=INTERCEPTO+MES	3206,6
REP=INTERCEPTO+ANO	3190,9
REP=INTERCEPTO+ANO	3190,9
REP=INTERCEPTO+ANO+MES	3185,6
REP=INTERCEPTO+ANO+PM	3192,1
REP=INTERCEPTO+ANO+MES	3185,6
REP=INTERCEPTO+ANO+MES+ANO:MES	3179,2
REP=INTERCEPTO+ANO+MES+PM	3186,2
REP=INTERCEPTO+ANO+MES+ANO:MES	3179,2
REP=INTERCEPTO+ANO+MES+ANO:MES+PM	3178,2
REP=INTERCEPTO+ANO+MES+PM+ANO:MES	3178,2
REP=INTERCEPTO+ANO+MES+PM+ANO:MES+ANO:PM	3169,1
REP=INTERCEPTO+ANO+MES+PM+ANO:MES+MES:PM	3169,5
REP=INTERCEPTO+ANO+MES+PM+ANO:MES+ANO:PM	3169,1
REP=INTERCEPTO+ANO+MES+PM+ANO:MES+ANO:PM+MES:PM	3165,2

Fonte: Elaborada pelo autor.

Para a espécie *P. laevicauda*, Tabela 10, as variáveis PM, MES, ANO e as interações de primeira ordem ANO:PM, MES:PM são importantes para modelar o parâmetro μ da distribuição beta binomial. No entanto, esse comportamento não se repete na modelagem do parâmetro σ. A adição da variável ANO para modelar o parâmetro sigma provoca uma redução no valor do AIC. Contudo, a inclusão desse termo ocasiona problemas na convergência do modelo, mesmo mudando o método de estimação e aumentando o número de interações. Por este motivo, o σ configurado como constante foi o mais adequado para sua modelagem.

Tabela 10 – Seleção de termos aditivos para os parâmetros da distribuição beta binomial utilizando a estratégia A. AIC – Critério de Informação de Akaike; REP – proporção de lagostas verdes capturadas com ovos ou espermoteca ou resto de espermoteca; ANO – ano em que as lagostas verdes foram capturadas; MES – mês que as lagostas verdes foram capturadas; PM - comprimento das lagostas verdes capturadas. Interações particulares são representadas por dois pontos.

Modelos	AIC
Panulirus laevicauda	
Parâmetro de distribuição: μ	
REP=INTERCEPTO	2660,3
REP=INTERCEPTO+MES	2380,6
REP=INTERCEPTO+PM	2576,8
REP=INTERCEPTO+ANO	2638,6
REP=INTERCEPTO+MES	2380,6
REP=INTERCEPTO+MES+PM	2267,7
REP=INTERCEPTO+MES+ANO	2351,1
REP=INTERCEPTO+MES+PM	2267,7
REP=INTERCEPTO+MES+PM+ANO	2210,6
REP=INTERCEPTO+MES+PM+MES:PM	2263,8
REP=INTERCEPTO+MES+PM+ANO	2210,8
REP=INTERCEPTO+MES+PM+ANO+MES:PM	2200,8
REP=INTERCEPTO+MES+PM+ANO+ANO:PM	2205,6
REP=INTERCEPTO+MES+PM+ANO+MES:PM	2200,8
REP=INTERCEPTO+MES+PM+ANO+MES:PM+ANO:PM	2192,3
Parâmetro de distribuição: σ	
REP=INTERCEPTO	2192,3
REP=INTERCEPTO+ANO	2174,7
REP=INTERCEPTO+MES	2192,3
REP=INTERCEPTO+PM	2192,8
REP=INTERCEPTO+ANO	2174,7
REP=INTERCEPTO+ANO+MES	2175,8
REP=INTERCEPTO+ANO+PM	2177,4

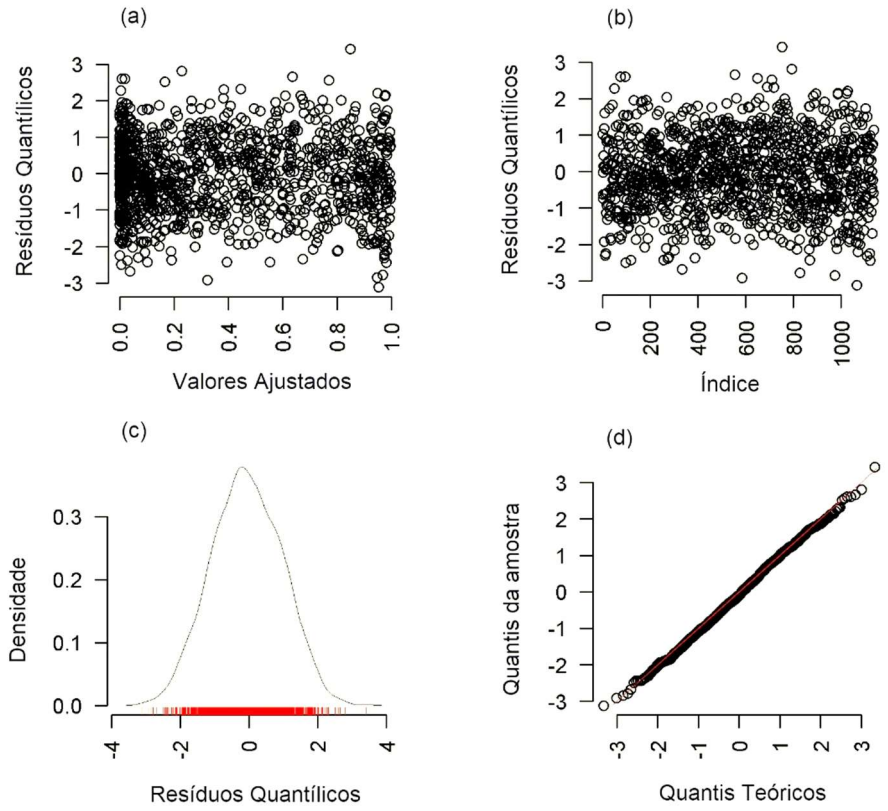
Fonte: Elaborada pelo autor.

A análise dos resíduos indica um bom ajuste dos modelos beta binomiais selecionados para ambas as espécies, visto que: a) há homocedasticidade (Figura 9(a) e Figura 10(a)), confirmada com a aplicação do teste de Breusch-Pagan ($p<0,05$); b) não há indícios de viés nas estimativas (Figura 9(b) e Figura 10(b)); e c) não se rejeita a hipótese de normalidade (Figura 9(c), Figura 10(c), Figura 9(d) e Figura 10(d)) no teste de Kolmogorov-Smirnov ($p>0,05$). As medidas descritivas dos resíduos quantílicos aleatorizados dos modelos beta binomiais ajustados para ambas as espécies confirmam esses resultados.

Para a espécie *P. argus,* a média dos resíduos quantílicos aleatorizados foi -0,05 (quase zero), a variância 1,01 (quase um), o coeficiente de assimetria 0,02 (próximo a zero, indicando distribuição simétrica), o coeficiente de curtose 3,04 (próximo a três, indicando distribuição mesocúrtica) e coeficiente de correlação de Filliben 1,00 ($p>0,05$, indicando normalidade). Para a espécie *P. laevicauda,* os resultados foram semelhantes, a média dos resíduos quantílicos aleatorizados foi -0,08, a variância 1,04, o coeficiente de assimetria -0,01, o coeficiente de curtose 3,07 e coeficiente de correlação de Filliben 1,00.

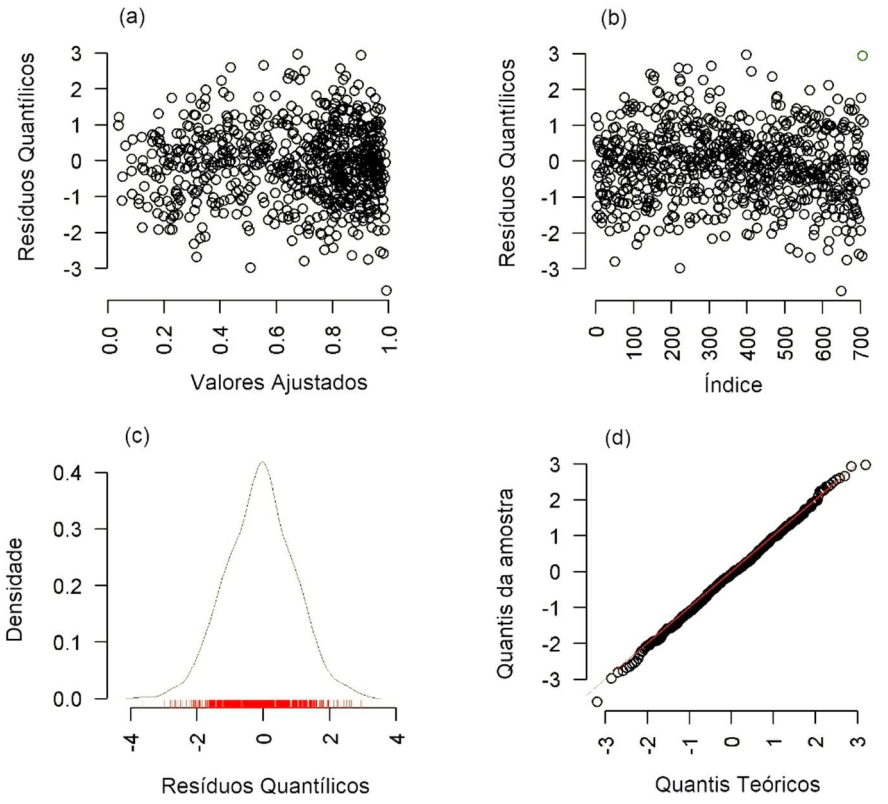
De acordo com Dunn e Smyth (1996), se um modelo se ajustar bem aos dados, os resíduos deverão apresentar distribuição aproximadamente normal padrão, ou seja, média próximo de zero, variância próximo de 1, coeficiente de assimetria próximo de 0 e coeficiente de curtose próximo de 3, mesmo quando a distribuição do modelo não é normal. Feng *et al.* (2017), em estudos de simulação, mostraram que o resíduo quantílico aleatorizado tem uma distribuição normal unificada sob o modelo verdadeiro e possui grande poder estatístico na detecção de muitas formas de inadequações do modelo.

Figura 9 – Diagnóstico do modelo ajustado para a espécie *Panulirus argus*.



Fonte: Elaborada pelo autor.

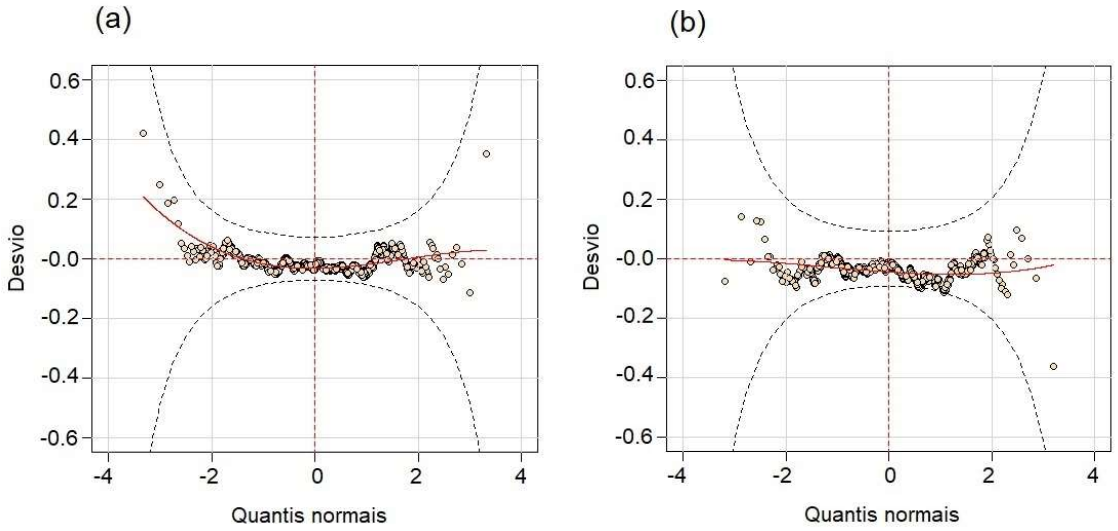
Figura 10 – Diagnóstico do modelo ajustado para a espécie *Panulirus laevicauda*.



Fonte: Elaborada pelo autor.

Para verificar a adequação dos modelos beta binomiais selecionados, foi utilizado ainda o gráfico denominado de *Worm Plo*t (WP) de VAN BUUREN e FREDRIKS (2001) (Figura 11). Para a espécie *P. argus*, todos os pontos estão dentro da região delimitada pelo intervalo de confiança de 95%, indicando um bom ajustamento do modelo beta binomial. Para a espécie *P. laevicauda*, nota-se alguns pontos sobre o limite da faixa de confiança inferior, no entanto, 95% dos pontos estão entre as duas curvas elípticas, indicando que a distribuição ajustada (ou os termos ajustados) do modelo são adequadas para explicar a variável resposta.

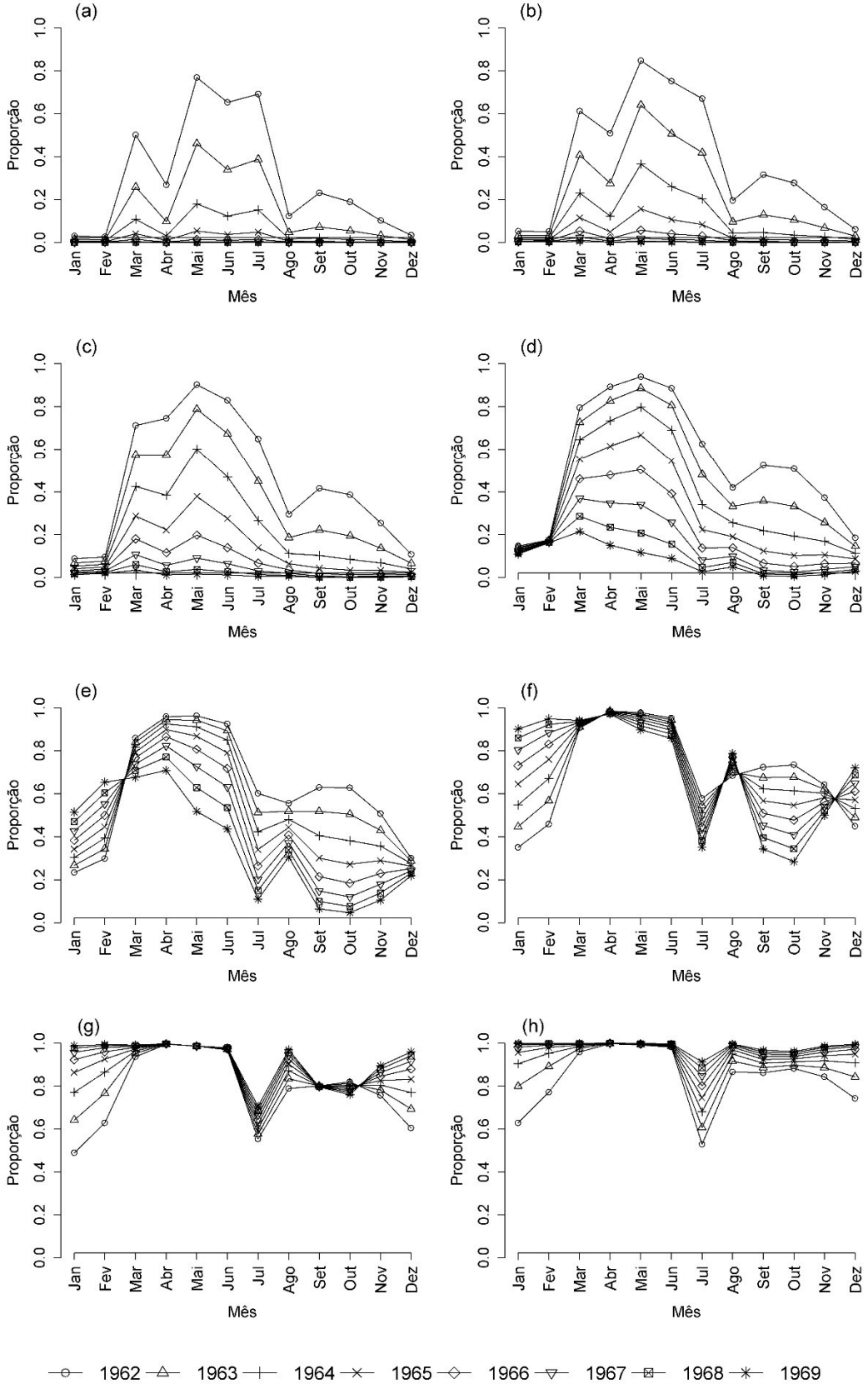
Figura 11 – Worm-plot dos modelos ajustados com distribuição beta binomial para as espécies *Panulirus argus* (a) e *Panulirus laevicauda* (b). Diferenças entre os z-escores dos valores empíricos padronizados e os z-escores dos valores esperados sob a distribuição teórica contra os z-escores dos valores teóricos.



Fonte: Elaborada pelo autor.

As predições da esperança calculadas com o modelo ajustado para as espécies *Panulirus argus* e *Panulirus laevicauda* foram realizadas considerando a amplitude total do comprimento da lagosta no momento da captura. Fixou-se uma variação de 3 cm no comprimento para mostrar as esperanças calculadas com o preditor linear (Figura 12 e Figura 13).

Figura 12 – Predições da esperança calculadas com o modelo ajustado para a espécie *Panulirus argus*, fixando um comprimento total da lagosta (PM) no momento da captura de 12,55cm (a), 15,55cm (b), 18,55cm (c), 21,55cm (d), 24,55cm (e), 27,55cm (f), 30,55cm (g) e 33,55cm (h).

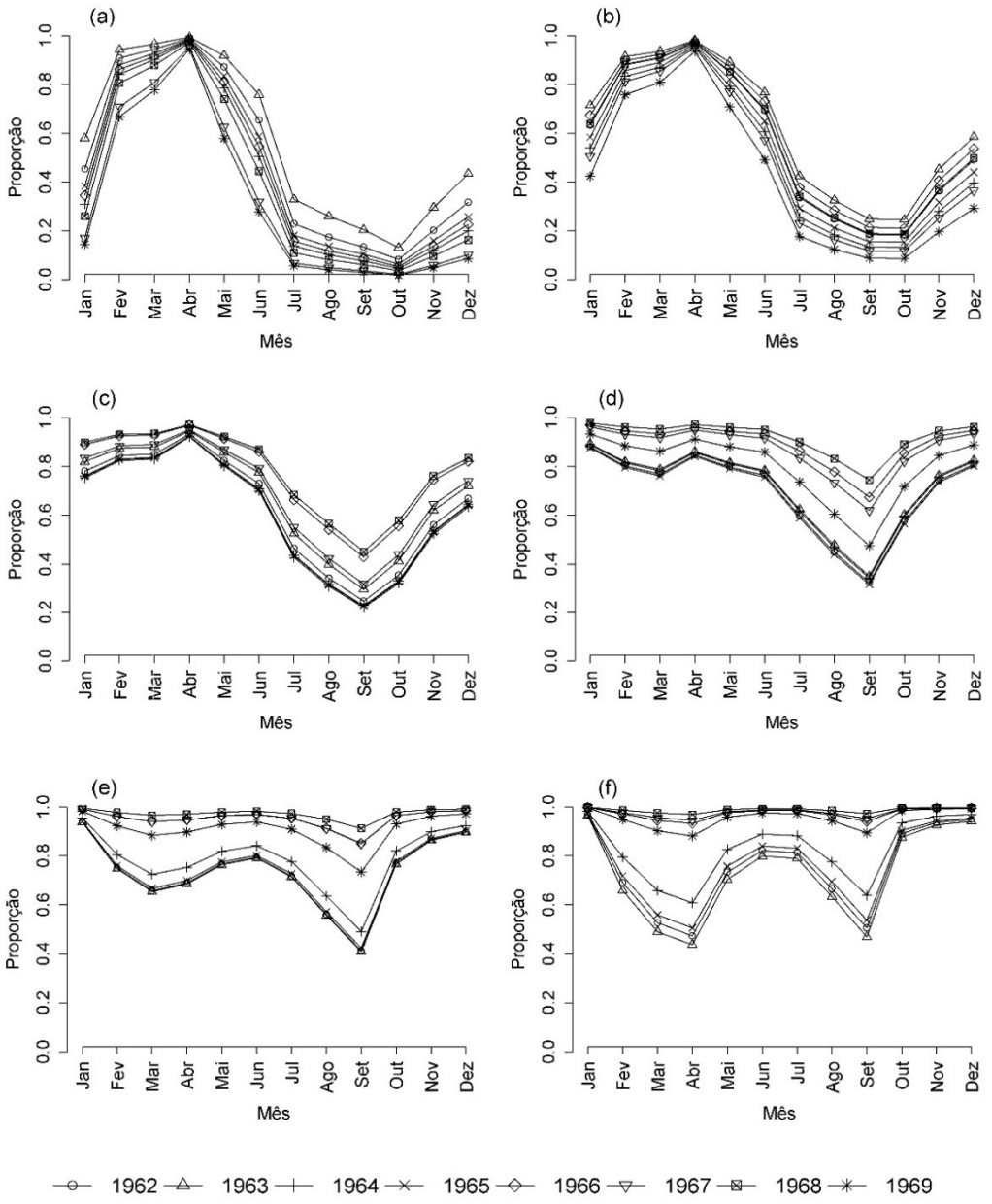


Fonte: Elaborada pelo autor.

Através das predições apresentadas na Figura 12, nota-se que a probabilidade de encontrar fêmeas da espécie *P. argus* em atividade reprodutiva é maior durante o primeiro semestre do ano, especialmente a partir de fevereiro. Nota-se ainda que essa probabilidade tende a crescer à medida que o comprimento da lagosta aumenta. Para os comprimentos de 12,55 cm a 24,55 cm, Figura 12(a) a Figura 12(e), há uma diminuição na proporção de exemplares com características reprodutivas de 1962 a 1969. Já para os comprimentos superiores de 27,55 cm a 33,55 cm, Figura 12(f) a Figura 12(h), o padrão é inverso, há um aumento na proporção de exemplares com características reprodutivas durante o período de 1962 a 1969.

Para as fêmeas da espécie *P. laevicauda* capturadas com comprimento de 12,55 cm o período de maior atividade reprodutiva ocorre durante o primeiro semestre do ano (Figura 13(a)). Nota-se que a medida que as lagostas verdes vão crescendo começa a surgir um outro período que ganha destaque, com início em meados de outubro (Figura 13(c)). De forma semelhante à espécie *P. argus*, também para a *P. laevicauda*, a probabilidade de que esteja em atividade reprodutiva é grande para exemplares capturados com comprimentos entre 24,55 cm e 27,55 cm (Figura 13(e) e Figura 13(f)). Nota-se também que a proporção de exemplares em atividade reprodutiva foi menor durante os anos de 1962 a 1965, para lagostas capturadas com comprimentos entre 24,55 cm e 27,55 cm (Figura 13(e) e Figura 13(f)).

Figura 13 – Predição da esperança calculadas com o modelo ajustado para a espécie *Panulirus laevicauda*, fixando um comprimento total da lagosta (PM) no momento da captura de 12,55cm (a), 15,55cm (b), 18,55cm (c), 21,55cm (d), 24,55cm (e) e 27,55cm (f).



Fonte: Elaborada pelo autor.

Ao avaliar em conjunto os documentos publicados por Paiva e Costa na década de 1960, verifica-se variações interanuais importantes no período de maior atividade reprodutiva das espécies *P. argus* e *P. laevicauda.* Para a espécie *P. argus*, houve dois períodos em que o fenômeno biológico se realizou com maior intensidade na população: o primeiro: de janeiro a abril (para os anos de 1966 e 1967) e de fevereiro a abril (para o ano de 1962); o segundo: de julho a setembro (para o ano de 1962), de

junho a setembro (para o ano de 1966) e de junho a agosto (para o ano de 1967). Para os demais anos, o período de maior atividade reprodutiva esteve compreendido entre os meses de janeiro a maio, variando o início e o término a depender do ano avaliado. De forma geral, fevereiro foi o mês com intensa atividade reprodutiva para todos os anos.

Há também variações interanuais importantes para a *P. laevicauda*. No entanto, no ano de 1962 e 1963, não foi possível delimitar períodos de intensa atividade reprodutiva, sendo registradas fêmeas em processo de reprodução em todos os meses do ano. A partir do ano de 1964, com exceção do ano de 1968, foram encontradas fêmeas se reproduzindo com maior intensidade em dois períodos do ano: o primeiro: fevereiro e março (para o ano de 1964), janeiro e fevereiro (para o ano de 1965), janeiro a abril (para o ano de 1966) e janeiro a junho (para o ano de 1967 e 1969); o segundo: de outubro a dezembro (para os anos de 1964, 1965, 1967 e 1968), de setembro a dezembro (para os anos de 1966) e de novembro e dezembro (para o ano de 1969). Em resumo, os meses de fevereiro e outubro a dezembro foram os mais ativos reprodutivamente para o período de 1964 a 1969.

As informações publicadas por diferentes autores sobre a época de reprodução das lagostas *P. argus* e *P. laevicauda* no Nordeste apresentam diferenças consideráveis. Coelho *et al.* (1963) constataram duas temporadas de acasalamento para ambas as espécies no litoral de Pernambuco, uma atingindo seu máximo em agosto-setembro e outra com pico em março. Paiva e Fonteles-Filho (1968), estudando a migração das lagostas vermelhas, observaram que o ciclo de reprodução destas ocorre entre os meses de março a maio e no mês de agosto. Nascimento (1984), estudando o período de reprodução, identificou dois picos de desova, um em março-maio e outro em agosto para a espécie *P. argus*. Soares e Cavalcante (1985) estudaram a dinâmica reprodutiva da lagosta e verificaram que fêmeas com massa espermatofórica e ovígeras podem ser encontradas com maiores frequências nos meses de janeiro a abril e setembro-outubro para a espécie *P. argus*. Soares (1990) identificou que o período de maior atividade reprodutiva para a lagosta vermelha ocorre durante o primeiro semestre do ano, especialmente entre os meses de fevereiro a maio, e que há ocorrência eventual de um segundo período de desova, geralmente em setembro e outubro. Silva *et al.* (2008), estudando a variação temporal das frequências dos estádios de maturidade e seu comprimento médio de primeira

maturação gonadal, constataram que a espécie *P. argus* apresenta desova contínua, sendo o terceiro trimestre o período de maior atividade reprodutiva.

Ao comparar os trabalhos de Paiva e Costa na década de 1960 com os trabalhos de Soares e Cavalcante (1985) e Soares (1990), há contradições quanto a identificação do mês de outubro como período de intensa atividade reprodutiva para a lagosta vermelha. No entanto, neste trabalho foi observado intensa atividade reprodutiva neste mês para lagostas vermelhas com comprimentos superiores a 27,55cm.

Para a espécie *P. laevicauda*, Ivo (1975) estabeleceu o período de abril a junho como de reprodução. Nascimento (1984) identificou dois picos de desova, um em março-julho e outro em outubro para a espécie *P. laevicauda*. Soares e Cavalcante (1985) verificaram que fêmeas com massa espermatofórica e ovígeras podem ser encontradas com maiores frequências nos meses de janeiro a maio para a espécie *P. laevicauda*. Soares e Fonteles-Filho (2000) utilizando a relação existente entre o desenvolvimento gonodal e os indicadores externos de maturidade sexual, mostrou que as fêmeas da espécie *P. laevicauda* se reproduz durante todo o ano, tendo um período de maior atividade reprodutiva nos meses de janeiro a maio.

Neste trabalho, fêmeas de *P. laevicauda* em reprodução são relativamente mais frequentes nos meses de julho a agosto à medida que atingem maiores comprimentos. De acordo com Ivo e Pereira (1996) as variações observadas por diferentes autores em relação a estas conclusões devem-se, provavelmente, a fatores casuais na amostra, ou ainda a variações ambientais e ou mesmo quanto ao local onde foi realizada a amostragem.

A escassez de estudos relacionados ao ciclo reprodutivo das lagostas capturadas no nordeste brasileiro dificulta o gerenciamento deste recurso pesqueiro, uma vez que as decisões sobre o estabelecimento de períodos de defesos são norteadas com base nessas informações. Os poucos estudos existentes sobre o período de reprodução das lagostas são, na grande maioria, apenas de caráter descritivo e levam em consideração a análise de registros feitos em determinados anos isoladamente.

Neste trabalho a análise da série temporal integrada com o uso do modelo aditivo generalizado para posição, escala e forma permitiu identicar a presença de variações interanuais importantes nos meses em que ocorreram picos de desova, como por exemplo, o aumento na proporção de exemplares da espécie *P. argus* com

características reprodutivas durante o período de 1962 a 1969 para os comprimentos superiores de 27,55 cm a 33,55 cm. Esse fato pode indicar que a procura por lagostas com maiores comprimentos aumentou nos últimos anos avaliados e, como lagostas maiores tem maior probabilidade de ao serem capturadas estarem em atividade reprodutiva, ocorre um aumento na proporção de lagostas maturas. Também foram registrados meses com atividade reprodutiva relativamente intensa que não fazem parte do período de defeso, como o mês de junho para a espécie *P.argus* e os meses de junho e novembro para a espécie *P. laevicauda*.

Diante desses resultados, o estudo do ciclo reprodutivo das lagostas torna-se ainda mais importante, visto que existiram períodos com intensa atividade reprodutiva que não fazem parte da paralisação da pesca. O Estabelecimento de um período de defeso correto é essencial, uma vez que garante a renovação da espécie.

Outro fato agravante são os sinais de sobreexploração. Atualmente os estoques de lagostas espinheiras são sobreexplorados em várias localidades, e compartilham um cenário formado por gestão ineficiente e decisões políticas com carência de embasamento científico (BUCARAM *et al.*, 2013; HARPER *et al.*, 2014; SPANIER *et al.*, 2015; BAISRE, 2017). Esses fatores contribuem tanto para o comércio ilegal quanto para a sobreexploração dos estoques locais (GIRALDES e SMYTH, 2016).

No caso da lagosta no Brasil, estudos desenvolvidos por Andrade (2015) confirmam que a pesca da lagosta vermelha passou por um período crítico desde os anos 2000 e comenta que apesar de todas as medidas empregadas pelo governo brasileiro para regular a captura de *P. argus*, há evidências claras de que os estoques estão sendo sobreexplotados, indicando que essas medidas são ineficazes ou não estão sendo adequadamente implementadas. O cenário negativo é também corroborado por Silva e Fonteles-Filho (2011) que estimaram que mais de 85% da captura total brasileira foi produzida ilegalmente. Em adição há mais de 6 anos, não existe um controle estatístico mínimo e informação insuficiente para a aplicação de um sistema de cotas de captura (Dias-Neto e Dias, 2015).

2.5 Conclusão

Na análise das variações das proporções de lagostas fêmeas em atividade reprodutiva, verificou-se que os ajustes de modelos lineares generalizados não foram satisfatórios, pois as distribuições dos resíduos não são normais, há heterocedasticidade e há indícios de viés nas estimativas. No entanto, os modelos aditivos generalizados para posição, escala e forma que permitem flexibilidade na modelagem estatística foram adequados para o estudo. Dentre as diversas distribuições utilizadas, a beta binomial foi a que melhor descreveu os dados.

O ciclo reprodutivo não apresenta periodicidade bem definida, com alguma desova ao longo de todos os meses. Em adição, há variações interanuais. Apesar da captura de lagostas com característica reprodutiva ter frequência relativa maior no período de defeso, a atividade é também relativamente intensa em outros períodos. O mês de junho é importante para a espécie *P. argus* e não está incluído no defeso. Os meses de junho e novembro são importantes para a reprodução da espécie *P. laevicauda* e não estão incluídos no defeso. Em adição, os meses de dezembro a fevereiro fazem parte do defeso, mas apresentam pequenas proporções relativas de fêmeas maturas de *P. argus.* Portanto, sugere-se uma reavaliação do período de defeso tradicionalmente utilizado a décadas no Brasil.

3 Padronização da CPUE do *Xiphias gladius* usando efeito fixo e aleatório

3.1 Introdução

No cenário nacional e internacional, além das lagostas já destacadas no capítulo 2, têm grande importância as pescarias de espadarte (*Xiphias gladius*) (FAO, 2012). O elevado valor de mercado, aliado à carne de ótima qualidade e à crescente demanda no cenário internacional fazem da espécie um dos peixes pelágicos mais visados pela indústria pesqueira mundial. De acordo com ICCAT (2015), a captura de *X. gladius* no Oceano Atlântico no ano de 2014 situou-se próxima de 21.000 toneladas. No Atlântico Sul, o Brasil foi o segundo país com maior contribuição do pescado, capturando 2.892 toneladas.

O espadarte pode ser encontrado em vastas áreas oceânicas no mundo inteiro, devido a sua capacidade de migração e ampla distribuição geográfica. Sendo assim, um estoque pesqueiro não pode ser gerido por uma nação. Agências internacionais são incumbidas de organizar e analisar as informações fornecidas por diversos países, com o intuito de gerar informações úteis para administrar as pescarias. No caso do oceano Atlântico, a agência responsável por essa gestão é a Comissão Internacional para a Conservação dos Atuns do Atlântico (*International Comission for the Conservation of Atlantic Tunas* – ICCAT). Para que a ICCAT possa definir recomendações de gestão para as pescarias, deve-se fazer o que se denomina de “avaliação de estoque” pesqueiro.

Nas avaliações de estoques, os principais objetivos são estimar a abundância e a produção potencial da população como balanço de recrutamento, crescimento somático e mortalidade (HILBORN e WALTERS, 1992; QUINN e DERISO, 1999). As análises de avaliação de estoques são úteis para os tomadores de decisão e a gestão da pesca. No entanto, o tamanho do estoque ou abundância é muitas vezes uma variável latente, no sentido de que não pode ser medida diretamente. Portanto, na maioria das análises de avaliação de estoques, a abundância é substituída por

índices, que supostamente refletem abundância (SPARRE e VENEMA, 1987; HILBORN e WALTERS, 1992; KING, 1985).

Muitas vezes, dados da taxa denominada de captura por unidade de esforço (CPUE) são utilizados para calcular índices de abundância relativa de recursos de pesca altamente migratórios, como atum e espécies afins. No entanto, a taxa de captura (ou CPUE) muda em resposta à abundância, mas também a vários fatores como a estratégia de pescadores ou tecnologia de pesca. As análises de CPUE para minimizar ou eliminar os efeitos de diferentes fatores e estimar uma série temporal de índices de abundância relativa são normalmente chamadas de "padronização". Com frequência a padronização da CPUE é baseada em estimativas dos parâmetros para o fator "ano", calculadas com Modelos Lineares Generalizados (GLM). O principal objetivo da incorporação do fator ano no modelo é detectar tendências ao longo do tempo na abundância. A aplicação do GLM possibilita a separação dos efeitos de muitos fatores que causam variabilidade na CPUE, permitindo a remoção dos mesmos, e a produção de séries de CPUE padronizadas, que refletiriam as variações temporais (*e.g*. anuais ou mensais) da abundância (MAUNDER e PUNT, 2004).

No processo de padronização da CPUE, uma questão metodológica importante que ainda demanda investigação é a forma como o fator “ano” é incorporado na modelagem. Se ano é considerado apenas como efeito fixo principal (abordagem A), as estimativas de CPUE padronizadas são extraídas diretamente de maneira simples. No entanto, se o ano é incluído no modelo como efeito fixo principal, mas também em interações de efeito fixo (abordagem B), as estimativas de CPUE padronizadas são mais complexas. Neste caso é necessária alguma suposição referente a pesos para calcular uma média ponderada sobre as interações (*e.g.* Quinn II *et al.*, 1982, Campbell, 2004). No entanto, se o ano é incluído como efeito fixo principal e também nas interações de efeito aleatório (abordagem C), os cálculos são novamente simplificados, porque as interações não são necessárias para estimar a CPUE padronizada. Inicialmente a abordagem A era bastante utilizada, com o passar do tempo a abordagem B substituiu a abordagem A e, atualmente, a abordagem C tem sido cada vez mais empregada. De acordo com Sahai e Ageel (2000), se um efeito deve ser considerado fixo ou aleatório dependerá da forma como os tratamentos experimentais (níveis de um fator) são selecionados e o tipo de inferências que se deseja fazer a partir da análise. Searle *et al.* (1992) comentam que as respostas claras à pergunta “fixas ou aleatórias” não são necessariamente a norma. Em estudos

experimentais é mais fácil decidir se um efeito principal será considerado fixo ou aleatório, no entanto, em estudos observacionais a decisão entre fixo ou aleatório pode não estar clara, especialmente se a interação analisada inclui um fator de interesse (*e.g.* ano) e outro fator que não é de interesse (*e.g.* trimestre). De Oliveira *et al.* (2015) destacam que a escolha do que deve ser fixo ou aleatório às vezes pode ser uma escolha conceitual. Comenta ainda que os efeitos aleatórios normalmente incluem blocos ou estudos observacionais que são replicados em espaço ou tempo, mas podem também englobar variações entre indivíduos, espécies ou regiões. Em seu estudo, optou por utilizar como efeito aleatório a bandeira da embarcação, pelo fato conceitual que a variabilidade nas taxas de captura do agulhão negro entre os barcos pode depender das características intrínsecas de cada frota de pesca, capitão e tripulação. Nesse contexto, a escolha de uma entre as abordagens A, B ou C pode não ser fácil.

O estudo dessa questão ainda em aberto no âmbito da avaliação de estoques é crucial para os avanços na gestão de recursos pesqueiros. Diante dessa problemática, este trabalho propõe a comparação entre as três abordagens para padronizar as CPUEs com aplicação para o caso espadarte (*Xiphias gladius*) capturado na pesca de atuns e afins no Atlântico Sul.

3.2 Revisão de Literatura

3.2.1 Biologia, características externas e população do espadarte

O *Xiphias gladius*, conhecido popularmente como espadarte, é uma espécie pelágica amplamente distribuída nos mares e oceanos do mundo (AMORIM e ARFELLI, 1984). Ocorre mais intensamente em águas temperadas, sendo também encontrado em águas subtropicais e tropicais (UEYANAGI *et al.*,1970). Apresenta ampla tolerância térmica, ocorrendo de 5º a 27ºC, desde a superfície até profundidades superiores a 600m (NAKAMURA, 1985). O espadarte é sexualmente dimórfico, com as fêmeas crescendo mais rápido e atingindo maiores comprimentos que os machos. São animais que atingem grande porte, havendo registros de exemplares com mais de 500kg de peso total (FOLSOM *et al.*, 1997; NAKAMURA, 1985; ICCAT, 2015).

Quanto às características externas, o espadarte, a partir da cabeça, apresenta corpo sem escamas, alongado e cilíndrico que afunila no pedúnculo caudal, olhos grandes, e arcos branquiais longos. A cabeça é caracterizada pelo prolongamento do rostrum, em forma de uma espada, a qual é longa e achatada, podendo atingir mais de 1/3 do comprimento total do corpo (NAKAMURA, 1985). Apresenta duas barbatanas dorsais bem separadas uma da outra, a primeira com 34 a 45 raios, bem rígida e triangular e a segunda menor com 4 a 5 raios. Possui duas barbatanas anais separadas, a primeira, nitidamente maior, com 3 a 14 raios, e a segunda com 3 ou 4 raios, sendo a posição desta, geralmente, anterior à segunda barbatana dorsal. Um par de barbatanas peitorais, também rígidas, situadas no flanco inferior do corpo, com 16 a 18 raios. A barbatana caudal apresenta-se larga e arqueada (forma de meia lua). Presença de uma quilha mediana no pedúnculo caudal. A linha lateral é ausente na fase adulta, mas visível em exemplares de até 1 metro (PALKO *et al.*, 1981; NAKAMURA, 1985). As colorações das barbatanas dorsais são pretas ou castanho escuras, enquanto que a coloração no dorso e flancos pode variar de azul-escuro a roxo quando vivos, a preto ou castanho depois de 30 minutos após sua morte, mudando gradualmente de tonalidade para castanho claro na parte ventral (Figura 14). Os peixes jovens são geralmente azulados na parte superior (HAZIN, 2006).

Figura 14 – Espadarte (*Xiphias gladius*).

Fonte: http://www.fishbase.org/photos/UploadedBy.php?autoctr=19777&win=uploaded

Quanto à estrutura da população de espadarte para o Oceano Atlântico e o Mar Mediterrâneo, a hipótese atualmente mais aceita é de que existem três estoques distintos, um no Atlântico Norte, outro no Atlântico Sul (separados, aproximadamente,

pela latitude de 5°N) e um no Mar Mediterrâneo. A diferenciação entre as três unidades de estoque é baseada em informações morfológicas e de crescimento por áreas. No ano 2014, as capturas de espadartes no Atlântico Norte e Sul foram, respectivamente, 10.801 t e 9.885 t, totalizando uma produção de 20.686 t. Os países que se destacam com maiores capturas desta espécie no Atlântico Norte são Espanha, Estados Unidos, Canadá, Portugal e Marrocos, enquanto no Atlântico Sul, os países são Espanha, Brasil, Japão, Taiwan e Namíbia (ICCAT, 2015).

3.2.2 Ciclo migratório e alimentação do espadarte

Folsom *et al.* (1997) relata que o comportamento migratório do espadarte não é totalmente compreendido, mas os peixes parecem se mover seguindo uma série de padrões circulares dependendo da estação. Durante os meses de verão eles podem ser encontrados em águas mais frias, situadas em latitudes mais ao norte. Durante os meses de inverno eles são encontrados mais perto do Equador onde as águas são mais quentes. O autor ainda comenta da possibilidade de existência de peixes relativamente residentes em algumas áreas. Em 2013, o Comitê da ICCAT revisou a análise dos movimentos horizontais que evidenciam padrões sazonais com peixes geralmente se movendo para o sul no inverno e retornando às águas temperadas na primavera. Também foram sugeridas áreas mais amplas de mistura entre algumas áreas do leste e oeste. Além das migrações horizontais, a espécie realiza migrações verticais diárias, encontrando-se em águas mais profundas durante o dia e na superfície durante a noite (CAREY e ROBISON, 1981; GUITART, 1964). De acordo com a ICCAT (2015), os resultados obtidos por satélite confirmam o conhecimento anterior referente ao comportamento migratório do espadarte, que estava disponível a partir dos dados da pesca.

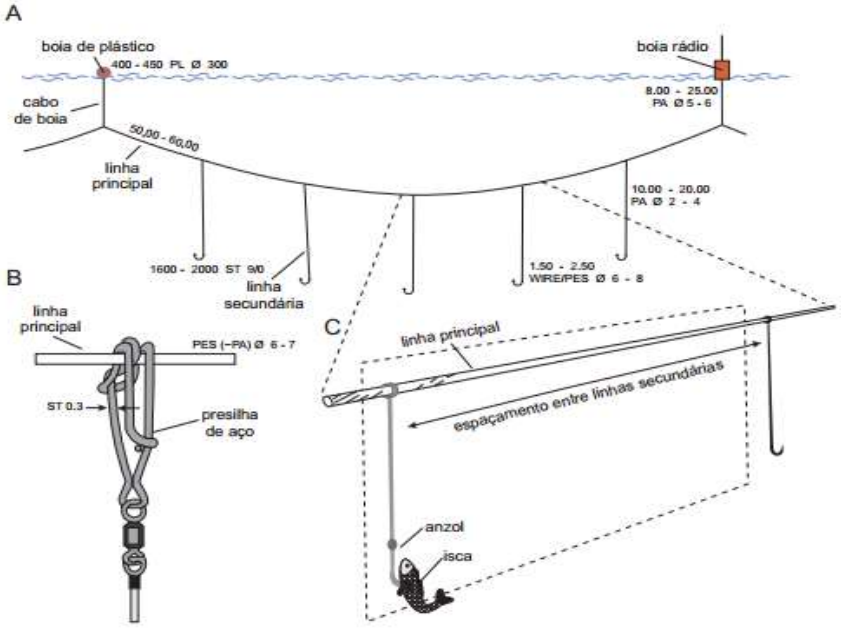
O espadarte se alimenta desde a superfície até camadas mais profundas, numa vasta gama de profundidade. Sua alimentação consiste de uma grande variedade de presas, incluindo peixes pelágicos, peixes de águas profundas e invertebrados (ICCAT, 2015). Os adultos com grande porte realizam frequentemente viagens de alimentação para o fundo, onde as temperaturas podem ser de 5° a 10°C e se alimentam de peixes demersais (espécies de peixes que vivem a maior parte do tempo em contato com o fundo do mar, quer seja arenoso ou rochoso) (NAKAMURA, 1985). Conforme observações de Abid e Idrissi (2006) a dieta do *X. gladius* varia

consideravelmente entre habitats e estações, sendo resultado principalmente de variações na dinâmica populacional das diferentes espécies de presas (LETELIER *et al.*, 2009).

3.2.3 Pesca com espinhel pelágico

A pesca com espinhel no Brasil foi implantada no ano de 1956 com a chegada de navios japoneses no porto de Recife, que vieram objetivando desenvolver a atividade no Atlântico Tropical (MENESES DE LIMA *et al.*, 2000). O espinhel do tipo japonês, desenvolvido para captura de atuns, caracterizava-se por sua linha principal de multifilamentos e pelo sistema de rolos em que era recolhido, armazenado e lançado. Esse tipo de aparelho de pesca perdurou entre as décadas de 1960 e 1980. Na atualidade, o espinhel de superfície (*long-line*), de uso comum no Brasil, é o de tipo americano, que consta de uma linha principal de monofilamento de poliamida, com 4 mm de diâmetro e que pode variar entre 80 e 200 km de comprimento. O espinhel de superfície é suspenso por boias com cabos de poliamida com 3 mm de diâmetro possuem entre 10 e 20 m de comprimento (Figura 15).

Figura 15 – Estrutura do espinhel de superfície de deriva: (A) samburá (composição entre boias subsequentes) de um espinhel pelágico de superfície; (B) sistema de fixação das linhas secundárias; (C) unidade básica de todo espinhel.



Fonte: CALAZANS, 2011.

Um número variado de linhas secundárias é disposto entre duas boias subsequentes, o que constitui a unidade básica do espinhel, o samburá (Figura 15A). As linhas secundárias são conectadas à principal, por meio de presilhas de aço, a intervalos de distância relativamente constantes de 50 a 60 m (Figura 15B e Figura 15C). Essas linhas secundárias, também de poliamidas, possuem 1,8 mm de diâmetro e de 10 a 20 m de comprimento; são usualmente equipadas com ponteiras adicionais, compostas por destorcedor de 20 g, linha de poliamida de 1,8 mm de diâmetro e 2,5 m de comprimento de anzol tipo J N^0. 9/0 (CALAZANS, 2011). Quanto às características operacionais na pesca com espinhel de superfície, destaca-se: 1) o uso de bastões luminosos (*light sticks*), que são colocados próximos aos anzóis, com a finalidade de atrair as espécies-alvos, aumentando assim a probabilidade de captura; 2) as iscas utilizadas, sendo as mais comuns lulas, bonitos e sardinhas; 3) os horários de lançamento do espinhel que, para atuns, é durante o dia e, para espadartes e tubarões, à noite (DIAS-NETO e DIAS, 2015; CALAZANS, 2011).

O espinhel horizontal de superfície é o equipamento mais utilizado na pesca oceânica e destina-se principalmente a captura de atuns, espadarte e tubarões. Uma das vantagens dessa modalidade é o reduzido impacto no ambiente físico. Apesar de ser relativamente seletivo, podem ser capturadas muitas espécies de peixes que não são alvo da pescaria e, em algumas situações aves, tartarugas e mamíferos, motivo pelo qual a modalidade pesqueira é parte central de grande polêmica na atualidade.

No início da pescaria brasileira de espinhel pelágico, o espadarte era capturado ocasionalmente como fauna acompanhante por embarcações que tinham como espécies-alvo albacoras ou tubarões. Os primeiros experimentos de pesca dirigidos à captura do espadarte foram realizados em 1980/81, ainda com o uso do espinhel de nylon multifilamento (AMORIM e ARFELLI, 1984). Em meados da década de 1990, em virtude da crescente demanda no mercado internacional, ocorreu o direcionamento da pesca para captura de espadartes de forma efetiva e bem estabelecida. Desde então surgiram diversas mudanças no aparelho e na estratégia de pesca, tais como: troca do tradicional nylon multifilamento para o nylon monofilamento; utilização de atrativos luminosos descartáveis (*light stick*) e lulas como isca, lançamento do espinhel durante a noite (devido a permanência do espadarte em águas mais rasas para alimentação noturna); e lançamento do espinhel de forma a propiciar que o anzol permaneça entre 30 e 200 m de profundidade (ARFELLI, 1996; AMORIM *et al.*, 1997; MENESES DE LIMA *et al.*, 2000).

3.2.4 Modelos lineares generalizados mistos

Os modelos lineares generalizados mistos (GLMM) são utilizados para descrever um conjunto de dados cuja estrutura de tratamentos envolve tanto fatores fixos como aleatórios, independentemente da média e do erro experimental. É uma extensão do GLM pela inclusão de efeitos aleatórios no preditor linear (BATES, 2010). Uma vantagem do GLMM é que se pode modelar não somente a média dos dados, mas também sua variância, com a inclusão de termos adicionais na relação entre a média da resposta e as variáveis explicativas, na forma de quantidades aleatórias ao invés de parâmetros. Os GLMMs são úteis para variáveis respostas que não são normalmente distribuídas, permitindo ligações não-lineares entre a média da variável resposta e os preditores, sendo capaz de modelar a sobredispersão e a correlação com a incorporação de efeitos aleatórios (BRESLOW e CLAYTON, 1993; MCCULLOCH e SEARLE, 2001).

3.2.4.1 Definição do modelo

Os componentes do GLMM são semelhantes ao do GLM (visto na seção 2.2.6), com as três partes:

a) preditor linear (η) – usado para modelar a relação entre os efeitos fixos e aleatórios, dado por:

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{Zu} \tag{43}$$

em que $\boldsymbol{X}$ é a matriz de incidência dos efeitos fixos de ordem $N \times p$; $\boldsymbol{Z}$ é a matriz de incidência dos efeitos aleatórios $N \times q$; $\boldsymbol{\beta}$ é o vetor de efeitos fixos desconhecidos de ordem $p \times 1$ e $\boldsymbol{u}$ é o vetor de efeitos aleatórios desconhecidos de ordem $q \times 1$.

A relação entre o preditor linear e o vetor de observações $\boldsymbol{y}$ é modelada como:

$$\boldsymbol{y}|\boldsymbol{u} \sim (g(\boldsymbol{\eta}), \boldsymbol{R}) \tag{44}$$

em que a notação, $\boldsymbol{y}|\boldsymbol{u} \sim (g(\boldsymbol{\eta}), R)$, especifica que a distribuição condicional de y dado u tem média, $g(\boldsymbol{\eta})$, e variância e covariância, $\boldsymbol{R}$.

b) função de ligação (g) – usada para modelar a relação entre o preditor linear e a média condicional da característica observada.

$$\boldsymbol{\mu_i} = g(\boldsymbol{\eta_i}) \tag{45}$$

A seleção de funções de ligação normalmente está associada à distribuição dos erros. Para a distribuição Poisson, a função de ligação é a função $g(\eta_i) = e^{\eta_i}$.

c) função de variância – utilizada para modelar a variabilidade residual.

$$v(\boldsymbol{\mu}_i, \emptyset) \tag{46}$$

Normalmente para um modelo linear generalizado, a variabilidade residual pode surgir devido à distribuição amostral da variável resposta (que é resolvida através das funções de variância) ou devido à sobredispersão. A variabilidade associada à sobredispersão pode ser resolvida através de várias maneiras: i) dimensionando a variabilidade residual como $var(y_i|\boldsymbol{u}) = \emptyset v(\mu_i)$, onde $\emptyset$ é o parâmetro de sobredispersão; ii) adicionando um efeito aleatório adicional $e_i \sim N(0, \emptyset)$ ao preditor linear para cada observação; iii) selecionando outra distribuição (*e.g.* substituir a distribuição Poisson de um parâmetro (μ) pela distribuição binomial negativa de dois parâmetros $(\mu, \emptyset)$) (KACHMAN, 2000).

3.2.4.2 Estimação por máxima verossimilhança

A estimação dos parâmetros do modelo linear generalizado misto pode ser feita através do método da máxima verossimilhança hierárquica proposto por Lee e Nelder (1996). Considere a distribuição condicional de $\boldsymbol{Y}$ dado $\boldsymbol{u}$, sendo $\boldsymbol{Y}$ o vetor de respostas assumido como se constituído de elementos condicionalmente independentes (não necessariamente) com densidade pertencente à família exponencial.

$$f_{Y_i|\boldsymbol{u}}(y_i|\boldsymbol{u}) = \exp\left\{\frac{w_i}{\emptyset}[y_i\theta_i - b(\theta_i)] + c(y_i; \emptyset)\right\} \tag{47}$$

Sabe-se que

$$E(Y_i) = b'(\theta_i) = \mu_i \tag{48}$$

$$V(Y_i) = a_i(\emptyset)b''(\theta_i) = \frac{\emptyset}{w_i}b''(\theta_i) = \frac{\emptyset}{w_i}V(\mu_i) \tag{49}$$

em que $\mu_i = \frac{\partial b(\theta_i)}{\partial \theta_i}$. Portanto, $E(Y_i|\boldsymbol{u}) = \mu_i$ e $g(\mu_i) = \boldsymbol{x_i'}\beta + z_i'\boldsymbol{u}$. Atribuindo-se ainda, uma distribuição para os efeitos aleatórios $\boldsymbol{u} \sim f_U(u)$, tem-se que a fórmula para verossimilhança pode ser escrita como:

$$L = \int \prod_i f_{Y_i|\boldsymbol{u}}(y_i|\boldsymbol{u})f_U(u)du = \prod_i \int f(\boldsymbol{y};u)du = \prod_{i=1}^{n} f_Y(\boldsymbol{y}) \tag{50}$$

em que a integração é sobre a distribuição de $\boldsymbol{u}$, de dimensões $q \times 1$.

3.2.4.2.1 Equações de verossimilhança para parâmetros de efeitos fixos

Reescrevendo a equação de verossimilhança (50) em uma forma mais simples, tem-se:

$$l = \log f_{\boldsymbol{Y}}(\boldsymbol{y}) \tag{51}$$

Dessa forma,

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = \frac{\partial}{\partial \boldsymbol{\beta}}\left[log \int f_{\mathrm{Y}|\boldsymbol{u}}(\boldsymbol{y}|\boldsymbol{u})f_U(u)du\right] = \frac{1}{f_Y(\boldsymbol{y})}\int \left[\frac{\partial}{\partial \beta} f_{\mathrm{Y}|\boldsymbol{u}}(\boldsymbol{y}|\boldsymbol{u})\right] f_U(u)du \tag{52}$$

pois $f_U(\boldsymbol{u})$ não envolve $\boldsymbol{\beta}$. Como $\frac{\partial}{\partial \beta} f_{\mathrm{Y}|\boldsymbol{u}}(\boldsymbol{y}|\boldsymbol{u}) = \left[\frac{\partial}{\partial \beta} log\, f_{\mathrm{Y}|\boldsymbol{u}}(\boldsymbol{y}|\boldsymbol{u})\right] f_{\mathrm{Y}|U}(\boldsymbol{y}|\boldsymbol{u})$, tem-se que:

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = \int \left[\frac{\partial l}{\partial \boldsymbol{\beta}} log\, f_{\mathrm{Y}|\boldsymbol{u}}(\boldsymbol{y}|\boldsymbol{u})\right] f_{\mathrm{U}|\mathrm{Y}}(\boldsymbol{u}|\boldsymbol{y})du \tag{53}$$

Mas,

$$\frac{\partial l}{\partial \beta} = \boldsymbol{X}'\boldsymbol{W}\Delta(\boldsymbol{y} - \boldsymbol{\mu}) \tag{54}$$

em que $\boldsymbol{W} = diag\{W_i\} = diag\left[a_i(\emptyset)V(\mu_i)\frac{\partial \eta_i}{\partial \mu_i}\right]^2$ e $\Delta = diag\left\{\frac{\partial \eta_i}{\partial \mu_i}\right\}$

Logo,

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = \int \boldsymbol{X}'\boldsymbol{W}^*(\boldsymbol{y} - \boldsymbol{\mu})\ f_{\mathrm{U|Y}}(u|y)du \tag{55}$$

em que $\boldsymbol{W}^* = diag\left[a(\emptyset)V(\mu_i)\frac{\partial \eta}{\partial \mu}\right]^{-1}$. Portanto,

$$\frac{\partial l}{\partial \beta} = \boldsymbol{X}'\boldsymbol{y}E[\boldsymbol{W}^*|\boldsymbol{y}] - \boldsymbol{X}'E[\boldsymbol{W}^*\boldsymbol{\mu}|\boldsymbol{y}] \tag{56}$$

$$\boldsymbol{X}'\boldsymbol{y}E[\boldsymbol{W}^*|\boldsymbol{y}] = \boldsymbol{X}'E[\boldsymbol{W}^*\boldsymbol{\mu}|\boldsymbol{y}] \tag{57}$$

3.2.4.2.2 Equações de verossimilhança para parâmetros de efeitos aleatórios

Seja φ denotando os parâmetros dos efeitos aleatórios. Fazendo-se

$$\frac{\partial l}{\partial \varphi} = \frac{1}{f(y)}\frac{\partial f(y)}{\partial \varphi} = \int f_{\mathrm{U|y}}(\boldsymbol{u}|\boldsymbol{y})\left[\frac{\partial}{\partial \varphi} log\ f_U(\boldsymbol{u})\right]du \tag{58}$$

$$= E\left[\frac{\partial}{\partial \varphi} log\ f_U(\boldsymbol{u})|\boldsymbol{y}\right] \tag{59}$$

que necessita da especificação da distribuição dos efeitos aleatórios para sua simplificação.

Um algoritmo iterativo para calcular as estimativas de máxima verossimilhança via “TMB” (*Template Model Builder*) foi proposta por Kristensen *et al.* (2015). O TMB calcula estimativas de ambos os parâmetros e efeitos aleatórios usando a aproximação de Laplace, sendo descrito da seguinte forma:

Seja $f(\boldsymbol{u}, \boldsymbol{\theta})$ a log-verossimilhança conjunta negativa dos dados e os efeitos aleatórios. Esta depende de efeitos aleatórios desconhecidos $\boldsymbol{u} \epsilon R^n$ e parâmetros $\boldsymbol{\theta} \epsilon R^m$, onde R está em um espaço n-dimensional e m-dimensional, respectivamente. Por meio do pacote TMB é possível obter estimativas de máxima verossimilhança e

cálculos de incerteza para $\boldsymbol{u}$ e $\boldsymbol{\theta}$. A estimativa de máxima verossimilhança para $\boldsymbol{\theta}$ maximiza

$$L(\boldsymbol{\theta}) = \int_{R^n} \exp\big(-f(\boldsymbol{u}, \boldsymbol{\theta})\big)\, du \tag{60}$$

em relação a $\boldsymbol{\theta}$. Note que os efeitos aletórios $\boldsymbol{u}$ foram integrados e a verossimilhança marginal $L(\boldsymbol{\theta})$ é a verossimilhança dos dados como uma função dos parâmetros. Considere $u(\boldsymbol{\theta})$ para denotar o minimizador de $f(\boldsymbol{u}, \boldsymbol{\theta})$ em relação a u, isto é,

$$\hat{u}(\boldsymbol{\theta}) = \arg\min_{u} f(\boldsymbol{u}, \boldsymbol{\theta}) \tag{61}$$

Considere $H(\theta)$ para denotar a Hessian de $f(\boldsymbol{u}, \boldsymbol{\theta})$ em relação a u e estimado em $\hat{u}(\boldsymbol{\theta})$, isto é:

$$H(\boldsymbol{\theta}) = f''_{uu}(\hat{u}(\boldsymbol{\theta}), \boldsymbol{\theta}) \tag{62}$$

A aproximação de Laplace para a verossimilhança marginal $L(\boldsymbol{\theta})$ é

$$L^*(\boldsymbol{\theta}) = \sqrt{2\pi}^{\,n} \det(H(\boldsymbol{\theta}))^{-\frac{1}{2}} \exp(-f(\hat{u}, \boldsymbol{\theta})) \tag{63}$$

Esta aproximação é amplamente aplicável, incluindo modelos que variam de modelos com efeitos mistos não-lineares a modelos complexos de espaço-tempo.

3.2.4.3 Testes de hipóteses

Para os modelos lineares generalizados mistos, os testes de hipóteses são aproximados e permitem avaliar tanto os efeitos fixos como os aleatórios em modelos encaixados, ou seja, cada modelo inclui mais termos que os anteriores, e têm a mesma função de ligação e distribuição. Dentre os diversos testes encontrados na literatura o teste da razão de verossimilhanças é o mais frequente. No entanto, Pinheiro e Bates (2000) comentam que para testar efeitos fixos usando o TRV, deve-se condicionar a especificação dos efeitos fixos às estimativas das variâncias e covariâncias dos efeitos aleatórios, visto que o teste tende a ser conservativo por usar como referência a distribuição qui-quadrado. Este teste condicional é dado pelos testes F e t usuais, como definidos nos modelos lineares, sendo condicionados à

$$\hat{\sigma}_R^2(\boldsymbol{\theta}) = s^2 = \frac{SQR}{N-p} \tag{64}$$

em que $\boldsymbol{\theta}$ refere-se aos parâmetros envolvidos nos efeitos fixos, SQR é a soma de quadrados do resíduo, N é a soma dos n_i e p a quantidade de parâmetros.

3.3 Dados e Metodologia

3.3.1 Dados de esforço de pesca e captura de espadarte no Atlântico Sul

O conjunto de informações utilizadas é a contida no "Banco Nacional de Atuns e Afins (BNDA)" do Comitê Científico Brasileiro de Atum e Afins do governo brasileiro. Foi analisado o espadarte (*Xiphias gladius*) capturado no Atlântico Sul pela frota brasileira entre 1990 e 2012. A frota brasileira inclui navios nacionais (BRA), mas também navios arrendados de mais de quinze países. Para as análises foram utilizadas as informações de barcos arrendados da Espanha (BRA-ESP), Honduras (BRA-HND), Japão (BRA-JPN), Panamá (BRA-PAN), China-Taipei (BRA-TAI) e São Vicente & Granadinas (BRA-VCT) (Tabela 11), para os quais o tamanho amostral é relativamente elevado. Foram também descartados registros com dados de captura ou esforço ausentes, ou com erros não amostrais (por exemplo, lances de pesca localizados em terra ou proximidades). Ao final apenas 34145 lances de pesca foram mantidos para a análise relativa à padronização da CPUE, e eventual estimativa de índice relativo de abundância.

Tabela 11 – Número de registros de lances de pesca por bandeira retidos para análise do espadarte capturado no Atlântico Sul no período de 1990 a 2012.

Bandeira do barco	Nomenclatura	Frequência
Brasil	BRA	5547
Brasil – Espanha	BRA-ESP	12559
Brasil – Honduras	BRA-HND	1079
Brasil – Japão	BRA-JPN	821
Brasil – Panamá	BRA-PAN	3567
Brasil – China Taipei	BRA-TAI	4704
Brasil – São Vicente & Granadinas	BRA-VCT	5868

Fonte: Elaborada pelo autor.

3.3.2 Modelos de regressão e variáveis utilizadas

As estimativas das CPUEs padronizadas incluindo unicamente o “ano” como efeito fixo principal (opção A), ou como efeito fixo principal e também em interações de efeito fixo (opção B), foram obtidas com o uso de modelos lineares generalizados (GLM). Para obter estimativas da CPUE padronizada com ano incluído como efeito fixo principal e também em interações aleatórias (opção C) foi utilizado o modelo linear generalizado misto (GLMM).

As variáveis utilizadas para obter as estimativas das CPUEs padronizadas foram: nº peixes capturados (S), bandeira do barco (F), ano (Y), trimestre (T), nº de samburás no espinhel (H), nº de anzóis por samburá e área de pesca (A): Norte (ao norte de 10ºS), Centro (entre 10ºS e 25ºS) e Sul (ao sul de 25ºS) (Figura 16). Estas áreas de pesca foram selecionadas com base na distribuição espacial dos registros, prezando por balanço com tamanhos amostrais relativamente elevados nos três setores. A unidade de esforço considerada na análise foi o número de anzóis, que foi utilizado como offset. Para modelagem da variável resposta (nº peixes capturados) com GLM e GLMM, foram testadas as distribuição Poisson e Binomial Negativa.

As estimativas dos componentes da variância foram obtidas pelo método da máxima verossimilhança (ML) (*e.g.* HARTLEY e RAO, 1967), utilizando a aproximação de Laplace (*e.g.* MAGNUSSON *et al.*, 2017). Para o GLMM, foi utilizada a matriz de variâncias e covariâncias não-estruturada (UN). A forma UN é heterogênea, com todas as variâncias e covariâncias desiguais.

3.3.3 Seleção de modelos e diagnósticos

A escolha da ordem de entrada das variáveis explicativas no modelo, foi realizada ajustando um modelo para cada variável explicativa separadamente e utilizando a deviância para classificação (NELDER e WEDDERBURN, 1972). Para as abordagens B e C, além dos fatores principais, foram consideradas interações de primeira ordem entre as variáveis. A partir de então a seleção de variáveis explicativas e interações nos modelos finais foi realizada com uma abordagem backward com base no AIC (AKAIKE, 1974), e no BIC (SCHWARZ, 1978). Para examinar se os modelos diferem do melhor modelo por um parâmetro e têm essencialmente os mesmos

valores da log-verossimilhança maximizada foi utilizado o Delta AIC (BURNHAM e ANDERSON, 2002).

Para avaliar o ajuste do modelo selecionado e influências dos registros individuais, foram utilizados gráficos de diagnóstico de resíduos e cálculos baseados na diagonal da matriz de projeção (*e.g.*, distâncias de Cook, alavancagem) (*e.g.* VENABLES e RIPLEY, 2002). Os testes de Kolmogorov-Smirnov e Breusch-Pagan (BREUSCH e PAGAN, 1979) também foram empregados para verificar se os resíduos apresentaram distribuição normal e comportamento homocedástico.

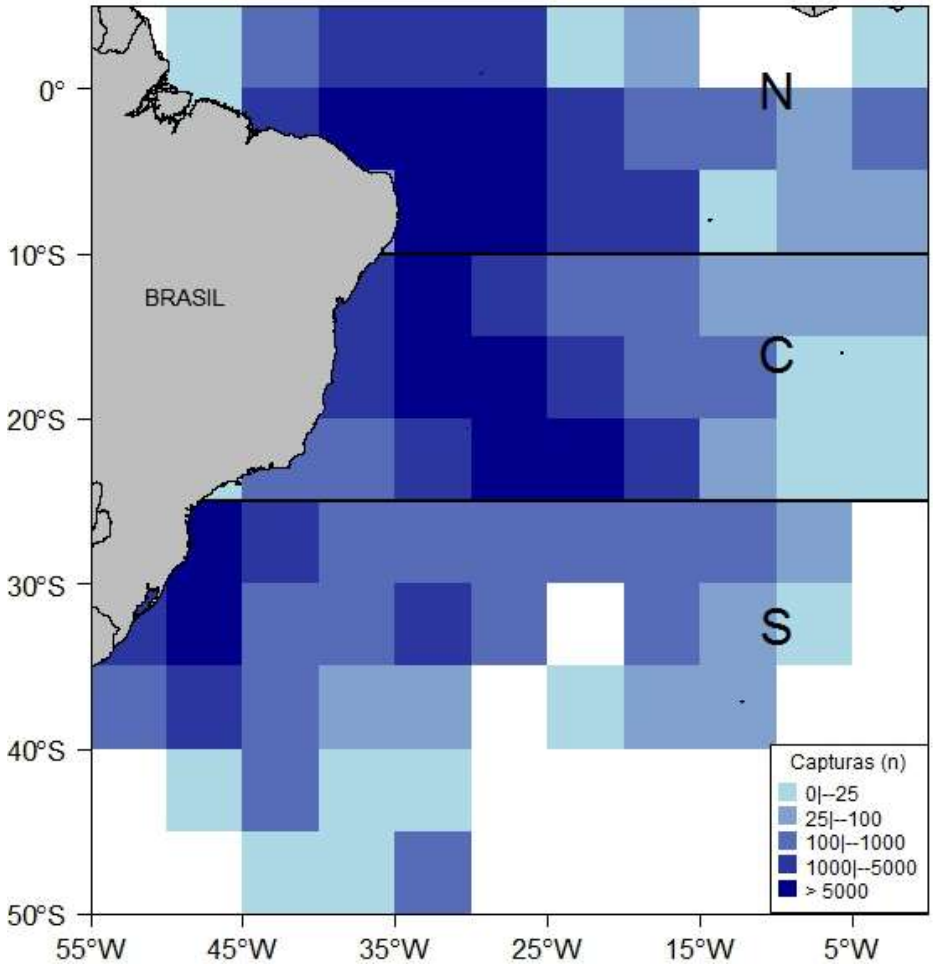
Para comparar as séries temporais das CPUEs padronizadas obtidas com as três opções (A, B e C) foi utilizado o teste não paramétrico de Kruskal-Wallis (KRUSKAL e WALLIS, 1952). Testes posteriores de comparação múltipla de Kruskal-Wallis foram empregados nos casos em que o teste geral indicou a existência de diferenças significativas entre as CPUEs padronizadas. Todas as análises foram realizadas com o programa R (R 3.4.1 Core Team, 2017) e as funções "*glm*" (GLM) e "*glmmTMB*" do package "glmmTMB" (Magnusson *et al*., 2017).

3.4 Resultados e Discussões

3.4.1 Análise Exploratória

As operações da frota brasileira cobrem grande parte do Atlântico Sul entre as longitudes 0º e 55ºW e as latitudes 5ºN e 50ºS (Figura 16). Os valores mostrados no mapa são a soma do número de peixes capturados entre 1990 e 2012, conforme relatado no conjunto de dados. Foram observadas capturas elevadas (> 5000) em todo o Atlântico Sul, mas principalmente nas áreas norte e central.

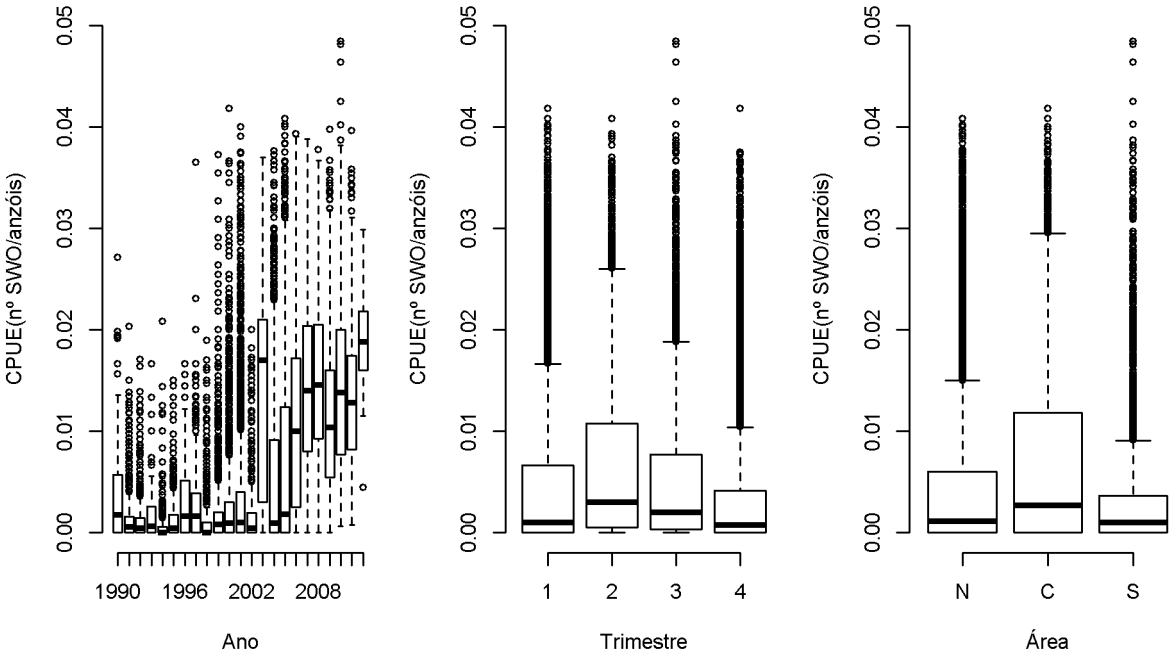
Figura 16 – Distribuição das capturas de *Xiphias gladius* por área durante o período de 1990 a 2012. Áreas de pesca: N – Norte; C – Centro e S – Sul.



Fonte: Elaborada pelo autor.

Diagramas de caixa da CPUE são apresentados na Figura 17. A variabilidade da CPUE em 1998 é elevada. Há também grande aumento das taxas de captura a partir de 2003. Os valores da CPUE no primeiro e no quarto trimestre tendem a ser menores do que os do segundo e terceiro trimestres. Verifica-se ainda que as taxas de captura foram, em geral, mais elevadas na área central. Os resultados obtidos com o teste não paramétrico de Kruskal-Wallis confirmam que as CPUEs de espadarte foram estatisticamente diferentes nas áreas e trimestres, considerando um nível de significância de 5%. Não foi observado diferenças estatísticas a 5% de probabilidade nas CPUEs entre os anos: 2008, 2010 e 2012; 2008, 2010 e 2011; 2003, 2007, 2009; 1990, 1996, 1997 e 2004;1993 e 2000; 1991, 1993, 1995, 1999 e 2002; 1995 e 1992.

Figura 17 – Captura por unidade de esforço (CPUE) por ano, trimestre e área (N - norte; C - centro; S - sul).



Fonte: Elaborada pelo autor.

3.4.2 Modelos lineares generalizados e modelos lineares generalizados mistos

A estrutura dos dez melhores modelos ajustados com distribuições Poisson e Binomial Negativa (cinco para cada distribuição) para calcular as CPUEs padronizadas do espadarte utilizando as três abordagens são apresentadas na Tabela 12. A seleção desses modelos foi baseada em AIC, BIC, na estatística de log-verossimilhança e diagnóstico dos resíduos, conforme descrito na sessão 3.3.3.

Tabela 12 – Estrutura dos cinco melhores modelos ajustados com distribuição Poisson e Binomial Negativa. S – número de espadartes capturados; F - bandeira do barco; Y – ano; A - área de pesca; T - trimestre; H – número de samburás (segmento entre uma boia a outra) no espinhel; N – número de anzóis por samburá. Interações particulares são representadas por dois pontos. Interações de primeira ordem entre cada combinação pareada de duas variáveis são representadas por parênteses e elevado a segunda potência ( )^2. Interações particulares são representadas por dois pontos. O sinal "menos (-)" na frente das interações indica que as mesmas foram eliminadas do modelo.

	Modelos	Distribuição
	Abordagem A	
1	S = F+Y+H+A+T+N	Poisson
2	S = F+Y+H+A+N	Poisson
3	S = F+Y+H+T+N	Poisson
4	S = F+Y+H+A+T	Poisson
5	S = F+Y+A+T+N	Poisson
6	S = F+Y+H+A+T+N	Binomial Negativa
7	S = F+Y+H+T+N	Binomial Negativa
8	S = F+Y+H+A+N	Binomial Negativa
9	S = F+Y+A+T+N	Binomial Negativa
10	S = F+Y+H+A+T	Binomial Negativa
	Abordagem B	
1	S = (F+Y+H+A+T+N)^2-Y:A-Y:T-F:Y	Poisson
2	S = (F+Y+H+A+T+N)^2-F:Y-Y:A-Y:T-H:A	Poisson
3	S = (F+Y+H+A+T+N)^2-F:Y-Y:A-Y:T-T:N	Poisson
4	S = (F+Y+H+A+T+N)^2-F:Y-Y:A-Y:T-H:N	Poisson
5	S = (F+Y+H+A+T+N)^2-F:Y-Y:A-Y:T-A:N	Poisson
6	S = (F+Y+H+A+T+N)^2-F:Y-Y:A-Y:T	Binomial Negativa
7	S = (F+Y+H+A+T+N)^2-F:Y-Y:A-Y:T-H:A	Binomial Negativa
8	S = (F+Y+H+A+T+N)^2-F:Y-Y:A-Y:T-T:N	Binomial Negativa
9	S = (F+Y+H+A+T+N)^2-F:Y-Y:A-Y:T-H:N	Binomial Negativa
10	S = (F+Y+H+A+T+N)^2-F:Y-Y:A-Y:T-A:T	Binomial Negativa
	Abordagem C	
1	S = (F+Y+H+A+T+N)^2-F:Y-Y:A-Y:T	Poisson
2	S = (F+Y+H+A+T+N)^2-F:Y-Y:A-Y:T-F:A	Poisson
3	S = (F+Y+H+A+T+N)^2-F:Y-Y:A-Y:T-F:H	Poisson
4	S = (F+Y+H+A+T+N)^2-F:Y-Y:A-Y:T-T:A	Poisson
5	S = (F+Y+H+A+T+N)^2-F:Y-Y:A-Y:T-H:A	Poisson
6	S = (F+Y+H+A+T+N)^2-F:Y-Y:A-Y:T	Binomial Negativa
7	S = (F+Y+H+A+T+N)^2-F:Y-Y:A-Y:T-T:A	Binomial Negativa
8	S = (F+Y+H+A+T+N)^2-F:Y-Y:A-Y:T-F:H	Binomial Negativa
9	S = (F+Y+H+A+T+N)^2-F:Y-Y:A-Y:T-F:A	Binomial Negativa
10	S = (F+Y+H+A+T+N)^2-F:Y-Y:A-Y:T-H:A	Binomial Negativa

Fonte: Elaborada pelo autor.
Nota: *Abordagem A* - Modelo com ano apenas como efeito fixo; *Abordagem B* - Modelo com ano como efeito fixo e também em interação de efeito fixo; *Abordagem C* - Modelo com ano como efeito fixo e também em interação de efeito aleatório.

O modelo de Poisson, candidato natural em se tratando de dados de contagem, foi o pior para todas as abordagens, quando consideradas as estatísticas AIC, BIC e log-verossimilhança (Tabela 13). Além disso, a deviância residual foi muito superior aos graus de liberdade residual, indicando que o modelo não tem ajuste satisfatório pelo teste qui-quadrado ($p<0,001$). Modelos com resposta binomial negativa foram utilizados para corrigir problemas de sobredispersão. Dentre os modelos ajustados para a Abordagem A, o modelo 6 (S =F+Y+A+T+N) (Tabela 12), foi o modelo mais parcimonioso e que melhor representou a padronização da CPUE do *Xiphias gladius* pescado pela frota brasileira no período de 1990 a 2012 (Tabela 13).

Para a abordagem B, que considera a inclusão do ano (Y) como efeito fixo principal e também em interações de efeitos fixos, o modelo com menor valor nas estatísticas AIC e log-verossimilhança foi o modelo 6 (S = (F+Y+H+A+T+N)^2-F:Y-Y:A-Y:T) com distribuição binomial negativa (Tabela 13). No entanto, o delta AIC entre o modelo 6 e modelo 7 (S = (F+Y+H+A+T+N)^2-F:Y-Y:A-Y:T-H:A) foi um, situando-se dentro do intervalo 0 e 2. Como sugerido por Burnham e Anderson (2002), neste caso, os modelos devem ser examinados para ver se eles diferem do melhor modelo por um parâmetro e têm essencialmente os mesmos valores da log-verossimilhança maximizada. Sendo assim, foi utilizado o teste da razão de verossimilhança para o modelo 6 (com menor valor AIC) (Tabela 14) e verificou-se que a interação entre as variáveis número de samburás (H) e área de pesca (A) não foi estatisticamente significativa ($p>0,05$) para explicar as variações na CPUE padronizada, podendo ser removida do modelo. A remoção dessa interação resulta no modelo 7 com distribuição binomial negativa.

Os modelos ajustados com a abordagem C, que considera a inclusão do ano (Y) como efeito fixo principal e também em interações de efeitos aleatórios foram os que apresentaram menores valores nas estatísticas de log-verossimilhança, AIC e BIC. Dentre os modelos selecionados utilizando esta abordagem, o modelo 6 (S = (F+Y+H+A+T+N)^2-F:Y-Y:A-Y:T) com resposta binomial negativa (Tabela 12) foi o que melhor descreveu os dados (Tabela 13).

Tabela 13 – Comparativo dos modelos ajustados para padronização da captura por unidade de esforço do *Xiphias gladius* capturado pela frota brasileira de 1990 a 2012. Modelos de 1 a 5 foram ajustados com distribuição Poisson; modelos de 6 a 10 foram ajustados com distribuição Binomial negativa. GL Residual – graus de liberdade residual; LogLik – estatística de log-verossimilhança; BIC – critério de informação Bayesiano; AIC – critério de informação de Akaike; Delta AIC – diferença entre o AIC do modelo que está sendo observado e o modelo com menor AIC.

Modelos	Deviância Residual	GL Residual	logLik	BIC	AIC	Delta AIC	Theta
			Abordagem A				
1	163787	34109	-124722	249820	249516	0	-
2	164100	34112	-124879	250102	249823	307	-
3	164310	34111	-124984	250322	250036	520	-
4	167741	34110	-126699	253764	253469	3953	-
5	168019	34110	-126839	254042	253747	4231	-
6	38583	34109	-86526	173439	173126	0	1,00
7	38566	34111	-86534	173435	173140	14	1,00
8	38583	34112	-86539	173434	173147	21	1,00
9	38607	34110	-86995	174366	174062	936	0,96
10	38561	34110	-87085	174544	174241	1115	0,95
			Abordagem B				
1	135296	34007	-110477	222394	221230	0	-
2	135357	34009	-110507	222435	221287	57	-
3	135409	34010	-110533	222476	221337	107	-
4	135410	34008	-110534	222498	221342	112	-
5	135756	34009	-110707	222833	221685	455	-
6	38351	34007	-82648	166747	165574	0	1,42
7	38348	34009	-82651	166731	165575	1	1,42
8	38362	34010	-82666	166751	165604	30	1,42
9	38370	34008	-82690	166820	165655	81	1,41
10	38309	34013	-82727	166842	165719	145	1,41
			Abordagem C				
1	179354	34097	-89677	179855	179450	0	-
2	179485	34098	-89743	179976	179579	129	-
3	179487	34098	-89743	179977	179581	131	-
4	179518	34098	-89759	180008	179612	162	-
5	179776	34098	-89888	180266	179870	420	-
6	159518	34096	-79759	160029	159616	0	2,55
7	159554	34097	-79777	160056	159650	34	2,54
8	159577	34097	-79788	160078	159673	57	2,54
9	159585	34097	-79793	160086	159681	65	2,54
10	159605	34097	-79803	160106	159701	85	2,53

Fonte: Elaborada pelo autor.

Nota: *Abordagem A* - Modelo com ano apenas como efeito fixo; *Abordagem B* - Modelo com ano como efeito fixo e também em interação de efeito fixo; *Abordagem C* - Modelo com ano como efeito fixo e também em interação de efeito aleatório.

Tabela 14 – Teste da razão de verossimilhança (LR) para o modelo 6 com distribuição binomial negativa e abordagem B. Os fatores são: F – bandeira; Y – ano; A – área; T – trimestre; H – número de samburás (segmento entre uma boia a outra) no espinhel; N – número de anzóis por samburá. GL – graus de liberdade. Interações particulares são representadas por dois pontos.

Variáveis	LR Chisq	GL	Pr(>Chisq)
	Abordagem B		
F	3478,8	6	< 2,2e-16*
Y	5488,8	22	< 2,2e-16*
H	914,3	1	< 2,2e-16*
A	53,4	2	2,6e-12*
T	9,2	3	0,02683*
N	1312,4	1	< 2,2e-16*
F:H	578,0	6	< 2,2e-16*
F:A	365,2	11	< 2,2e-16*
F:T	219,9	18	< 2,2e-16*
F:N	1537,6	6	< 2,2e-16*
Y:H	2003,7	22	< 2,2e-16*
Y:N	487,2	22	< 2,2e-16*
H:A	4,6	2	0,10244
H:T	31,1	3	7,9e-07*
H:N	83,0	1	< 2,2e-16*
A:T	157,7	6	< 2,2e-16*
A:N	68,9	2	1,1e-15*
T:N	35,4	3	1,0e-07*

Fonte: Elaborada pelo autor.
Nota: *Abordagem B* - Modelo com ano como efeito fixo e também em interação de efeito fixo. Nível de significância: p<0,05 (*).

Os resultados dos testes de hipóteses de razão de verossimilhança (abordagens A e B) e do teste de Wald (abordagem C) para os modelos selecionados (Tabela 15) indicam que as variáveis explicativas consideradas são importantes para modelar a variação das taxas de captura, com exceção da área de pesca e do trimestre no GLMM (abordagem C).

Tabela 15 – Análise da deviância dos modelos ajustados. Testes da razão de verossimilhança (LR) foram utilizados para as abordagens A e B, e teste Wald foi utilizado para a abordagem C. Os fatores são: F – bandeira; Y – ano; A – área; T – trimestre; H – número de samburás (segmento entre uma boia a outra) no espinhel; N – número de anzóis por samburá. GL – graus de liberdade. Interações particulares são representadas por dois pontos.

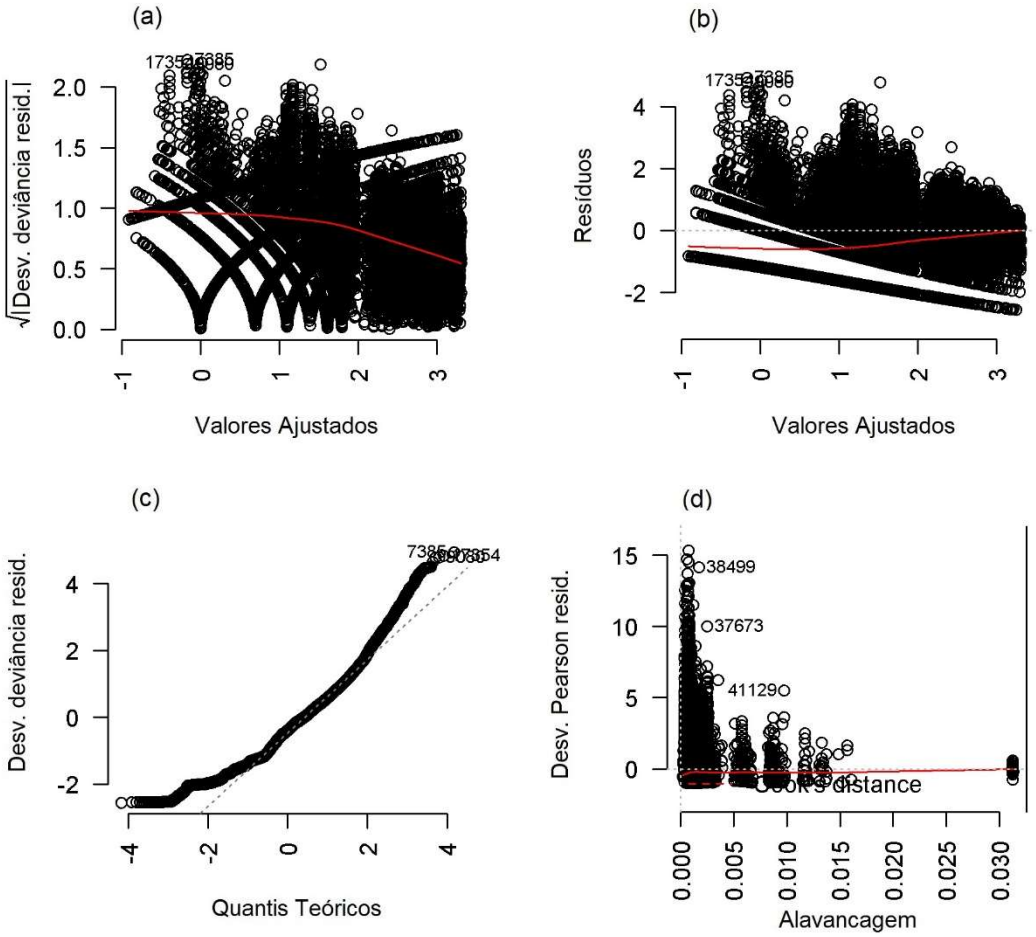
Variáveis	LR Chisq	GL	Pr(>Chisq)
	Abordagem A		
F	6129,1	6	< 2,2e-16*
Y	4487,2	22	< 2,2e-16*
H	1136,5	1	< 2,2e-16*
A	17,6	2	0,00015*
T	26,2	3	8,6e-06*
N	950,8	1	< 2,2e-16*
	Abordagem B		
F	3484,5	6	< 2,2e-16*
Y	5472,7	22	< 2,2e-16*
H	914,0	1	< 2,2e-16*
A	53,4	2	2,6e-12*
T	8,6	3	0,03468*
N	1331,6	1	< 2,2e-16*
F:H	593,3	6	< 2,2e-16*
F:A	498,8	11	< 2,2e-16*
F:T	230,6	18	< 2,2e-16*
F:N	1548,5	6	< 2,2e-16*
Y:H	2029,2	22	< 2,2e-16*
Y:N	495,9	22	< 2,2e-16*
H:T	38,7	3	2,0e-08*
H:N	89,0	1	< 2,2e-16*
A:T	160,5	6	< 2,2e-16*
A:N	82,1	2	< 2,2e-16*
T:N	40,2	3	9,8e-09*
	Abordagem C		
	Chisq	GL	Pr(>Chisq)
Abordagem C			
F	40,4	6	3,8e-07*
Y	222,6	22	< 2,2e-16*
A	0,1	2	0,95620
T	2,67	3	0,44540
H	91,1	1	< 2,2e-16*
N	115,9	1	< 2,2e-16*

Fonte: Elaborada pelo autor.
Nota: *Abordagem A* - Modelo com ano apenas como efeito fixo; *Abordagem B* - Modelo com ano como efeito fixo e também em interação de efeito fixo; *Abordagem C* - Modelo com ano como efeito fixo e também em interação de efeito aleatório. Nível de significância: p<0,05 (*).

Na análise dos resíduos foram identificados problemas nos modelos selecionados para cada abordagem. As distribuições dos resíduos para a abordagem A apresentam comportamentos heterocedásticos (Figura 18(a)), confirmado pelo teste

de Breusch-Pagan (p<0,05). Na Figura 18(b), nota-se que há indícios de viés nas estimativas. Os resíduos não possuem uma distribuição normal, apresentam violações nas caudas (Figura 18(c)), confirmado pelo teste de Kolmogorov-Smirnov (p<0,05). Altos valores de resíduos são associados a um baixo valor de alavancagem (diagonal da matriz de projeção) (Figura 18(d)). As distâncias de Cook foram inferiores a 0,5.

Figura 18 – Diagnóstico dos resíduos (abordagem A).

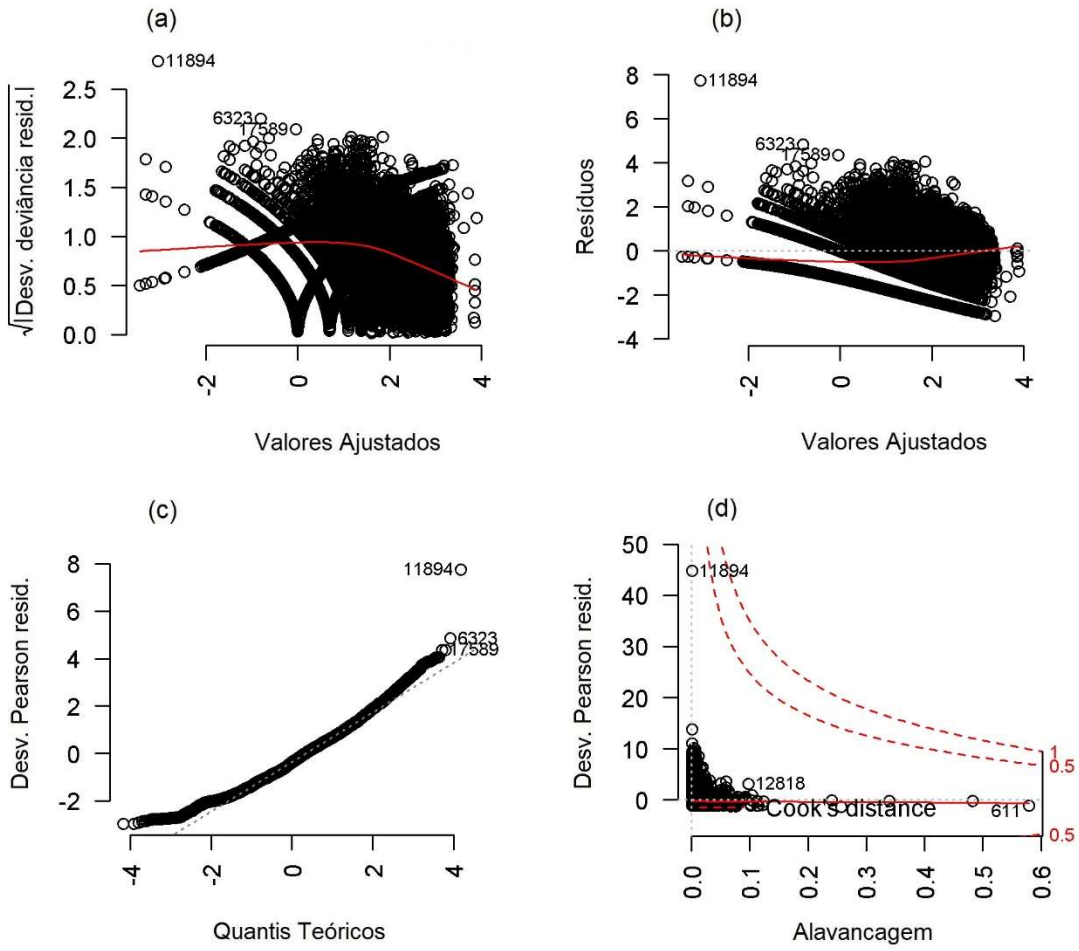


Fonte: Elaborada pelo autor.

No caso da abordagem B, também há violações das suposições do modelo. Os resíduos são heterocedásticos (Figura 19(a)), confirmado pelo teste de Breusch-Pagan (p<0,05). Há indícios de viés nas estimativas (Figura 19(b)), a distribuição de probabilidade dos resíduos não é normal, confirmado pelo teste de Kolmogorov-Smirnov (p<0,05) (Figura 19(c)) e, as observações que correspondem a resíduos

relativamente elevados têm pouco poder de alavancagem (Figura 19(d)). Além disso, nota-se presença de outlier que destoa de forma acentuada do restante dos dados.

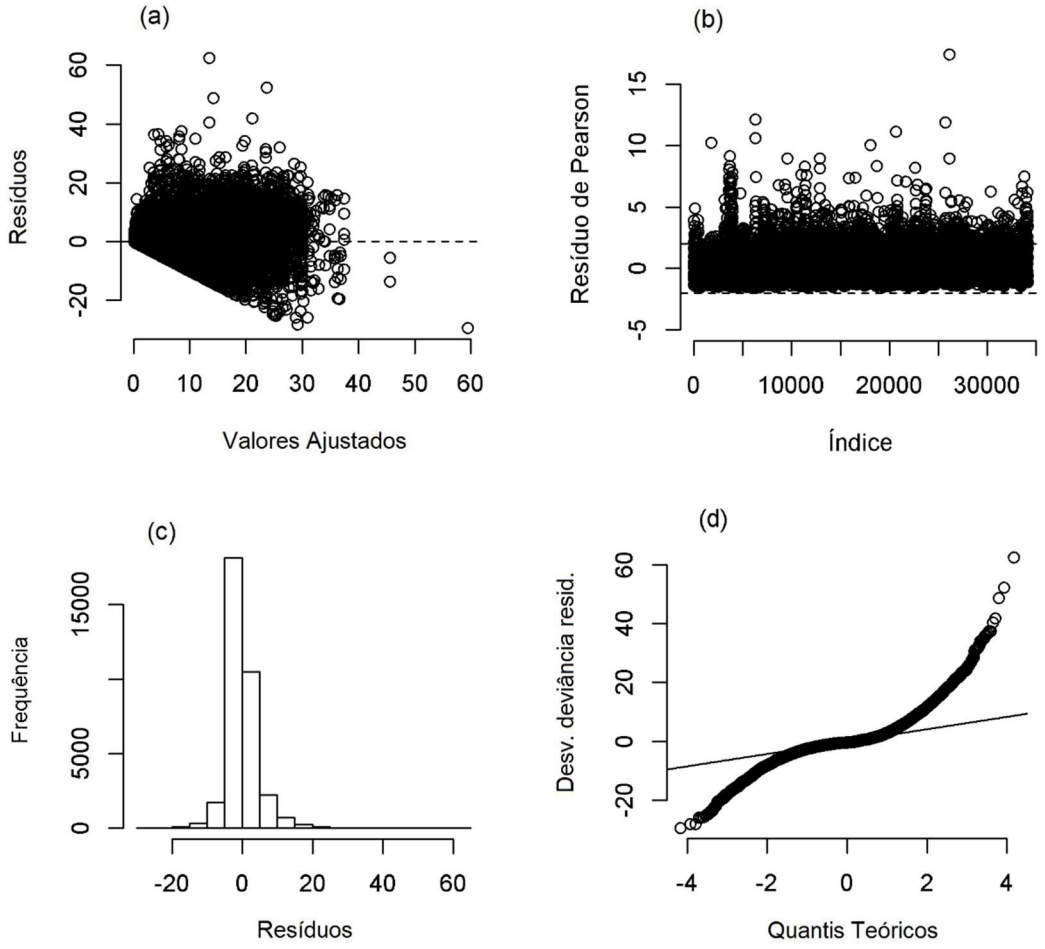
Figura 19 – Diagnóstico dos resíduos (abordagem B).



Fonte: Elaborada pelo autor.

Na abordagem C os problemas são semelhantes. Resíduos heterocedásticos, confirmado pelo teste de Breusch-Pagan ($p<0,05$), há indícios de viés nas estimativas (Figura 20(a)) e a distribuição de probabilidade dos resíduos não é normal, confirmado pelo teste de Kolmogorov-Smirnov ($p<0,05$) (Figura 20(c) e Figura 20(d)). 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.

Figura 20 – Diagnóstico dos resíduos (abordagem C).



Fonte: Elaborada pelo autor.

O resumo das estimativas do modelo ajustado com o ano apenas como efeito principal (abordagem A) estão na Tabela 16. Observe que a maioria das estimativas indica rejeição da hipótese nula (parâmetros iguais a zero). As estimativas dos parâmetros do fator bandeira foram negativas, mas a estimativa do nível BRA-ESP é a exceção. Por conseguinte, existem evidências de que a esperança da CPUE dos navios alugados de Espanha é maior do que para os navios nacionais (nível de base). As esperanças das CPUEs dos barcos nacionais (BRA) e os barcos arrendados da Espanha (BRA-ESP) são por sua vez superiores às dos barcos arrendados de outros países. As estimativas das esperanças de CPUE para 1991, 1995 a 1999 e 2003 em diante foram maiores do que a esperança para o ano 1990 que é o nível de base. Observe também que a esperança das taxas de captura para a região norte foi maior do que para a região sul e não houve diferenças significativas entre as regiões norte

e centro. As estimativas indicam ainda que as esperanças foram maiores para o segundo trimestre se comparadas às calculadas para os demais períodos do ano. A perspectiva para o número de samburás e anzóis no espinhel foi negativa, no sentido de que as esperanças das CPUEs diminuem à medida em que se aumentam as quantidades de samburás e anzóis.

Tabela 16 – Estimativa dos parâmetros (abordagem A). F – bandeira; Y – ano; A –área; T – trimestre; H – número de samburás (segmento entre uma boia a outra) no espinhel; N – número de anzóis por samburá; Significâncias: (***) - 0,001, (**) - 0,01 e (*) - 0,05

	Estimativa	Erro padrão	z valor	Pr(>|z|)
(Intercepto)	-3,887702	0,11141	-34,897	< 2e-16 ***
(F) BRA-ESP	0,182986	0,01978	9,252	< 2e-16 ***
(F) BRA-HND	-0,462589	0,03853	-12,004	< 2e-16 ***
(F) BRA-JPN	-0,633556	0,06139	-10,320	< 2e-16 ***
(F) BRA-PAN	-1,666051	0,03155	-52,808	< 2e-16 ***
(F) BRA-TAI	-0,850941	0,04267	-19,942	< 2e-16 ***
(F) BRA-VCT	-1,495393	0,03893	-38,416	< 2e-16 ***
(Y) 1991	0,344538	0,11472	3,003	0,00267 **
(Y) 1992	0,011306	0,11165	0,101	0,91934
(Y) 1993	-0,069485	0,16623	-0,418	0,67595
(Y) 1994	-0,639092	0,11929	-5,358	8,4e-08
(Y) 1995	0,395724	0,10954	3,612	0,00030 ***
(Y) 1996	-0,487425	0,13382	-3,642	0,00027 ***
(Y) 1997	-0,465835	0,11108	-4,194	2,7e-05 ***
(Y) 1998	-0,517164	0,11380	-4,544	5,5e-06 ***
(Y) 1999	0,025784	0,10549	0,244	0,80690 ***
(Y) 2000	0,010008	0,10457	0,096	0,92375
(Y) 2001	-0,018393	0,10475	-0,176	0,86062
(Y) 2002	-0,013110	0,10843	-0,121	0,90376
(Y) 2003	0,836437	0,14018	5,967	2,4e-09 ***
(Y) 2004	0,472768	0,10782	4,385	1,2e-05 ***
(Y) 2005	0,576923	0,10528	5,480	4,3e-08 ***
(Y) 2006	1,068792	0,10579	10,103	< 2e-16 ***
(Y) 2007	1,325584	0,10880	12,183	< 2e-16 ***
(Y) 2008	1,095038	0,11188	9,787	< 2e-16 ***
(Y) 2009	0,822575	0,10654	7,721	1,2e-14 ***
(Y) 2010	1,065331	0,11417	9,331	< 2e-16 ***
(Y) 2011	0,978360	0,10934	8,948	< 2e-16 ***
(Y) 2012	1,297049	0,20883	6,211	5,3e-10 ***
(A) C	0,012105	0,01538	0,787	0,43135
(A) S	-0,078645	0,02018	-3,898	9,7e-05 ***
(T) 2	0,041837	0,01800	2,324	0,02013 *
(T) 3	0,019231	0,01940	0,991	0,32167
(T) 4	-0,049800	0,01764	-2,824	0,00475 **
(H)	-0,138823	0,00417	-33,293	< 2e-16 ***
(N)	-0,003348	0,00011	-29,742	< 2e-16 ***

Fonte: Elaborada pelo autor.
Nota: *Abordagem A* - Modelo com ano apenas como efeito fixo.

As estimativas dos parâmetros dos modelos ajustados com o ano incluído como efeito fixo principal e também nas interações de efeito fixo (abordagem B) estão na Tabela 17. Diferentemente da abordagem A, se considerarmos somente os efeitos principais, as CPUEs esperadas para barcos com bandeira BRA-PAN seriam maiores que as esperadas por barcos de outras bandeiras. Da mesma forma, uma estimativa simplista dos efeitos principais indicaria que as esperanças das CPUEs para o ano de 1993 seriam superiores ao ano de 1990 (nível de base), e que as esperanças para as taxas de captura nos demais anos não seriam significativamente diferentes das de 1990. No entanto, não é adequada a interpretação individual do efeito principal somente, quando a variável está também incluída em interações de efeito fixo, como é o caso na abordagem B. Geralmente, a interpretação das estimativas é facilitada com algum tipo de cálculo de média ponderada (por exemplo, mínimos quadrados) para estimar o efeito marginal separado de um fator. Para extrair as soluções marginais, assumimos pesos iguais para todos os meses, bandeiras, áreas, números de anzóis e de samburás ao lidar com as interações ano: mês, ano: bandeira, ano: área, ano: anzol e ano: samburá. No entanto, é importante ressaltar que em algumas situações não há uma resposta clara a respeito da questão de quais pesos devem ser usados. Assumir um peso igual a um mês pode ser apropriado (D.S. BUTTERWORTH apud MAUNDER e PUNT, 2004), mas para outros fatores, como área, assumir pesos similares nem sempre é uma escolha natural (MAUNDER e PUNT, 2004).

Tabela 17 – Estimativa dos parâmetros (abordagem B). F – bandeira; Y – ano; A – área; T – trimestre; H – número de samburás (segmento entre uma boia a outra) no espinhel; N – número de anzóis por samburá. Interações particulares são representadas por dois pontos. Significâncias: (***) - 0,001, (**) - 0,01, (*) - 0,05 e (.) - 0,1.

	Estimativa	Erro padrão	z valor	Pr(>|z|)
(Intercepto)	-5,16600	1,40000	-3,691	0,00022 ***
(F) BRA-ESP	0,44170	0,15700	2,813	0,00490 **
(F) BRA-HND	-3,72500	0,39430	-9,448	< 2e-16 ***
(F) BRA-JPN	-1,46300	0,71780	-2,038	0,04150 *
(F) BRA-PAN	6,79000	0,28620	23,725	< 2e-16 ***
(F) BRA-TAI	0,43210	0,29590	1,460	0,14417
(F) BRA-VCT	-1,50000	0,26630	-5,633	1,8e-08 ***
(Y) 1991	0,47700	1,42500	0,335	0,73781
(Y) 1992	-1,60700	1,39200	-1,154	0,24833
(Y) 1993	4,55700	1,99700	2,282	0,02250 *
(Y) 1994	0,40320	1,42300	0,283	0,77696
(Y) 1995	0,05969	1,39900	0,043	0,96596
(Y) 1996	0,00126	1,47500	0,001	0,99932

(continuação)

	Estimativa	Erro padrão	z valor	Pr(>|z|)
(Y) 1997	-1,93400	1,39900	-1,382	0,16696
(Y) 1998	0,87100	1,39300	0,625	0,53177
(Y) 1999	-1,57400	1,38300	-1,138	0,25496
(Y) 2000	-1,53900	1,38100	-1,115	0,26482
(Y) 2001	0,77360	1,38300	0,560	0,57579
(Y) 2002	1,21000	1,39200	0,869	0,38482
(Y) 2003	-3,55600	1,96500	-1,810	0,07036 .
(Y) 2004	-0,49280	1,41100	-0,349	0,72690
(Y) 2005	1,40600	1,40500	1,001	0,31706
(Y) 2006	0,83010	1,40500	0,591	0,55455
(Y) 2007	-1,81100	1,49700	-1,209	0,22662
(Y) 2008	-0,47170	1,46900	-0,321	0,74804
(Y) 2009	-0,18550	1,48500	-0,125	0,90055
(Y) 2010	1,72000	1,56500	1,099	0,27161
(Y) 2011	1,27000	1,52500	0,833	0,40495
(Y) 2012	4,00500	10,9000	0,367	0,71341
(H)	-0,05560	0,22380	-0,248	0,80384
(A) C	-0,13220	0,06745	-1,960	0,04996 *
(A) S	-0,10130	0,07692	-1,317	0,18798
(T) 2	0,62040	0,12590	4,929	8,3e-07 ***
(T) 3	0,81080	0,13330	6,082	1,2e-09 ***
(T) 4	0,60600	0,11510	5,266	1,4e-07 ***
(N)	0,00395	0,00222	1,757	0,07887 .
(F) BRA-ESP: (H)	0,03627	0,01799	2,016	0,04377 *
(F) BRA-HND: (H)	0,22020	0,02939	7,492	6,8e-14 ***
(F) BRA-JPN: (H)	0,04399	0,04182	1,052	0,29275
(F) BRA-PAN: (H)	-0,46000	0,02734	-16,828	< 2e-16 ***
(F) BRA-TAI: (H)	-0,00473	0,02024	-0,234	0,81511
(F) BRA-VCT: (H)	0,00126	0,02031	0,062	0,95059
(F) BRA-ESP: (A) C	0,20390	0,04541	4,491	7,1e-06 ***
(F) BRA-HND: (A) C	-0,17000	0,15180	-1,120	0,26270
(F) BRA-PAN: (A) C	0,21310	0,10090	2,112	0,03466 *
(F) BRA-TAI: (A) C	-0,26480	0,06459	-4,100	4,1e-05 ***
(F) BRA-VCT: (A) C	-0,30240	0,06207	-4,872	1,1e-06 ***
(F) BRA-ESP: (A) S	0,27590	0,05557	4,964	6,9e-07 ***
(F) BRA-HND: (A) S	-0,05103	0,10720	-0,476	0,6341
(F) BRA-JPN: (A) S	2,16000	0,17940	12,045	< 2e-16 ***
(F) BRA-PAN: (A) S	1,60200	0,16780	9,550	< 2e-16 ***
(F)BRA-TAI: (A) S	-0,00624	0,08003	-0,078	0,93785
(F)BRA-VCT: (A) S	0,63150	0,09680	6,524	6,8e-11 ***
(F) BRA-ESP: (T) 2	-0,10070	0,04619	-2,180	0,02923 *
(F)BRA-HND: (T) 2	0,10370	0,09574	1,083	0,27879
(F) BRA-JPN: (T) 2	0,09202	0,17320	0,531	0,59521
(F) BRA-PAN: (T) 2	0,29090	0,07788	3,736	0,00019 ***
(F) BRA-TAI: (T) 2	0,37060	0,12750	2,906	0,00366 **
(F) BRA-VCT: (T) 2	0,43180	0,14300	3,019	0,00254 **
(F) BRA-ESP: (T) 3	-0,23340	0,04930	-4,733	2,2e-06 ***
(F) BRA-HND: (T) 3	0,21780	0,12310	1,769	0,07684 .
(F) BRA-JPN: (T) 3	-0,05736	0,17580	-0,326	0,74416
(F) BRA-PAN: (T) 3	0,36520	0,07577	4,820	1,4e-06 ***
(F) BRA-TAI: (T) 3	0,62070	0,11610	5,344	9,1e-08 ***
(F) BRA-VCT: (T) 3	0,64310	0,12320	5,218	1,8e-07 ***
(F) BRA-ESP: (T) 4	-0,35210	0,05026	-7,006	2,5e-12 ***

(continuação)

	Estimativa	Erro padrão	z valor	Pr(>|z|)
(F) BRA-HND: (T) 4	0,26980	0,09552	2,824	0,00474 **
(F) BRA-JPN: (T) 4	0,10590	0,18240	0,580	0,56167
(F) BRA-PAN: (T) 4	-0,23070	0,07876	-2,929	0,00340 **
(F) BRA-TAI: (T) 4	0,04703	0,09866	0,477	0,63356
(F) BRA-VCT: (T) 4	0,13890	0,09392	1,479	0,13904
(F) BRA-ESP: (N)	-0,00180	0,00032	-5,677	1,4e-08 ***
(F) BRA-HND: (N)	0,00784	0,00104	7,545	4,5e-14 ***
(F) BRA-JPN: (N)	-0,00221	0,00117	-1,882	0,05984 .
(F) BRA-PAN: (N)	-0,02302	0,00069	-33,497	< 2e-16 ***
(F) BRA-TAI: (N)	-0,00417	0,00059	-7,108	1,2e-12 ***
(F) BRA-VCT: (N)	-0,00052	0,00060	-0,862	0,38873
(Y) 1991: (H)	-0,15470	0,22260	-0,695	0,48690
(Y) 1992: (H)	0,22530	0,22110	1,019	0,30827
(Y) 1993: (H)	-0,68590	0,33670	-2,037	0,04163 *
(Y) 1994: (H)	0,03132	0,22240	0,141	0,88803
(Y) 1995: (H)	0,03173	0,22200	0,143	0,88632
(Y) 1996: (H)	0,08063	0,22940	0,351	0,72527
(Y) 1997: (H)	0,28280	0,22370	1,265	0,20602
(Y) 1998: (H)	-0,07110	0,22200	-0,320	0,74881
(Y) 1999: (H)	0,16610	0,22160	0,749	0,45361
(Y) 2000: (H)	0,12730	0,22160	0,574	0,56577
(Y) 2001: (H)	-0,08312	0,22160	-0,375	0,70762
(Y) 2002: (H)	-0,01168	0,22170	-0,053	0,95799
(Y) 2003: (H)	0,40750	0,26670	1,528	0,12651
(Y) 2004: (H)	0,07022	0,22390	0,314	0,75387
(Y) 2005: (H)	-0,00668	0,22360	-0,030	0,97618
(Y) 2006: (H)	0,03926	0,22350	0,176	0,86058
(Y) 2007: (H)	0,39310	0,23270	1,689	0,09114 .
(Y) 2008: (H)	0,23190	0,22800	1,017	0,30900
(Y) 2009: (H)	0,34970	0,22840	1,531	0,12584
(Y) 2010: (H)	-0,08794	0,23880	-0,368	0,71271
(Y) 2011: (H)	0,02506	0,23220	0,108	0,91405
(Y) 2012: (H)	-0,24230	1,00700	-0,241	0,80990
(Y) 1991: (N)	0,00265	0,00233	1,136	0,25583
(Y) 1992: (N)	-0,00181	0,00226	-0,802	0,42236
(Y) 1993: (N)	0,00328	0,00353	0,927	0,35394
(Y) 1994: (N)	-0,00542	0,00230	-2,356	0,01847 *
(Y) 1995: (N)	-0,00148	0,00225	-0,655	0,51274
(Y) 1996: (N)	-0,00649	0,00247	-2,632	0,00849 **
(Y) 1997: (N)	-0,00115	0,00224	-0,515	0,60689
(Y) 1998: (N)	-0,00366	0,00224	-1,630	0,10305
(Y) 1999: (N)	-0,00010	0,00221	-0,046	0,96353
(Y) 2000: (N)	0,00094	0,00221	0,428	0,66877
(Y) 2001: (N)	-0,00255	0,00221	-1,155	0,24822
(Y) 2002: (N)	-0,00554	0,00228	-2,426	0,01526 *
(Y) 2003: (N)	0,01107	0,00527	2,099	0,03579 *
(Y) 2004: (N)	0,00307	0,00230	1,334	0,18217
(Y) 2005: (N)	-0,00349	0,00226	-1,543	0,12278
(Y) 2006: (N)	-0,00161	0,00226	-0,712	0,47648
(Y) 2007: (N)	0,00192	0,00246	0,778	0,43668
(Y) 2008: (N)	-0,00020	0,00268	-0,074	0,94085
(Y) 2009: (N)	-0,00537	0,00253	-2,122	0,03383 *
(Y) 2010: (N)	-0,00182	0,00259	-0,702	0,48249

(continuação)

	Estimativa	Erro padrão	z valor	Pr(>|z|)
(Y) 2011: (N)	-0,00280	0,00262	-1,069	0,28524
(Y) 2012: (N)	-0,00771	0,02297	-0,336	0,73704
(H) : (T) 2	-0,03627	0,01320	-2,747	0,00601 **
(H) : (T) 3	-0,07538	0,01270	-5,934	2,9e-09 ***
(H) : (T) 4	-0,05251	0,01097	-4,785	1,7e-06 ***
(H) : (N)	-0,00046	0,00005	-9,216	< 2e-16 ***
(A) C: (T) 2	-0,17260	0,03991	-4,324	1,5e-05 ***
(A) S: (T) 2	-0,05942	0,06488	-0,916	0,35974
(A) C: (T) 3	-0,06163	0,04381	-1,407	0,15954
(A) S: (T) 3	0,42730	0,05769	7,407	1,3e-13 ***
(A) C: (T) 4	0,13170	0,03781	3,482	0,00050 ***
(A) S: (T) 4	0,16260	0,06271	2,593	0,00951 **
(A) C: (N)	0,00032	0,00020	1,564	0,11793
(A) S: (N)	-0,00191	0,00025	-7,583	3,4e-14 ***
(T) 2: (N)	-0,00139	0,00028	-4,957	7,2e-07 ***
(T) 3: (N)	-0,00183	0,00033	-5,522	3,3e-08 ***
(T) 4: (N)	-0,00084	0,00027	-3,091	0,00199 **

Fonte: Elaborada pelo autor.
Nota: *Abordagem B* - Modelo com ano como efeito fixo e também em interação de efeito fixo.

As estimativas dos parâmetros do modelo misto com ano incluído como efeito fixo principal e também em interações de efeito aleatório (abordagem C) estão na Tabela 18. De maneira geral as interpretações dos coeficientes estimados para as bandeiras não diferem muito das realizadas para a abordagem A, as CPUEs esperadas para barcos alugados da Espanha são maiores que as esperadas por barcos de outras bandeiras. Apenas os barcos arrendados de Honduras e Japão apresentaram CPUEs semelhantes aos barcos nacionais. No entanto, no caso do fator ano, as estimativas de parâmetros foi significativamente diferente de zero a partir do ano de 2003, indicando que há evidências de que as CPUEs aumentaram em relação ao ano de 1990. Em adição, há evidências de que a esperança da CPUE para as regiões norte, centro e sul não diferem entre si. As estimativas também indicam que a esperança para o primeiro trimestre é semelhante aos demais períodos do ano.

Tabela 18 – Estimativa dos parâmetros (abordagem C). F – bandeira; Y – ano; A –área; T – trimestre; H – número de samburás (segmento entre uma boia a outra) no espinhel; N – número de anzóis por samburá. Interações particulares são representadas por dois pontos. Significâncias: (***) - 0,001, (**) - 0,01, (*) - 0,05 e (.) - 0,1.

	Estimativa	Erro padrão	z valor	Pr(>|z|)
(Intercepto)	-4,58038	0,38254	-11,974	< 2e-16 ***
(F) BRA-ESP	0,470809	0,23169	2,032	0,04215 *
(F) BRA-HND	-0,12135	0,27045	-0,449	0,65365
(F) BRA-JPN	-0,08772	0,30536	-0,287	0,77390
(F) BRA-PAN	-0,88198	0,25417	-3,47	0,00052 ***
(F) BRA-TAI	-0,72867	0,26814	-2,717	0,00658 **
(F) BRA-VCT	-1,8409	0,26551	-6,933	4,1e-12 ***
(Y) 1991	0,204066	0,32637	0,625	0,53180
(Y) 1992	0,094077	0,32751	0,287	0,77392
(Y) 1993	0,120371	0,4263	0,282	0,77766
(Y) 1994	-0,16553	0,33086	-0,5	0,61686
(Y) 1995	0,248312	0,32086	0,774	0,43899
(Y) 1996	-0,04954	0,38906	-0,127	0,89868
(Y) 1997	0,081469	0,34336	0,237	0,81245
(Y) 1998	-0,04431	0,34348	-0,129	0,89735
(Y) 1999	0,344402	0,31932	1,079	0,28078
(Y) 2000	0,316806	0,31686	1,000	0,31740
(Y) 2001	-0,04979	0,31696	-0,157	0,87518
(Y) 2002	0,221753	0,32469	0,683	0,49463
(Y) 2003	0,718117	0,38888	1,847	0,06480 .
(Y) 2004	0,957282	0,31548	3,034	0,00241 **
(Y) 2005	1,023149	0,31177	3,282	0,00103 **
(Y) 2006	1,368268	0,31210	4,384	1,2e-05 ***
(Y) 2007	1,269885	0,31300	4,057	4,9e-05 ***
(Y) 2008	1,371912	0,31608	4,34	1,4e-05 ***
(Y) 2009	1,133626	0,31295	3,622	0,00029 ***
(Y) 2010	1,348995	0,31638	4,264	2,0e-05 ***
(Y) 2011	1,232554	0,31275	3,941	8,1e-05 ***
(Y) 2012	1,327403	0,37151	3,573	0,00035 ***
(A) C	0,03643	0,15496	0,235	0,81414
(A) S	-0,00953	0,15572	-0,061	0,95121
(T) 2	0,124076	0,10704	1,159	0,24638
(T) 3	0,130148	0,10786	1,207	0,22758
(T) 4	0,168415	0,10802	1,559	0,11898
(H)	-0,11427	0,01181	-9,679	< 2e-16 ***
(N)	-0,00404	0,00037	-10,927	< 2e-16 ***

Fonte: Elaborada pelo autor.
Nota: *Abordagem C* - Modelo com ano como efeito fixo e também em interação de efeito aleatório.

As estimativas das componentes de variância para a abordagem C estão na Tabela 19. Como as interações entre bandeira e área de pesca (F:A), bandeira e trimestre (F:T), bandeira e número de samburás (F:H), bandeira e número de anzóis por samburá (F:N), ano e número de samburás (Y:H), ano e número de anzóis por

samburá (Y:N), área de pesca e trimestre (A:T), área de pesca e número de samburás (A:H), área de pesca e número de anzóis por samburás (A:N), trimestre e número de samburá (T:H), trimestre e número de anzóis por samburá (T:N) e número de samburá e número de anzóis por samburá (H:N) são consideradas como efeito aleatório, o interesse reside em estimar a sua variabilidade (por meio de seus componentes de variância).

Tabela 19 – Estimativas dos componentes de variâncias para os efeitos aleatórios (Abordagem C). F – bandeira; Y – ano; A – área; T – trimestre; H – número de samburás (segmento entre uma boia a outra) no espinhel; N – número de anzóis por samburá. Interações particulares são representadas por dois pontos.

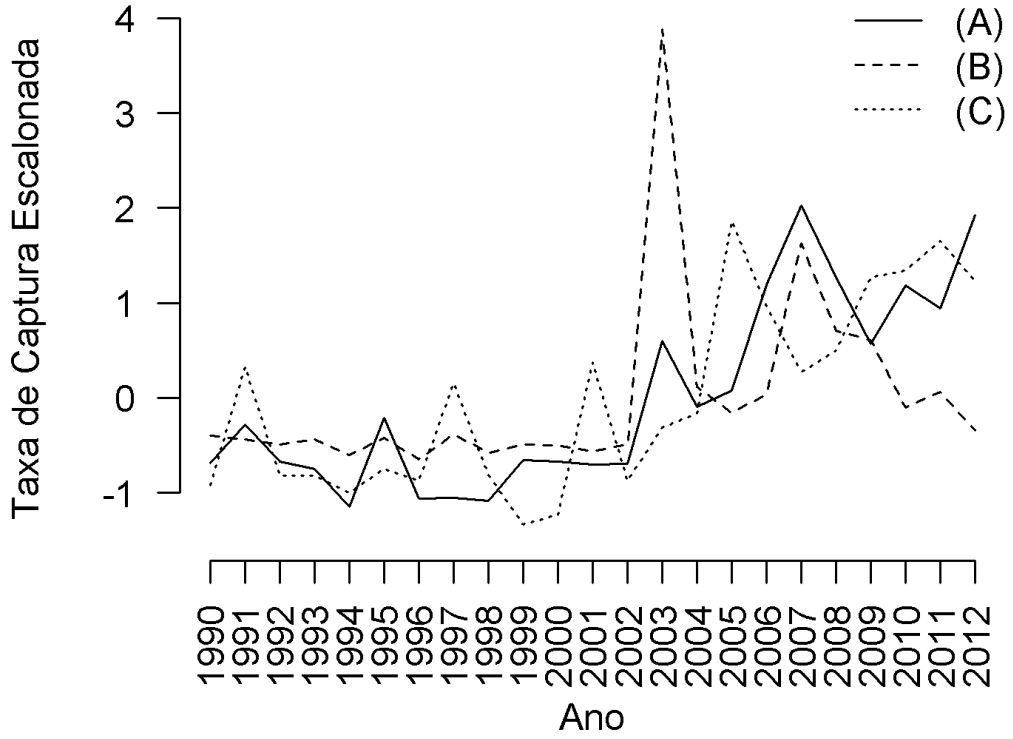
Grupo	Variância	Desvio Padrão
H:N	0,06	0,25
Y:H	0,06	0,26
T:H	0,01	0,11
A:H	0,02	0,13
F:H	0,03	0,16
Y:N	0,15	0,38
T:N	0,05	0,22
F:N	0,21	0,46
A:N	0,05	0,23
F:T	0,02	0,13
F:A	0,06	0,24
A:T	0,01	0,07

Fonte: Elaborada pelo autor.
Nota: Número de observações: 34145; Grupos: H:N = 3866; Y:H = 2681; T:H = 2599; A:H = 2465; F:H = 2432; Y:N = 2150; T:N = 1607; F:N = 1362; A:N = 1311; F:T= 28; F:A = 20; A:T = 12. *Abordagem C* - Modelo com ano como efeito fixo e também em interação de efeito aleatório.

As CPUEs padronizadas calculadas com base nas abordagens A, B e C são apresentadas na Figura 21. As estimativas foram escalonadas como $z = (x - \bar{x})/s$ para facilitar a comparação. De modo geral, nota-se que no período compreendido entre 1990 e 2001 não houve tendências gerais definidas para as três séries avaliadas, no entanto a partir do ano 2002 verifica-se uma tendência crescente nas séries avaliadas com as abordagens A e C. CPUEs padronizadas mais elevadas foram verificadas nos anos de 2007 para a abordagem A, no ano de 2003 para a

abordagem B e em 2005 para a abordagem C. Contudo, a estimativa do índice padronizado com a abordagem B para o ano de 2003 é bastante imprecisa.

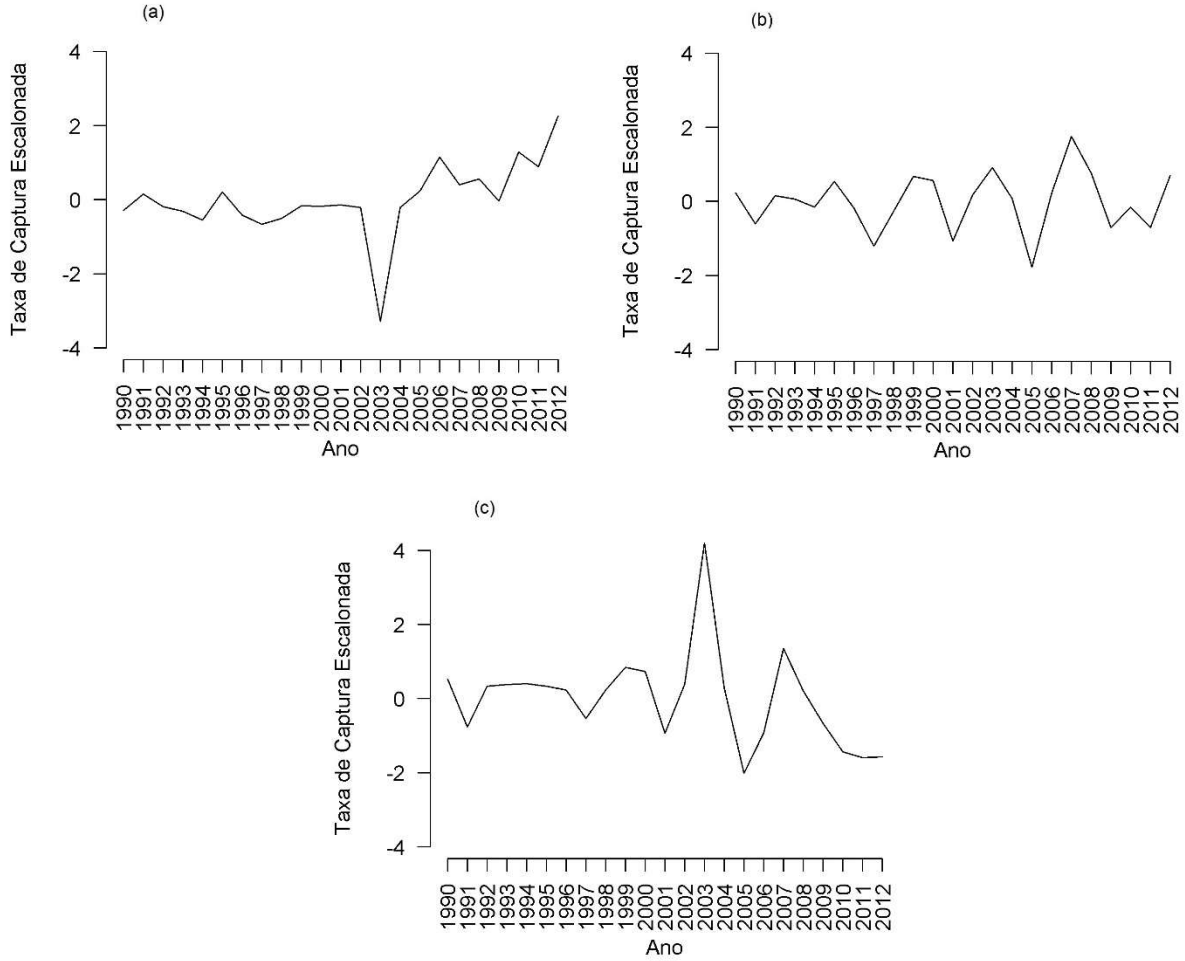
Figura 21 – Taxa de captura por unidade de esforço (CPUE) padronizada calculada com as abordagens A, B e C. As estimativas foram escalonadas calculando a diferença entre o valor original e a média da série, dividido pelo desvio padrão da série.



Fonte: Elaborada pelo autor.

Diferenças dois a dois entre as CPUEs padronizadas escalonadas são apresentadas na Figura 22. As séries de CPUEs padronizadas com as abordagens A e B, Figura 22(a), foram as que apresentaram comportamentos mais semelhantes, diferenças mais acentuadas foram verificadas nos anos de 2003 e 2012. Ao comparar as séries produzidas com abordagens A e C, Figura 22(b), nota-se maiores variações entre elas dos anos de 1999 a 2008. As diferenças entre as abordagens B e C, Figura 22(c), são acentudas a partir do ano 2003, com destaque para esse ano.

Figura 22 – Diferença entre as taxas de captura por unidade de esforço (CPUE) padronizada calculada com as abordagens A e B (a), com as abordagens A e C (b) e com as abordagens B e C(c).



Fonte: Elaborada pelo autor.

Como a inspenção visual é subjetiva, foi utilizado o teste não paramétrico de Kruskal-Wallis para comparar as CPUEs padronizadas com as três abordagens em cada ano avaliado. Os resultados são apresentados na Tabela 20. A escolha pelo teste de Kruskal-Wallis foi realizada após identificar, com o teste de Kolmogorov-Sminorv ($p<0,05$), que as predições do modelo para cada abordagem não são normalmente distribuídas, sendo necessário o uso de um teste estatístico não-paramétrico. A partir dos resultados obtidos com o teste de Kruskal-Wallis e considerando as comparações anuais entre as CPUEs padronizadas com as três abordagens, foi comprovado que apenas nos anos de 1991, 1998 e 2011 não houve diferenças significativas a 5% de probabilidade entre as abordagens A, B e C. Em adição, nos anos de 1990, 2001, 2003, 2004, 2007, 2008 e 2012 houve diferenças

significativas entre as CPUEs padronizadas com as três abordagens. Para detectar quais abordagens se diferenciam entre si foi realizado um pós-teste para comparações múltiplas não paramétricas. Não foram constatadas diferenças significativas entre as CPUEs padronizadas com as abordagens A e B nos anos de 1991 a 1995, 1998 a 2000, 2002, 2005, 2006 e 2011. Para as abordagens A e C não houve diferenças significativas entre as CPUEs padronizadas nos anos 1991, 1994, 1996, 1998, 2009 e 2011. Para as abordagens B e C não houve diferenças significativas nas estimativas das CPUEs padronizadas nos anos de 1991, 1996 a 1998, 2005, 2010 e 2011. De forma geral, verificou-se que as estimativas das CPUEs padronizadas com as abordagens A e B foram as mais semelhantes.

Tabela 20 – Teste de Kruskal-Wallis para comparação anual entre as abordagens A, B e C. GL – graus de liberdade.

Ano	Qui-quadrado	Gl	p-valor	Abordagens
1990	42,10	2	7,2e-10*	A^{a}, B^{b}, C^{c}
1991	4,39	2	0,11130	A^{a}, B^{a}, C^{a}
1992	20,72	2	3,2e-05*	A^{a}, B^{a}, C^{b}
1993	25,25	2	3,3e-06*	A^{a}, B^{a}, C^{b}
1994	6,58	2	0,03728*	A^{ab}, B^{a}, C^{b}
1995	40,22	2	1,9e-09*	A^{a}, B^{a}, C^{b}
1996	5,10	2	0,07814	A^{a}, C^{ab}, B^{b}
1997	29,48	2	4,0e-07*	A^{b}, B^{a}, C^{a}
1998	1,56	2	0,45770	A^{a}, B^{a}, C^{a}
1999	105,33	2	< 2,2e-16*	A^{a}, B^{a}, C^{b}
2000	82,83	2	< 2,2e-16*	A^{a}, B^{a}, C^{b}
2001	18,98	2	7,6e-05*	A^{a}, B^{b}, C^{c}
2002	24,16	2	5,7e-06*	A^{a}, B^{a}, C^{b}
2003	163,39	2	< 2,2e-16*	A^{a}, B^{b}, C^{c}
2004	46,58	2	7,7e-11*	A^{a}, B^{b}, C^{c}
2005	6,90	2	0,03179*	A^{b}, B^{ab}, C^{a}
2006	11,33	2	0,00347*	A^{a}, B^{a}, C^{b}
2007	98,52	2	< 2,2e-16*	A^{a}, B^{b}, C^{c}
2008	53,96	2	1,9e-12*	A^{a}, B^{b}, C^{c}
2009	32,25	2	9,9e-08*	A^{b}, B^{a}, C^{b}
2010	7,66	2	0,02164*	A^{a}, B^{b}, C^{b}
2011	0,65	2	0,72290	A^{a}, B^{a}, C^{a}
2012	49,01	2	2,3e-11*	A^{a}, B^{b}, C^{c}

Fonte: Elaborada pelo autor.
Nota: *Abordagem A* - Modelo com ano apenas como efeito fixo; *Abordagem B* - Modelo com ano como efeito fixo e também em interação de efeito fixo; *Abordagem C* - Modelo com ano como efeito fixo e também em interação de efeito aleatório. ^{a,b,c}Abordagens estatisticamente diferentes. Nível de significância: $p<0,05$ (*).

Na maioria dos artigos sobre padronização da CPUE, assume-se que as estimativas dos parâmetros para os níveis do fator ano refletem a variação anual da biomassa (MAUNDER e PUNT, 2004). Existem três alternativas para lidar com os efeitos das interações entre o ano e outros fatores ao estimar a série temporal da CPUE padronizada: 1) não as incluir como efeito fixo (abordagem A) ou 2), as incluir como efeitos aleatórios (Abordagem C), o que é de certa forma uma outra maneira de ignorá-las; ou 3) considerá-las (abordagem B), sendo que algum tipo de média ponderada pode ser calculada sobre os termos de interação para obter uma solução marginal para ano. Nas reuniões da ICCAT foram utilizadas as abordagens relativas à inclusão do ano nas interações (fixas ou aleatórias) (*e.g.* fixas - HAZIN *et al.*, 2007, CARNEIRO *et al.*, 2015, TSAI e LIU, 2016; aleatória - CHANG, 2003, AROCHA *et al.*, 2016, WALTER e LAURETTA, 2016). Neste trabalho, os resultados dos testes de hipóteses indicaram que as CPUEs de espadarte calculadas com as abordagens A, B e C foram significativamente diferentes em vários anos avaliados. Esta é uma indicação de que, a escolha da metodologia afeta o entendimento sobre a situação do estoque.

De acordo com Venables e Dichmont (2004) o uso de GLMMs é recomendado na ciência da pesca, uma vez que os registros de captura e esforço provenientes das mesmas embarcações em momentos diferentes, provavelmente, são correlacionados. A propriedade que os termos aleatórios têm de induzir correlações entre as observações é possivelmente o efeito mais importante de um modelo de efeitos mistos, pois permitem que uma medida de integração de dados ocorra na análise, a denominada propriedade de "*borrowing strength*". Nos GLMMs, os termos aleatórios não contribuem para a parte fixa da média, mas os componentes de variância associados a eles aumentam a variabilidade das previsões de forma adequada (VENABLES e DICHMONT, 2004; CRAWLEY, 2007). A análise dos índices de abundância pelo GLM pode subestimar o nível de variabilidade, uma vez que ignora a natureza agrupada dos reboques dentro das embarcações (HELSER et al., 2004, BAUM e BLANCHARD, 2010). Hinde e Demetrio (1998) destacam que 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.

Os GLMMs têm sido cada vez mais empregados nas padronizações da CPUE. Comumentemente, a distribuição binomial negativa é utilizada para dados de

contagem sobredispersos tanto nos GLMM como nos GLM (*e.g.* IJIMA, 2017). Neste trabalho, os modelos binomiais negativos foram os mais adequados para obtenção das séries padronizadas de CPUEs do espadarte capturados pela frota brasileira no Atlântico Sul. Estudos realizados por Vieira (2015) também revelaram a superioridade do modelo binomial negativo com função de ligação logarítmica para a padronização da CPUE dos espadartes capturados no Atlântico Sul, após o ajuste de 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.

Considerando a estimativa de CPUE padronizada de espadarte como um índice de abundância relativo, os resultados obtidos com a abordagem A e C sugerem que a biomassa do estoque de espadarte do Atlântico Sul aumentou a partir do ano 2003, indicando que a pesca atuneira no Atlântico Sul tem sido realizada de forma sustentável. Os resultados obtidos com a abordagem B sugere que a abundância do estoque variou ao longo dos anos estudados, mas sem afetar fortemente o estoque da espécie. Esses resultados são concordantes 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 de 1990 a 2012. Para se obter uma interpretação do estado atual do estoque, seria necessário estudos incluindo os anos mais recentes. Outra observação importante é que nesta análise foi utilizado dados de uma frota somente (Brasil). Pesquisas similares com outros conjuntos de dados e frotas devem ser encorajadas para investigar se as diferenças entre as estimativas são tão altas quanto as que foram obtidas com a frota brasileira.

3.5 Conclusão

Os resultados obtidos nesse trabalho com o teste de Kruskal-Wallis comprovaram que apenas nos anos de 1991, 1998 e 2011 não houve diferenças significativas entre as três abordagens ($p<0,05$), indicando que a escolha de abordagem pode conduzir a resultados equivocados sobre a situação do estoque quando analisado apenas os anos isoladamente. Verificou-se ainda, que as

estimativas anuais das CPUEs padronizadas com as abordagens A e B foram as mais semelhantes.

As tendências temporais das estimativas de CPUEs padronizadas com as abordagens A e B não são tão diferentes. Para as abordagens A e C os valores das CPUEs aumentaram a partir de 2003. Já para a abordagem B os valores de CPUEs aumentam em 2003, mas diminuem em 2010, 2011 e 2012. No entanto, os modelos selecionados para a estimativa da CPUE padronizada utilizando as abordagens A, B e C merecem atenção e são questionáveis, visto que pressupostos foram violados e identificados pelo diagnóstico dos resíduos. Diante deste fato, sugere-se como trabalho futuro a utilização dos modelos aditivos generalizados para posição, escala e forma (GAMLSS) que permitem flexibilidade na modelagem estatística.

4 Considerações Finais

Os modelos aditivos generalizados para posição, escala e forma representam um instrumento poderoso na modelagem de dados de natureza discreta com comportamento altamente assimétrico, curtótico e heretocedástico. Essa técnica de modelagem estatística permite o ajuste de uma ampla família de distribuições discretas para a variável resposta. Considera-se que esta classe de modelos é um passo na direção realista e modelagem flexível na ciência pesqueira. Dentre os diversos modelos que podem ser utilizados, modelos com variável resposta seguindo distribuição beta binomial podem ser aplicados em várias áreas do conhecimento, visto que dados com natureza discreta assimétrica são comuns em situações reais.

Neste trabalho, os modelos ajustados com distribuição beta binomial foram eficientes para modelar a proporção de lagostas fêmeas em atividade reprodutiva. O entendimento do ciclo e da periodicidade de reprodução das lagostas é fundamental para a tomada de medidas administrativas no manejo da pescaria, como por exemplo, a definição dos períodos de defeso tradicionalmente utilizados na gestão pesqueira no Brasil. Historicamente tem-se centrado o período de defeso de maneira fixa na primeira metade do ano, com vistas, entre outras coisas, a proteger o estoque no período de desova. No entanto, os modelos beta binomiais selecionados permitiram identificar que apesar da captura de lagostas com característica reprodutiva ser mais frequente no primeiro semestre, ocorreram meses com intensa atividade reprodutiva que não fazem parte do defeso, como por exemplo, o mês de junho para a espécie *P. argus* e, os meses de junho e novembro para a espécie *P. laevicauda*. Outro fato que chamou atenção foi os meses de dezembro a fevereiro que apesar de fazerem parte da paralisação das atividades pesqueiras ocorreram pequenas proporções de fêmeas maturas de *P. argus.* Diante dessas questões, sugere-se uma reavaliação do período de defeso tradicionalmente utilizado a décadas no Brasil.

Percebeu-se que a abordagem via GAMLSS conduziram a resultados mais adequados aos dados positivos assimétricos da lagosta em comparação aos modelos lineares generalizados, pois as suposições de normalidade e homogeneidade das variâncias foram satisfeitas.

Quanto ao espadarte, a necessidade da avaliação do seu estoque pesqueiro é indispensável para assegurar uma exploração adequada e garantir a sobrevivência da espécie. Tradicionalmente, a avaliação do estoque é feita padronizando a CPUE com base em estimativas dos parâmetros para os níveis do fator "ano", calculados com Modelos Lineares Generalizados ou Modelos Lineares Generalizados Mistos. Apesar de ser comum o uso de três abordagens para a forma de tratar interações que envolvem ano, não há muita informação sobre as consequências da escolha por uma ou outra abordagem. Os resultados obtidos com o teste de Kruskal-Wallis indicaram que a escolha de abordagem pode conduzir a resultados equivocados sobre a situação do estoque quando analisado apenas os anos isoladamente. Quando avaliadas as tendências temporais das CPUEs padronizadas com as abordagens A e C as interpretações não são tão diferentes.

Os modelos selecionados para a estimativa da CPUE padronizada utilizando as abordagens A, B e C merecem atenção e são questionáveis, visto que pressupostos foram violados e identificados pelo diagnóstico dos resíduos. Diante deste fato, sugere-se como trabalho futuro a utilização dos modelos aditivos generalizados para posição, escala e forma (GAMLSS) para estimar as CPUEs padronizadas com as abordagens A, B e C, visto que esses modelos permitem flexibilidade na modelagem estatística e são adequados para modelar variáveis respostas do tipo contagem que sejam sobredispersas, assimétricas, curtóticas e heterocedásticas.

Referências

ABID, N.; IDRISSI, M. Swordfish. In: ICCAT 2006 - 2009 Manual. International Commission for the Conservation of Atlantic Tuna publications. p. 181-198, 2006.

AKAIKE, H. A new look at the statistical model identification. IEEE Transactions on Automatic Control., Boston, v.19, n.6, p.716-723, Dec. 1974.

AKAIKE, H. Information measures and model selection. Bulletin of the International Statistical Institute, v. 50, 277-290, 1983.

AMADOR, J. P.; LOPES, S. J.; PEREIRA, J. E. DA S.; SOUZA, A. M.; TOEBE, M. Análise das pressuposições e adequação dos resíduos em modelo de regressão linear para valores individuais, ponderados e não ponderados, utilizando procedimentos do SAS®. Ciência e Natura, v. 33, n. 2, p. 7, 2011.

AMORIM, A. F.; ARFELLI, C. A. Estudo biológico-pesqueiro do espadarte, Xiphias gladius Linnaeus, 1758, no sudeste e sul do Brasil (1971 a 1981). Boletim do Instituto de Pesca, 11: 35-62, 1984.

AMORIM, A. F.; BRAGA, F. M. S.; FAGUNDES, L.; COSTA, F. E. S.; ARFELLI, C. A. The evolution of tuna fishery in Santos- São Paulo, Southern Brazil (1971-1995). Collective Volume of Scientific Papers ICCAT, 46(4): 425-429, 1997.

ANDRADE, H. A. Stock assessment of the red spiny lobster (*Panulirus argus*) caught in the tropical southwestern Atlantic. Latin American Journal of Aquatic Research, v. 43, n. 1, 2015.

ANDRADE, H. A. Taxa de captura para o bonito-listrado (*Katsuwonus pelamis*) do sudoeste do oceano Atlântico Sul. Boletim do Instituto de Pesca, 2007.

ANON. Ministério do Desenvolvimento, Indústria e Comércio Exterior -MDIC (Sistema AliceWeb). Disponível em: http://aliceweb.desenvolvimento.gov.br/. Acesso em: 16 de abril de 2013.

ARAÚJO, P. V. do N.; AMARAL, R. F. do. Mapping of coral reefs in the continental shelf of Brazilian Northeast through remote sensing. Journal of Integrated Coastal Zone Management / *Revista de Gestão Costeira Integrada*, Lisboa , v. 16, n. 1, p. 5-20, 2016 .

ARFELLI, C. A. Estudo da pesca e aspectos da dinâmica populacional de espadarte, Xiphias gladius L. 1758, no Atlântico Sul. Rio Claro. 175f. 1996 (Tese de Doutorado. Universidade Estadual Paulista - UNESP).

AROCHA, F., ORTIZ, M., MARCANO, J. H. Standardized catch rates for blue shark (Prionace glauca) from the venezuelan pelagic longline fishery in the caribbean sea and adjacent waters of the north Atlantic ocean: period 1994-2013. Collect. Vol. Sci. Pap. ICCAT, 72(4), 1033-1043, 2016.

BAISRE, Julio A. An overview of Cuban commercial marine fisheries: the last 80 years. Bull. Mar. Sci, v. 94, 2017.

BATES, Douglas M. lme4: Mixed-effects modeling with R. 2010.

BAUM, J. K.; BLANCHARD, W. Inferring shark population trends from generalized linear mixed models of pelagic longline catch and effort data. Fish. Res., v. 102, n. 3, p. 229-239, 2010.

BERNASCONI, J. F.; PERIER, M. R.; DI GIACOMO, E. E. Standardized catch rate of cockfish, Callorhinchus callorynchus, in a bottom trawl fishery of Patagonia: Is it possible its use as a predictor of abundance trend?. Braz. j. oceanogr., São Paulo, v. 63, n. 2, p. 147-160, 2015 .

BERTELSEN, R. D. Characterizing daily movements, nomadic movements, and reproductive migrations of Panulirus argus around the Western Sambo Ecological

Reserve (Florida, USA) using acoustic telemetry. *Fisheries Research*, v. 144, p. 91-102, 2013.

BOX, G. E. P. E COX, D. R. An analysis of transformations (with discussion). Journal of the Royal Statistical Society B 26, 211-252, 1964.

BRESLOW, N. E; CLAYTON, D. G. Approximate inference in generalized linear mixed models. Journal of the American Statistical Association 88, p. 9-25, 1993.

BREUSCH, TREVOR S.; PAGAN, ADRIAN R. A simple test for heteroscedasticity and random coefficient variation. Econometrica: Journal of the Econometric Society, p. 1287-1294, 1979.

BUCARAM, S. J.; WHITE, J. W.; SANCHIRICO, J. N.; WILEN, J. E. Behavior of the Galapagos fishing fleet and its consequences for the design of spatial management alternatives for the red spiny lobster fishery. Ocean & Coastal Management, v. 78, p. 88-100, 2013. ISSN 0964-5691.

BURNHAM, K. P.; ANDERSON, D. R. Model Selection and Multimodel Inference: a Practical Information-Theoretic Approach (second ed.), Springer-Verlag, New York, 2002. ISBN 0-387-95364-7

BUSTOS, B.; CUBILLOS, L. A. Cambios interanuales en la talla de madurez de sardina común, Strangomera bentincki, en la zona centro-sur de Chile (2007-2012). Revista de biología marina y oceanografía, v. 51, n. 2, p. 317-325, 2016.

CALAZANS, D. K. de. Estudos oceanográficos: do instrumental ao prático. Pelotas: Editora Textos, 2011. 462 p. ISBN: 978-85-99333-06-8

CAMPBELL, R. CPUE standardization and the construction of indices of stock abundance in a spatially varying fishery using general linear models. Fish. Res. 70: 209-227, 2004.

CAREY, F. G., ROBISON, B. H. Daily patterns in the activities of swordfish, Xiphias gladius, observed by acoustic telemetry. Fish. Bull. 79, 277-292, 1981.

CARNEIRO, V., FIALHO, E., ANDRADE, H. A. Updated standardized catch rates for skipjack tuna (Katsuwonus pelamis) caught in the southwest of South Atlantic Ocean. Collect. Vol. Sci. Pap. ICCAT, 71(1), 306-316, 2015.

CASTIGLIONI, D. da S.; SILVA-CASTIGLIONI, D. da; OLIVEIRA, P. J. A. de. Reproductive biology of *Ucides cordatus* (Linnaeus, 1763) (Crustacea, Brachyura, Ucididae) in two mangrove areas from the south coast of state of Pernambuco, Brazil. *Revista de Gestão Costeira Integrada*, Lisboa, v. 13, n. 4, p. 433-444, 2013.

CAVALCANTE, P. P. L; FURTADO NETO, M. A. de A.; COSTA, J. M. da; NEVES, S. da S. Ordenamento da Pesca da Lagosta: Uma experiência desordenada. Fortaleza: Ministério da Pesca e Aquicultura; Universidade do Ceará, 2011. 256p. ISBN:978-85-7564-557-4

CEPENE. Boletim Técnico-Científico do CEPENE - Centro de Pesquisa e Gestão de Recursos Pesqueiros do Litoral Nordeste. Tamandaré, PE: Cepene, v. 19, n. 1, 2013. ISSN 0104-6411

CHACE Jr., F. A.; DUMONT, W. H. Spiny lobsters – identification, world distribution, and U. S. trade. Comm. Fish. Rev., Washington, v.11, n. 5. p. 1-12, 1949.

CHANG, S. K. Analysis of Taiwanese white marlin catch data and standardization of catch rates. ICCAT Col. Vol. Sci. Pap, 55(2), 453-466, 2003.

COELHO, P. A. Bases para regulamentação da pesca de lagostas. Boletim Estatístico da Pesca, v.2, p.3-6, 1962.

COELHO, P. A.; MOURA, S. J. C.; SILVA, V. R. C. & COÊLHO, R.R. Nota sobre a reprodução das lagostas *Panulirus argus* (Latreille) e *P. laevicauda* (Decapoda: Palinuridae) no litoral do Estado de Pernambuco. Trab. Oceanogr. Univ. Rec., v. 3, n. 1, p. 61-67, 1963.

COLE, T. J.; GREEN, P. J. Smoothing reference centile curves: the LMS method and penalized likelihood. Statist. Med., 11, 1305–1319, 1992.

CORDEIRO, G. M. e LIMA NETO, E. A. Modelos Paramétricos. Recife: Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, 2006.

COSTA, L. S.; SILVA, J. R. da; SILVA, F. Análise da produção pesqueira na região norte do Brasil: uma abordagem através de modelos de regressão e componentes principais. Observatorio de la Economía Latinoamericana, p. 1-15, 2013.

COX, D. R.; SNELL, E. J. A general definition of residuals. Journal of the Royal Statistical Society. Series B (Methodological), p. 248-275, 1968.

CRAWLEY, M. J. The R book. (Ed.). Chichester: John Wiley and Sons, 2007. 942 p.

DE OLIVEIRA, Í. A.; HAZIN, H. G.; HAZIN, F. H. V.; TRAVASSOS, P. E. P.; DA SILVA, G. B.; MOURATO, B. L.; CARVALHO, F. Distribuição do agulhão negro no atlântico sul e equatorial e potencial de estratégia de manejo espacial. Boletim do Instituto de Pesca, v. 41, n. 3, p. 607-617, 2015.

DEMÉTRIO, C. G. B. Modelos Lineares Generalizados em experimentação agronômica. Piracicaba: ESALQ/USP, 2002. 121p.

DIAS-NETO, J. Gestão do uso dos recursos pesqueiros marinhos no Brasil. 242p. IBAMA, Brasília, DF, Brasil. ISBN, p. 85-7300, 2010.

DIAS-NETO, J.; DIAS, J. de F. O. O uso da biodiversidade aquática no Brasil: Uma avaliação com foco na pesca. Brasília: Ibama, 288p, 2015. ISBN 978-85-7300-379-6.

DOBSON, A. J. An Introduction to Generalized Linear Models, 2 ed. Boca Raton: Chapman and Hall, 2008. 307p.

DUNN, P. K.; SMYTH, G. K. Randomized quantile residuals. Journal of Computational and Graphical Statistics, v. 5, p. 236-244, 1996.

FAO. FAO yearbook. Fishery and Aquaculture Statistics. 2014/FAO annuaire. 2016a. ISBN 978-92-5-009268-3

FAO. Fishery and Aquaculture Statistics Yearbook. 2012. ISSN 2070-6057

FAO. The State of World Fisheries and Aquaculture. Rome, 2016. 200 p. ISBN 978-92-5-109185-2

FARIA, A.; SILVA, D. Os palinurideos do Brasil (Crustacea – Macrura). Rev. Dep. nac. Prod. anim., Rio de Janeiro, v.4, n. 4/6, p. 1-45, 1937.

FENG, C.; SADEGHPOUR, Al.; LI, L. Randomized Quantile Residuals: an Omnibus Model Diagnostic Tool with Unified Reference Distribution. arXiv preprint arXiv:1708.08527, 2017.

FILLIBEN, J. J. The Probability plot correlation coefficient test for normality. Technometrics, v. 17, n. 1, p. 111-117, Feb. 1975.

FOLSOM, W. B., WEIDNER, D. W., WILDMAN, M. R. World swordfish fisheries: An analysis of swordfish fisheries, market trends, and trade patterns Past-Present Future. NOAA Tech. Memo. NMFS. 1997.

FONTELES-FILHO, A. A. State of the lobster fishery in Northeast Brazil. Spiny lobster: Fisheries and Culture, Second Edition, p. 121-134, 2000.

FONTELES-FILHO, A. A.; IVO, C. T. C. Comportamento migratório da lagosta *Panulirus argus (Latreille*), em frente ao Estado do Ceará, Brasil. Bol. Inst. Oceanogr. [online]. vol.29, n.2, p.173-176, 1980. ISSN 0373-5524. http://dx.doi.org/10.1590/S0373-55241980000200035.

FONTELES-FILHO, A.A. Síntese sobre a lagosta-vermelha (*Panulirus argus*) e a lagosta-verdade (*Panulirus laevicauda*) do Nordeste do Brasil, p. 257-265, in Haimovici, M. (org.), A prospecção pesqueira e abundância de estoques marinhos no

Brasil nas décadas de 1960 a 1990: levantamento de dados e avaliação crítica. Ministério do Meio Ambiente, 329 p., Brasília, 2007.

GALLARDOA, M. de L. Á.; LÓPEZ, A. E. G.; RAMOS, M.; MUJICA, A.; MUÑOZ, P.; SELLANES, J.; YANNICELLI, B. Reproductive patterns in demersal crustaceans from the upper boundary of the OMZ off north-central Chile. Continental Shelf Research, v. 141, p. 26–37, 2017.

GIRALDES, B. W.; SMYTH, D. M. Recognizing Panulirus meripurpuratus sp. nov. (Decapoda: Palinuridae) in Brazil - Systematic and biogeographic overview of Panulirus species in the Atlantic Ocean. Zootaxa, 4107(3): 353-366, 2016.

GONZÁLEZ, A.; VEGA, R.; YÁÑEZ, E. Operational interactions between the South American sea lion *Otaria flavescens* and purse seine fishing activities in northern Chile. Revista de Biología Marina y Oceanografía, v. 50, n. 3, 2015.

GUITART-MANDAY, G. Biology pesquera del imperador o pez espada, *Xiphias gladius Linnaeus* (Teleostomi: Xiphiidae) en las aguas de Cuba. Poeyana 1, 1-37, 1964.

HARPER, S.; GUZMAN, H. M.; ZYLICH, K.; ZELLER, D. Reconstructing Panama's Total Fisheries Catches from 1950 to 2010: Highlighting Data Deficiencies and Management Needs. Marine Fisheries Review, v. 76, p. 51-65, 2014.

HARTLEY, H. O.; RAO, J. N. K. Maximum likelihood estimation for the mixed analysis of variance model, Biometrika, v. 54, 93-108, 1967.

HASTIE, T. J.; TIBSHIRANI, R. J. Generalized Additive Models. Chapman and Hall, London, 1990.

HAZIN, H. G. Influência das variáveis oceanográficas na dinâmica populacional e pesca do espadarte, *Xiphias gladius Linnaeus 1758*, capturados pela frota brasileira. 2006. Tese de Doutorado. Faculdade de Ciências do Mar e do Ambiente, Universidade do Algarve, Portugal, 216p.

HAZIN, H. G.; HAZIN, F.; TRAVASSOS, P.; CARVALHO, F. C.; ERZINI, K. 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(6). p. 2039-2047, 2007.

HELSER, T. E.; PUNT, A. E.; METHOT, R. D. A generalized linear mixed model analysis of a multi-vessel fishery resource survey. Fish. Res., v. 70, n. 2, p. 251-264, 2004.

HILBORN, R., WALTERS, C.J. Quantitative Fisheries Stock Assessment. Chapman & Hall, London, 1992.

HINDE, J.P. ; DEMÉTRIO, C.G.B. Overdispersion: models and estimation. Computation Statistics and Data Analysis, v.27, p.151-170. 1998.

IBAMA. Plano de Gestão para uso sustentável de lagostas no Brasil. Ministério do Meio Ambiente. Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis. Brasília, 2008.

ICCAT. ICCAT Report 2014-2015 (II). Executive Summary SWO-ATL. p.158-176, 2015.

ICCAT. Informe del comité permanente de investigación y estadísticas (SCRS). Madrid, España. 314 p, 2012.

IGARASHI, M. A. Sinopse da Situação Atual, Perspectivas e Condições de Cultivo para Lagostas Palinuridae. Ciência Animal Brasileira, v. 8, n. 2, p. 151-166, abr./jun. 2007.

IGARASHI, M.A. Avanços no desenvolvimento tecnológico e status atual do cultivo de lagostas (*Panulirus argus*) do ovo ao tamanho comercial. Semina: Ciências Agrárias, v.31, n.1, p. 269-280, 2010.

IJIMA, H. CPUE standardization of the Indian Ocean swordfish (*Xiphias gladius*) by Japanese longline fisheries: Using negative binomial GLMM and zero inflated negative binomial GLMM to consider vessel effect. IOTC-2017-WPB15-19. 32p, 2017.

IVO, C. T. C. Novo estudo sobre o crescimento e idade da lagosta *Panulirus laevicauda* (Latreille), em águas costeiras do Estado do Ceará (Brasil). Arq. Ciên. Mar., v. 15, n. 1, p. 29-32, 1975.

IVO, C. T. C.; FILHO, A. A. F.; SILVA, A. A. da; VIEIRA, R. H. S. F. Cadeia produtiva da lagosta nas regiões Norte e Nordeste do Brasil. Editora RDS. Fortaleza, 2012.

IVO, C. T. C.; PEREIRA, J. A. Sinopse das principais observações sobre as lagostas *Panulirus argus (Latreille)* e *Panulirus laevicauda (Latreille)*, capturadas em águas costeiras do Brasil, entre os estados do Amapá e do Espírito Santo. Boletim Técnico Científico do CEPENE, Tamandaré, v.4, n.1, p. 7-94, 1996.

KACHMAN, S. D. An introduction to generalized linear mixed models. In: Proceedings of a Symposium at the Organizational Meeting for a NCR Coordinating Committee on “Implementation Strategies for National Beef Cattle Evaluation”. Athens. p. 59-73, 2000.

KHAN, M. M. *Xiphias gladius* collected in Pakistan. 2013. Disponível em: <http://www.fishbase.org/photos/UploadedBy.php?autoctr=19777&win=uploaded> Acesso em: 03 abr. 2018.

KING, M. Fisheries Biology, Assessment and Management. Fishing News Books, Oxford, 341p., 1995.

KRISTENSEN, K., NIELSEN, A., BERG, C. W., SKAUG, H., & BELL, B. TMB: automatic differentiation and Laplace approximation, 2015. *arXiv preprint arXiv:1509.00660*.

KRUGER, R. F.; WENDT, L. D.; RIBEIRO, P. B. The effect of environment on development and survival of pupae of the necrophagous fly Ophyra albuquerquei

Lopes (Diptera, Muscidae). Rev. Bras. entomol., São Paulo, v. 55, n.3, p.401-405, 2011.

KRUSKAL, W.H.; WALLIS, W.A. 1952 Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association, 47(260): 583-621.

LEE, Y.; NELDER, J. A. Hierarchical generalized linear models. Journal of the Royal Statistical Society. Series B (Methodological), p. 619-678, 1996.

LETELIER, S.; MELÉNDEZ, R.; CARREÑO, S.; LOPEZ, S.; BARRIA, P. Alimentación y relaciones tróficas del pez espada (*Xiphias gladius Linnaeus*, 1758), frente a Chile centro-norte durante 2005. Lat. Am. J. Aquat. Res., v. 37, n. 1, p. 107-119, 2009.

LIMA, D. P. de; SILVA, L. M. A. da; LIRA, A. C. S. de. Biologia populacional de *Macrobrachium jelskii* (MIERS, 1778) (Crustacea, Decapoda, Palaemonidae) em uma planície inundável na Amazônia Oriental, Brasil. Biota Amazônia, v. 3, n. 2, p. 11-22, 2013.

LIMA, L. T. B. de.; OLIVEIRA, M. R.; NÓBREGA, M. F.; CARVALHO, M. M. de.; CHELLAPPA, S.; OLIVEIRA, J. E. L. Biologia reprodutiva de Bagre marinus (Mitchill, 1815) (Siluriformes: Ariidae) das águas costeiras do Rio Grande do Norte, Brasil. Biota Amazônia (Biote Amazonie, Biota Amazonia, Amazonian Biota), v. 6, n. 4, p. 81-86, 2016.

LIMA, S. A. DE O.; ANDRADE, H. A. Variações nas capturas das lagostas vermelha (*Panulirus meripurpuratus*), verde (*Panulirus laevicauda* LATREILLE, 1817) e sapata (*Scyllarides brasiliensis* RATHBUN, 1906) na costa de Pernambuco. Boletim do Instituto de Pesca, v. 43, n. 2, p. 194-206, 2017.

LIMA, W. M. G. D.; MENDES, N. C. B.; SILVA, B. B. D. Estudo da produção pesqueira e fecundidade de lagostas no Norte do Brasil, Municípios de Bragança e Augusto Corrêa-PA. Biota Amazônia (Biote Amazonie, Biota Amazonia, Amazonian Biota), v. 4, n. 3, p. 48-56, 2014.

LOBATO JUNIOR, D. Análise de diagnóstico em modelos de regressão normal e logísticas. Tese (doutorado). Lavras: UFLA, 2010.

MAGNUSSON, A.; SKAUG, H.; NIELSEN, A.; BERG, C.; KRISTENSEN, K.; MAECHLER, M.; BENTHAM, K. V.; BOLKER, B.; BROOKS, M. Generalized Linear Mixed Models using Template Model Builder, 2017.

MAIA, H. A.; FORMIGONI, H.; SILVA, A. A. da. Negocio de auditoria y cumplimiento del nivel del divulgacion obligatoria durante la convergencia de las normas internacionales de contabilidad en brasil. Revista Brasileira de Gestão de Negócios (Brazilian Journal of Business Management), n.18, p.335, 2012.

MATSUURA, Y. Recursos pesqueiros: avaliação de recursos no Brasil. In: OGAWA M.; KOIKE, J. (Eds.). Manual de pesca. Fortaleza: Gráfica Batista, 1987. p. 67-112.

MAUNDER, M. N.; PUNT, A. E. Standardizing catch and effort data: a review of recent approaches. Fisheries Research, v. 70, n. 2, p. 141-159, 2004.

MAUNDER, M. N.; LANGLEY, A. D. Integrating the standardization of catch-per-unit-of-effort into stock assessment models: testing a population dynamics model and using multiple data types. Fisheries Research, v. 70, n. 2, p. 389-395, 2004.

MAXWELL, K. E.; MATTHEWS, T. R.; BERTELSEN, R. D.; DERBY, C. D. Age and size structure of Caribbean spiny lobster, *Panulirus argus*, in a no-take marine reserve in the Florida Keys, USA. Fisheries Research, v.144, p. 84-90, 2013.

MCCULLAGH, P.; NELDER, J. A. Generalized linear models. Monographs on Statistics and Applied Probabibilty. Chapman and Hall, London, 1989.

MCCULLOCH, E. C.; SEARLE, S. R. Linear and Generalized Linear Mixed Models. Wiley, New York, 2001.

MEDEL, C. A.; SALGADO, S. C. Does the BIC Estimate and Forecast Better than the AIC?. Revista de Análisis Económico–Economic Analysis Review, v. 28, n. 1, p. 47-64, 2013.

MENESES DE LIMA, J.H.; KOTAS, J.E.; LIN, C.F. A historical review of the Brazilian longline fishery and catch of swordfish (1972-1997). Collective Volume of Scientific Papers ICCAT, 54(4): 1329-1257, 2000.

MESQUITA, A. L. L. de. Aspectos cronológicos da reprodução da lagosta *Panulirus argus (Latreille)*, no estado do Ceará (Brasil). Arq. Ciên. Mar, Fortaleza, v. 13, n. 2, p. 77-88, 1973.

MILTON, D. A.; SATRIA, F.; PROCTOR, C. H.; PRASETYO, A. P.; UTAMA, A. A.; FAUZI, M. Environmental factors influencing the recruitment and catch of tropical *Panulirus* lobsters in southern Java, Indonesia. Continental Shelf Research, v. 91, p. 247-255, 2014.

MOURATO, B. L., HAZIN, H., HAZIN, F., CARVALHO, F., DE AMORIM, A. F. Assessing Atlantic sailfish catch rates based on Brazilian sport fishing tournaments (1996-2014). Bol. Inst. Pesca, São Paulo, v. 42, n. 3, p. 625-634, 2016.

NAKAMURA, I. FAO species catalogue. In: FAOs (Ed.), Billfishes of the World: An annotated and illustrated catalogue of marlins, sailfishes, spearfishes and swordfishes known to date. FAO Fish. Synop., Rome, pp. 1-65, 1985.

NASCIMENTO, I. V. Reprodução das lagostas *Panulirus argus* e *Panulirus laevicauda* (Latr.) provenientes de desembarques comerciais em Natal, estado do Rio Grande do Norte. Recife: SUDENE, Divisão de Recursos Pesqueiros, p. 25-34, 1984. (Série: Brasil: Estudos de Pesca, n. 11).

NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. Journal of the Royal Statistical Society, A 135, p. 370-384, 1972.

OLIVEIRA, V. S. DE.; ANDRADE, H. A.; HAZIN, F. H. V.; VIEIRA, A. C. Pesca da lagosta com covos na costa central de Pernambuco: taxas de captura da lagosta e fauna acompanhante. Boletim do Instituto de Pesca, v. 41, n. 2, p. 373-385, 2015.

ONO, K.; PUNT, A. E.; HILBORN, R. Think outside the grids: an objective approach to define spatial strata for catch and effort analysis. Fisheries Research, v. 170, p. 89-101, 2015.

PAIVA, M. P., COSTA, R. S. Comportamento biológico da lagosta *Panulirus laevicauda* (Latreille). Est. Biol. Mar. Univ. Fed. Ceará, Fortaleza, v. 8, n. 1, p. 1-6, 1968a.

PAIVA, M. P., COSTA, R. S. Estudos de biologia da pesca de lagostas no Ceará – Dados de 1964. Arq. Est. Biol. Mar. Univ. Fed. Ceará, Fortaleza, v.5, n. 2, p. 127-150, 1965.

PAIVA, M. P., FONTELES-FILHO, A. A. Sobre as migrações e índices de exploração da *Panulirus laevicauda* (Latreille). Arq. Est. Biol. Mar. Univ. Fed. Ceará, Fortaleza, v. 8, n. 1, p. 15-23, 1968.

PAIVA, M. P.; COSTA, R. S. da. Estudo de biologia da pesca de lagosta no Ceará – Dados de 1966. Arq. Est. Biol. Mar. Univ. Ceará, Fortaleza, v.7, n. 2, p.147-171, 1967.

PAIVA, M. P.; COSTA, R. S. da. Estudo de biologia da pesca de lagosta no Ceará – Dados de 1967. Arq. Est. Biol. Mar. Univ. Ceará, Fortaleza, v.8, n. 2, p. 187-210, 1968.

PAIVA, M. P.; COSTA, R. S. da. Estudo de biologia da pesca de lagosta no Ceará – Dados de 1968. Arq. Est. Biol. Mar. Univ. Ceará, Fortaleza, v.9, n. 2, p. 133-146, 1969.

PAIVA, M. P.; COSTA, R.S. da. Estudo de biologia da pesca de lagosta no Ceará – Dados de 1965. Arq. Est. Biol. Mar. Univ. Ceará, Fortaleza, v.6, n. 2, p. 167-193, 1966.

PAIVA, M. P; COSTA, R. S. da. Estudo de biologia da pesca de lagosta no Ceará – Dados de 1962. Arq. Est. Biol. Mar. Univ. Ceará, Fortaleza, v.3, n.1, p. 27-52, 1963.

PAIVA, M. P; COSTA, R. S. da. Estudo de biologia da pesca de lagosta no Ceará – Dados de 1963. Arq. Est. Biol. Mar. Univ. Ceará, Fortaleza, v.4, n. 2, p. 45-70, 1964.

PAIVA, M. P; COSTA, R. S. Estudo de biologia da pesca de lagosta no Ceará – Dados de 1969. Arq. Est. Biol. Mar. Univ. Ceará, Fortaleza, v.10, n. 2, p. 131-142, 1970.

PAIVA, M. P; SILVA, A. B. da. Estudo de biologia da pesca de lagosta no Ceará – Dados de 1961. Arq. Est. Biol. Mar. Univ. Ceará, Fortaleza, v.2, n.2, p. 21-34, 1962.

PAIVA, M.P. Estudo de biologia da pesca de lagosta no Ceará – Dados de 1965. Arq. Est. Biol. Mar. Univ. Ceará, Fortaleza, v.6, n.2, p. 147-165, 1966.

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.

PAULA, G. A. Modelos de regressão: com apoio computacional. São Paulo: IME-USP, 2013.

PÉREZ-GONZÁLEZ, R.; VALADEZ, L. M.; RODRÍGUEZ-DOMÍNGUEZ, G.; ALBERTO, E. Seasonal Variation in Brood Size of the Spiny Lobster *Panulirus gracilis* (Decapoda: Palinuridae) in Mexican Waters of the Gulf of California. Journal of Shellfish Research, v.31, n.4, p. 935-940, 2012.

PINHEIRO, J. C.; BATES, D. M. Linear mixed-effects models: basic concepts and examples. Mixed-effects models in S and S-Plus, p. 3-56, 2000.

PORTO, V. R. Alguns aspectos reprodutivos em gônodas de lagosta *Panulirus laevicauda* (Latreille). Fortaleza: UFC, Dissertação (Graduação), 20p. 1992.

POSTUMA, F.A.; GASALLA, M. D. L. A. Reproductive activity of the tropical arrow squid Doryteuthis pleiaround São Sebastião Island (SE Brazil) based on a 10-year fisheries monitoring. Fisheries Research, v. 152, p. 45– 54, 2014.

QUINN II, T. J.; HOAG, S. H.; SOUTHWARD, G. M. Comparison of two methods of combining catch-per-unit-effort from geographic regions. Can. J. Fish. Aquat. Sci. 39: 837-846, 1982.

QUINN, T. J.; DERISO, R. B. Quantitative Fish Dynamics. Oxford University Press, New York, 1999.

RIGBY, R. A.; STASINOPOULOS D. M. Generalized additive models for location, scale and shape (with discussion). *Appl. Statist.*, v. 54, part 3, pp 507-554, 2005.

RIGBY, R. A.; STASINOPOULOS, D.M. A semi-parametric additive model for variance heterogeneity. Statistical Computing 6, 57-65, 1996a.

RIGBY, R. A.; STASINOPOULOS, D.M. Mean and dispersion additive models. In Statistical Theory and Computational Aspects of Smoothing. Eds: Härdle, W. and Schimek, M.G., 215-230. Heidelberg: Physica, 1996b.

ROSSI, R. M.; GUEDES, T. A.; JANEIRO, V.; MARTINS, E. N. Estimação do fator de condição de peixes da espécie *Tracydoras paraguayensis*: uma perspectiva bayesiana. Acta Scientiarum - Animal Sciences, Maringá, v. 29, n. 1, p. 85-92, 2007. DOI:10.4025/actascianimsci.v29i1.263

SAHAI, H.; AGEEL, M. The Analysis of Variance: Fixed Random and Mixed Models. Boston: Birkhauser, 2000.

SANTANA, W.; IVO, C. T. C.; DIAS-NETO, J.; DUARTE, L. F. A.; PINHEIRO, M. A. A.; BOOS, H.; PINHEIRO, A. P.; ALMEIDA, A. O.; HERNÁEZ, P.; COELHO, P. A. Avaliação Lagostas-de-espinho (Decapoda: Palinuridae), Cap. 21: p. 268-283. In: Pinheiro, M.A.A. & Boos, H. (Org.). Livro Vermelho dos Crustáceos do Brasil: Avaliação 2010-2014. Porto Alegre, RS: Sociedade Brasileira de Carcinologia - SBC, 466 p., 2016.

SARMENTO, J. L. R.; PIMENTA FILHO, E. C.; ABREU, U. G. P. de.; RIBEIRO, M. N.; SOUZA, J. E. R. de. Prolificidade de caprinos mestiços leiteiros no semiárido nordestino. R. Bras. Zootec., Viçosa, v. 39, n. 7, p. 1471-1476, 2010.

SCHWARZ, G. Estimating the dimension of a model. Annals of Statistics. 6: 461-464., 1978.

SEARLE, S. R.; CASELLA, G.; MCCULLOCH, C. E. Variance components. New York: John Wiley & Sons, 501p, 1992.

SHEARD, K. The Marine Crayfishes (Spiny Lobsters), Family Palinuridae, of Western Australia. Bulletin of the Commonwealth Scientific and Industrial Research Organization, Melbourne, n. 247, p.1-45, 1949.

SHINOZAKI-MENDES, R. A.; MANGHI, R. F.; LESSA, R. Comparative study of the molting cycle of wild and reared swimming crabs Callinectes danae (Crustacea: Portunidae). Journal of applied ichthyology, v. 30, n. 3, p. 502-506, 2014

SIEGEL, S.; CARMONA, S. I. C. Estatística não-paramétrica para ciências do comportamento. 2. ed. São Paulo: Artmed, 2006. 448 p.

SILVA, A. C.; FONTELES-FILHO, A. A. Avaliação do defeso aplicado a pesca da lagosta no nordeste do Brasil. Editora Expressão Gráfica, Fortaleza, 110 pp, 2011.

SILVA, A. P. da. Pesca artesanal brasileira. Aspectos conceituais, históricos, institucionais e prospectivos. Palmas: Embrapa Pesca e Aquicultura-Boletim de Pesquisa e Desenvolvimento (INFOTECA-E), 2014.

SILVA, K. C. DE A.; BOTELHO, K. B.; CINTRA, I. H. A.; ARAGÃO, J. A. N. Parâmetros reprodutivos da lagosta-vermelha, *Panulirus argus* (LATREILLE, 1804), na costa norte do Brasil. Bol. Téc. Cient. CEPNOR, Belém, v. 8, n. 1, p. 27-39, 2008.

SMITH, T. G. W. The Spiny Lobster Industry of Florida. State of Florida Board of Conservation Educational Series, Miami, n.11, p.1-36, 1958.

SMITH, T. G. W. The Spiny Lobster Industry of the Caribbean and Florida. Caribbean Research Council Fishers Series, Port-of-Spain, n.3, p.1-58, 1948.

SOARES, C. N. C. Reproductive season of the Caribbean spiny lobster *Panulirus argus* in the coast of Northern Brazil. Fishbyte, Manila, v.8, n. 3, p. 27-28, 1990.

SOARES, C. N. C., CAVALCANTE, P. P. L. Caribbean spiny lobster (*Panulirus argus*) and smoothtail spiny lobster (*Panulirus laevicauda*) reproductive dynamics on the Brazilian Northeastern coast. *FAO Fish Rep*., Rome, n. 327, p. 200-217, 1985.

SOARES, C. N. C.; FONTELES-FILHO, A. A. Época de reprodução da lagosta-verde, *Panulirus laevicauda (latreille)*, no estado do Ceará, Brasil. Arq. Ciên. Mar, Fortaleza, v. 33, p. 43-50, 2000.

SPANIER, E., LAVALLI, K. L., GOLDSTEIN, J. S., GROENEVELD, J. C., JORDAAN, G. L., JONES, C. M., PHILLIPS, B. F.; BIANCHINI, M. L.; KIBLER, R. D.; MALLOL, D. D. S.; GOÑI, R.; MEEREN, G. I. V. D.; AGNALT, A. L.; BEHRINGER, D. C.; KEEGAN, W. F.; JEFFS, A. A concise review of lobster utilization by worldwide human populations from prehistory to the modern era. ICES Journal of Marine Science, v. 72, n. suppl_1, p. i7-i21, 2015.

SPARRE, P.; VENEMA, S. C. Introdução à avaliação de mananciais de peixes tropicais. Parte 1- Manual. Roma: FAO. 404p, 1987. (Documento técnico sobre as pescas. nº306/1)

STASINOPOULOS D. M.; RIGBY R.A. Generalized additive models for location scale and shape (GAMLSS) in R. *Journal of Statistical Software*, Vol. 23, Issue 7, 2007.

STASINOPOULOS D. M.; RIGBY R.A.; HELLER G.; VOUDOURIS V.; DE BASTIANI F. *Flexible Regression and Smoothing: Using GAMLSS in R*, Chapman and Hall/CRC, 2017.

SUNDELÖF, A.; BARTOLINO, V.; ULMESTRAND, M.; CARDINALE, M. Multiannual fluctuations in reconstructed historical time-series of a European lobster (Homarus gammarus) population disappear at increased exploitation levels. PloS one, v. 8, n. 4, 2013.

TANAKA, K. R.; BELKNAP, S. L.; HOMOLA, J. J.; CHEN, Y. A statistical model for monitoring shell disease in inshore lobster fisheries: A case study in Long Island Sound. PloS one, v. 12, n. 2, p. e0172123, 2017.

TRIOLA, M. F. Introdução à Estatística - Atualização da Tecnologia, 11 ed. Rio de Janeiro: LTC, 2013. 706 p.

TSAI, W. P.; LIU, K. M. Updated and revised standardized catch rates of blue sharks caught by the taiwanese longline fishery in the atlantic ocean. Collect. Vol. Sci. Pap. ICCAT, 72(5), 1134-1149, 2016.

UEYANAGI, S. et al. Distribution, spawning, and relative abundance of billfishes in the Atlantic Ocean. Bull. Far. Seas Fisheries Research Laboratory, v. 3, p. 15-45, 1970.

VAN BUUREN, S.; FREDRIKS, M. Worm plot: a simple diagnostic device for modelling growth reference curves. Statistics in Medicine, v. 20, 1259-1277, 2001.

VENABLES, W. N.; DICHMONT, C. M. GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research. Fish. Res., v. 70, n. 2/3, p. 319-337, 2004.

VENABLES, W. N.; RIPLEY, B. D. Modern Applied Statistics with S. Fourth Edition. Springer, New York. 495p. 2002.

VENABLES, W. N.; RIPLEY, B. D. Modern Applied Statistics with S. Fourth Edition. Springer, New York. 495p. 2002.

VIEIRA, A. A. Modelagem da taxa de captura para o espadarte capturado pela frota espinheleira brasileira no Atlântico Sul: enfoque transversal versus longitudinal. Tese de Doutorado. Universidade Federal Rural de Pernambuco. Recife, 2015. 106p.

VRIEZE, S. I. Model selection and psychological theory: A discussion of the differences between the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Psychological Methods. 2012; 17(2):228-243. doi:10.1037/a0027127.

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.

WALD, A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. Transactions of the American Mathematical society, v. 54, n. 3, p. 426-482, 1943.

WALTER, J.; LAURETTA, M. Standardized catch rates for bigeye tuna (*Thunnus obesus*) from the United States pelagic longline fishery. Collect. Vol. Sci. Pap. ICCAT, 72(2), 311-342, 2016.

WANG, H.; SHIH, C.; HSU, C. Standardized catch per unit effort of Pacific bluefin tuna (Thunnus orientalis) by general linear model for Taiwanese small-scale longline fishery in the southwestern North Pacific Ocean. In: Pacific Bluefin Tuna Working Group Meeting, ISC, 2014.

Apêndice

APÊNDICE A – Artigo científico enviado para o ICCAT

ESTIMATIONS OF STANDARDIZED CATCH RATES OF SWORDFISH (*XIPHIAS GLADIUS*) CAUGHT BY BRAZILIAN FLEET AS CALCULATED USING FIXED AND RANDOM EFFECTS

Evelyn Souza Chagas Oliveira[1], Vívian Graziela Oliveira Carneiro[2], Silvaneide Luzinete Rodrigues[3], Humber Agrelli Andrade[4]

SUMMARY

Estimations of standardized CPUE were calculated following three approaches: A) year was included in the models as main fixed effect only; B) year was included in the models as main fixed effect and also in fixed effect interactions; and C) year was included in the models as main fixed effect and in random effect interactions. We have used Generalized Linear Models (GLM) and Generalized Linear Mixed Models (GLMM) with Poisson distribution and logarithm link function. The response variable was the catch (number of fish), explanatory variables were year, area, flag and quarter, and logarithm of effort was included as offset Convergence of GLMM was difficult to achieve probably due to the lack of balance of the Brazilian dataset Time trend of the three standardized CPUE time series were not different. However, it is important to highlight that in this preliminary study we have analyzed only part of Brazilian dataset using simples model with few explanatory variables.

KEYWORDS

Catch/effort, Stock assessment, Abundance, Biomass, Fishing effort

1. Introduction

In the stock assessments the main objectives are to estimate abundance and the potential production of the population as the balance of recruitment, somatic growth and mortality (Hilborn and Walters, 1992; Quinn and Deriso, 1999). Stock assessment analyses are useful for decision makers and the fishery management. However, the stock size or abundance is often a latent variable, in the sense it can not be directly measured. Therefore, in most of stock assessment analyses abundance is replaced by indices, which are supposed to be a proxy of abundance (Sparre and Venema, 1987; Hilborn and Walters, 1992; King, 1985).

Often the commercial Catch per Unit Effort (CPUE) data is used to calculate relative abundance indices for highly migratory fishery resource like tuna and tuna like species. However, the catch rate (or CPUE) changes due to the abundance but also due to several factors like fishermen strategy or fishery technology. The analyses of CPUE to eliminate the effects of different factors and to estimate a time series of indices of relative abundance is usually called "standardization". Often the standardization of CPUE is based on estimations of parameters of the factor "year " as calculated using Generalized Linear Models (GLM) (Maunder and Punt, 2004).

If year is considered only as fixed main effect (approach A), the estimations of standardized CPUEs are straightforward. However, if year is included in the model as fixed main effect but also in fixed effect interactions (approach B), the estimations of standardized CPUEs are more complex. In this case it is necessary some supposition concerning the weights to calculated a weighted average over the interactions (*e.g*. least square means) (*e.g.* Quinn II *et al.*, 1982; Campbell, 2004). However, if year is included as main fixed effect and also in random effect interactions (approach C), the calculations are simple, because the interactions are not necessary to estimate the standardized CPUE. In the past the approach A was popular, then the approach B replaced the approach A, and finally, the popularity of the approach C has increased. However, "Whether an effect should be considered fixed or random will depend on way the experimental treatments (levels of a factor) are selected and the kind of inferences one wishes to make from the analysis" (Sahai and Ageel, 2000). "… clear answers to the question 'fixed or random?' are not necessarily the norm" (Searle *et al.*, 1992). While in experimental studies it is more easy to decide if a main effect will considered fixed or random, in observational studies the decision between fixed or random may be not clear, specially if we are talking about an interaction that includes a factor of interest (*e.g.* year) and other factor that is not of interest (*e.g.* quarter). Hence the choice of one among the approaches A, B or C may be not easy. In this paper the objective was not to investigate and discuss which of the three approaches is best in the statistical and fishery theoretical grounds. We just called attention to the point that it maybe be difficult to answer the question "fixed or random?" when calculating standardized CPUE. We believe that investigations on the issue should be encouraged in the future. Here we show the results of an exercise to investigate what are the differences between standardized CPUEs calculated using the three approaches (A, B, and C) often used in the recent years. Our case of study is the swordfish (*X. gladius*) caught in the South Atlantic by the Brazilian fleet.

2. Materials and Methods

2.1. Data

Dataset we have analyzed is the "Banco Nacional de Atuns e Afins (BNDA)" of the Tuna and Tuna-Like Brazilian Scientific Committee of the Brazilian government. In this working paper we have analyzed the swordfish caught in the South Atlantic by the Brazilian fleet between 1990 a 2012. Brazilian fleet includes national vessels (BRA) but also vessels leased from more than fifteen countries. However we retained for analyses leased boats from Spain (BRA-ESP), Honduras (BRA-HND), Japan (BRA-JPN), Panama (BRA-PAN), China-Taipei (BRA-TAI) e Saint Vicent & Grenadines (BRA-VCT) (**Table 1**), because the sample size (number of longline sets) are high for these flags. Longline sets with missing values for

catch or effort were discarded. Non sampling errors (*e.g.* fishing sets located on or too close to the land) were also discarded. After the exploratory analyses to identify and discard errors and suspected data only 35055 fishing sets were retained for the analysis concerning the standardization of the CPUE.

2.2. Analysis

In order to estimate standardized CPUE using the approach A (year is considered only as main fixed effect) and B (year is included in the models as main fixed effect but also in fixed effect interactions) we have used conventional GLM models. Generalized linear models were first published by Nelder and Wedderburn (1972). The GLMs have three components:
a) random – y is usually a vector of length n with independent realizations of a random variable Y with expectation μ. The distribution of Y is one of the exponential family with canonical structure;
b) linear predictor – systematic linear part of the model $\eta = X\beta$, in which $\eta_i = X_i\beta$ is the component concerning the *i*th observation; and
c) link function *g( )* – a function to link the expectation of Y and the linear predictor as $g(\mu_i) = \eta_i = x_i^T\beta$. Details on the structure and calculations of GLM can be find in textbooks like Dobson (2008).

In the approach C the fator year is included in the models as main fixed effect and also as part of random interactions. In order to estimate standardized CPUE we have a Generalized Linear Mixed Model (GLMM). The components of the GLMM are similar to those of the GLM (see above), but the linear predictor is $\eta = X\beta + Zu$, in which Z is a design matrix like X, and u is component to account for the random effect. In the GLMM the variance function $v(\mu_i, \emptyset)$ is used to model the residual variability, and the estimations are calculated using maximum likelihood (ML) algorithms. Details concerning GLMM can be found in Breslow and Clayton (1993), and in Pinheiro and Bates (2000).

We have assumed that the response variable (S), which is the catch as reported in number of fish, follows a Poisson distribution. We have used a logarithmic link function. The explanatory variables were: flag (F), year (Y), quarter (T) and fishing area (A). The levels of factor area were: North (northward of 10°S), Central (between 10°S and 25 ^{o}S), South (southward of 25 ^{o}S) (**Figure 1**). These levels of area were selected based on the data balance and on preliminary attempts to fit the GLMM. GLMM did not converge when we tried out other alternatives concerning levels of area. The unit of effort is the number of hooks of the longline. Logarithm of effort was used as offset It is important to stress that the intention was just to compare approaches concerning the way we deal with year in interactions. We were not trying to calculate a useful standardized CPUE, hence we did not carried out tests concerning different explanatory variables, probability distributions and the excess of zeros.

In order to choose the order the explanatory variables are included in the models, the first step was to fit one simple model with only one of the factors year, quarter and area at a time. Residual deviance (Nelder and Wedderburn, 1972) was the selected criterion to order the explanatory variables. In addition to the main effects we tried out to include all the first order interactions in the models when using the approaches B and C. In the approach A only the factors quarter, flag and area were included in interactions. However, due to convergence issues only two interactions were considered in the GLMM (approach C). Finally, the selection of explanatory variables and interactions in the three approaches, were based on the Akaike Information Criterion (AIC) (Akaike, 1974), and on the Bayesian Information Criterion (BIC) (Schwarz, 1978).

Conventional residual analysis (e.g. fitted x residual graphs, leverage and Cook's distances calculations) were used to assess the quality of the model fittings (Dobson, 2008; Venables e Ripley, 2002). Pseudo-R^2 (McFadden, 1974) were calculated and the three standardized CPUEs (A, B e C) were compared. All the analyses and calculations were carried out using the software R (R 3.3.1 Core Team, 2016) and the functions *glm( )* (GLM) and *glmer( )* of the package *lmer4* (Bates *et al.*, 2015).

3. Results

Brazilian fleet operations cover a large part of the South Atlantic between longitudes 0º and 55ºW, and the latitudes 5ºN and 50ºS (**Figure 1**). The values showed in the map are summation of the number of fish caught from 1990 to 2012 as reported in the dataset High catches (> 5000) were observed all over the South Atlantic, but mainly in the north and central areas. Boxplots of CPUE are shown in **Figure 2**. Notice the high variability of CPUE in 1998. Notice also that catch rates have increased after 2003 onwards. Overall values of CPUE in the first and fourth quarter tend to smaller than those of the second and third quarter. Catch rates were in general higher in the central area.

Selected models based on AIC and BIC for the approaches A, B, and C are shown in **Table 2**. Results of likelihood ratio hypothesis tests (approaches A and B) and of Wald test (approach C) for the selected models (**Table 3**) indicate that all the explanatory variables are important to model the variation of catch rates, but he factor quarter of the GLMM (approach C) is an exception.

Diagnostic of residuals of GLM model fitted following approach A are in **Figure 3**. Residuals were approximately homoscedastic (**Figure 3A**). There were not evidences that the model is biased in the sense the expectations of residuals are close to zero (**Figure 3B**). Distributions of the residuals are approximately normal but there are violations in the tails (**Figure 3C**). High values of residuals are associated a high leverage values (diagonal of hat matrix) (**Figure 3D**). Overall Cook's distances were lower than 0.5. Diagnostic of residuals calculated for the models fitted following the approach B (**Figure 4**) and C (**Figure 5**) were similar to those calculated for the model fitted based on the approach A (**Figure 3**).

Summary of estimations of the model fitted with year as main effect only (approach A) are in **Table 4**. Notice that most of estimations indicate that we reject the null hypothesis (parameters equal to zero). Estimations of parameters of factor flag were negative, but the estimation of level BRA-ESP is the exception. Hence there are evidence that the expectation of CPUE of vessels leased from Spain are higher than for national boats (base level). However the CPUE of national boats (BRA) and leased boats from Spain (BRA-ESP) are higher than the CPUEs of boats leased from other countries. Estimations of expectations of CPUE for 1991, 1998 and 2003 onwards were higher than the expectation of CPUE of the base level (year 1990). Notice also that the expectation of catch rates for the north area was higher than for the other regions. The estimations also indicate that the expectation was higher for the first quarter than for the other periods of the year.

Estimations of the parameters of models fitted with year included as main fixed effect and also in fixed effect interactions (approach B) are in **Table 5**. Similarly, the estimations of parameters of the mixed model with year included as main fixed effect and also in random effect interactions (approach C) are in **Table 6**. Estimations of the variance components for the approach C are in **Table 7**. Interpretations of the estimations showed in **Tables 5 and 6** (approaches B and C) can be made based on the values, signals and hypothesis tests just like we did in the above paragraph for the **Table 4** (approach A). However, it is necessary to account for the interaction when assessing the estimations of the model fitted following the approach B. Usually some kind of weighted average calculation (e.g. least square means) is necessary to estimate the separated effect of a main factor also included in the model as interaction. In order to extract the separated effect of year factor as calculated with the approach B we assumed equal weights to all months, flags and areas when dealing with interactions year:month, year:flag and year:area. We choose this alternative in this exercise. However, it is important to stress in some situations there are not a clear answer concerning the question about which weights should be used. To assume equal weight to month may be appropriate (D. S. Butterworth apud Maunder and Punt, 2004), but to select weights for other factors, like area, may be not as straightforward (Maunder and Punt, 2004).

Standardized CPUEs calculated as calculated based on approaches A, B and C) are shown in **Figure 6**. Estimations were scaled as $z = (x - \bar{x})/s$ to make comparison easier. Overall there are not large among the three time series, but estimations for 1991, 1994 and 2008 are the exceptions. The three standardized catch rates as calculated decreased slightly between 1990 and 2001, but the estimations increased in the beginning of 2000's. Standardized catch rates in the end of the three time series were high.

4. Discussion

The convergence of the mixed model (GLMM) arose as an important issue. Some decisions concerning the model structure (*e.g.* levels of factor area, number of interactions included in the model) were selected in an attempt to achieve convergence. Convergence of mixed models may be not easy when the dataset and the design matrix are no not balanced. Estimators of maximum likelihood (ML) are usually sufficient statistics, consistent and asymptotically normally distributed and efficient (Harville, 1977), but ML estimators of the variance components estimations of the fixed effects. In spite their biases, ML estimators are popular due to their asymptotic properties (Searle *et al.*, 1992), and because there are empirical evidence that they perform better than other estimators (Swallow and Monaham, 1984). However the estimations are computationally demanding. In addition convergence difficulties increase as the dimension of the models increase specially if the data is not well balanced (*e.g.* Brazilian dataset). Studies concerning comparisons of different methods to estimate the variance components of mixed models have been made since mid 1990's (*e.g.* Resende *et al.*, 1996; Marcelino and Lemma, 2000). However, by the moment the ML estimator is probable one of the best alternative despite of the convergence issue.

Often the estimations of factor year is assumed to reflect annual variation of biomass in most of the papers concerning standardization of CPUE (Maunder and Punt, 2004). There are two alternatives to cope with effects of interactions between year and other factors when estimating the standardized CPUE time series: 1) To ignore them by do not including year in fixed effect interactions (approach A) or by including year only in random effect interactions (approach C); or 2) To take them into account (approach B), but some kind of weighted average might be calculated over the interaction terms. In the ICCAT meetings both approaches concerning the inclusion of year in the interactions (fixed or random) have been used (*e.g.* fixed - Hazin *et al.*, 2007; Carneiro *et al.*, 2015; Tsai and Liu, 2016; random - Chang, 2003; Arocha *et al.*, 2016; Walter and Lauretta, 2016). However, the motivations for the choice for one or another alternative (to ignore or to take interactions into account) to standardize CPUE have been not discussed. In most of the papers concerning standardization of CPUE it is not clear if the choice of one among the three approaches (A, B or C) was based on theoretical grounds or if it was a convenient choice to make the calculations easier. If the time trend of estimations of standardized CPUEs time series (approaches A, B and C) are very different, decisions and diagnostics concerning the status of the stock may change.

Here we showed the results of a simple exercise we did with part of the Brazilian dataset We assessed the main time trend signal of estimation of standardized CPUE, but we did analyzed the precision of the estimations (*e.g.* standard errors). The results indicate that time trends of standardized CPUEs as calculated based in the approaches A, B and C were not that different. This is an indication that in some situations the choice of the methodology will not strongly affect the understanding about the status of the stock. However, it is important to stress that we tried out a simple model (few explanatory variables) for a fleet (*i.e*. Brazilan) only. Similar investigations with other datasets and fleets should be encouraged to investigate if the differences among the estimations are as low as they were in this exercise with Brazilian fleet.

5. Remarks

- Convergence of mixed models (GLMM) was an issue in the analysis of Brazilian dataset which is unbalanced;

- In some situations to ignore or to include interactions with year as fixed or random effect does not result in quite different standardized CPUE time series.

References

Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control., Boston, v.19, n.6, p.716-723.

Arocha, F., Ortiz, M., Marcano, J. H. 2016. Standardized catch rates for blue shark (Prionace glauca) from the venezuelan pelagic longline fishery in the caribbean sea and adjacent waters of the north Atlantic ocean: period 1994-2013. Collect. Vol. Sci. Pap. ICCAT, 72(4), 1033-1043.

Bates, D.; Maechler, M.; Bolker, B.; Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Breslow, N. E and Clayton, D. G. 1993. Approximate inference in generalized linear mixed models. Journal of the American Statistical Association 88, p. 9-25.

Campbell, R. 2004. CPUE standardization and the construction of indices of stock abundance in a spatially varying fishery using general linear models. Fish. Res. 70: 209-227.

Carneiro, V., Fialho, E., Andrade, H. A. 2015. Updated standardized catch rates for skipjack tuna (*Katsuwonus pelamis*) caught in the southwest of South Atlantic Ocean. Collect. Vol. Sci. Pap. ICCAT, 71(1), 306-316.

Chang, S. K. 2003. Analysis of Taiwanese white marlin catch data and standardization of catch rates. ICCAT Col. Vol. Sci. Pap, 55(2), 453-466.

Dobson, A. J. 2008. An Introduction to Generalized Linear Models, 2 ed. Boca Raton: Chapman and Hall, 307p.

Harville, D. A. 1977. Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. Journal of the American Statistical Association, Alexandria, v.72, n.358, p. 320-338.

Hazin, F. H., Hazin, H. G., Travassos, P., da Mata Oliveira, I. 2007. Standardized catch per unit of effort of white marlin, *Tetrapturus albidus*, and blue marlin, Makaira nigricans, caught by Brazilian tuna longline fleet Col. Vol. Sci. Pap. ICCAT. SCRS/2006/065.

Hilborn, R. and Walters, C.J. 1992. Quantitative Fisheries Stock Assessment.Chapman & Hall, London.

King, M. 1995. Fisheries biology, assessment and management. Fishing News Books, Oxford, 341p.

Marcelino, S. D. do R. and Iemma, A. F. 2000. Métodos de estimação de componentes de variância em modelos mistos desbalanceados. Scientia Agricola, Piracicaba, v. 57, n. 4, p. 643-652. ISSN 1678-992X. URL http://dx.doi.org/10.1590/S0103-90162000000400008.

Maunder, M. N. and Punt, A. E. 2004. Standardizing catch and effort data: a review of recent approaches. Fisheries Research, v. 70, n. 2, p. 141-159.

McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. In F. Zabembka (ed.), Frontiers in Econometrics. New York: Academic Press.

Nelder, J. A. and Wedderburn, R.W.M. 1972. Generalized linear models. Journal of the Royal Statistical Society, A 135, p. 370-384.

Pinheiro, J. C. and Bates, D. M. 2000. Linear mixed-effects models: basic concepts and examples. Mixed-effects models in S and S-Plus, p. 3-56.

Quinn II, T. J.; Hoag, S. H.; Southward, G. M. 1982. Comparison of two methods of combining catch-per-unit-effort from geographic regions. Can. J. Fish. Aquat. Sci. 39: 837-846.

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

R CORE TEAM. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Resende, M. D. V. de; Prates, D. F.; Yamada, C. K.; Jesus, A. de. 1996. Estimação de componentes de variância e predição de valores genéticos pelo método da máxima verossimilhança restrita (REML) e melhor predição linear não viciada (BLUP) em Pinus. Boletim de Pesquisa Florestal, v. 32, n. 33, p. 18-45.

Sahai, H.. and Ageel, M. 2000. The Analysis of Variance: Fixed Random and Mixed Models. Boston: Birkhauser.

Schwarz, G. 1978. Estimating the dimension of a model. Annals of Statistics. 6: 461-464.

Searle, S.R.; Casella, G.; McCulloch, C.E. 1992. Variance components. New York: John Wiley & Sons, 501p.

Sparre, P. and Venema, S. C. 1987. Introdução à avaliação de mananciais de peixes tropicais. Parte 1-Manual. Roma: FAO. 404p. (Documento técnico sobre as pescas. nº306/1)

Swallow, W. H. and Monahan, J. F. 1984. Monte Carlo comparison of ANOVA, MINQUE, REML, and Ml estimators of variance components. Technometrics, vol. 26, 47-57.

Tsai, W. P. and Liu, K. M. 2016. Updated and revised standardized catch rates of blue sharks caught by the taiwanese longline fishery in the atlantic ocean. Collect. Vol. Sci. Pap. ICCAT, 72(5), 1134-1149.

Venables, W. N. and Ripley, B. D. 2002. Modern Applied Statistics with S. Fourth Edition. Springer, New York. 495p.

Walter, J. and Lauretta, M. 2016. Standardized catch rates for bigeye tuna (*Thunnus obesus*) from the United States pelagic longline fishery. Collect. Vol. Sci. Pap. ICCAT, 72(2), 311-342.

Table 1 – Number of fishing sets by flag retained for analysis.

FLAG	Frequencies
BRA	8205
BRA-ESP	9581
BRA-HND	1238
BRA-JPN	994
BRA-PAN	2594
BRA-TAI	5526
BRA-VCT	6917

BRA= Brazil; BRA-ESP = Brazil – Spain; BRA-HND = Brazil – Honduras; BRA-JPN = Brazil – Japan; BRA-PAN = Brazil – Panama; BRA-TAI = Brazil – China Taipei; BRA-VCT = Brazil – St Vicente & Grenadines.

Table 2 – Models selected to standardize the catch per unit effort of *Xiphias gladius* caught by Brazilian fleet from 1990 to 2012. S – number of swordfish caught; F – vessel flag; Y – year; A –area; T – quarter. AIC – Akaike Information Criterion; BIC – Bayesian Information Criterion; DF – degrees of freedom; Dev – Deviance; Res. Dev. – Residual Deviance; Loglik - Log-likelihood statistic.

Model	AIC	BIC	Dev.	GL	Dev. Res.	GL Res.	Pseudo-R^2	Loglik
GLM								
(A) S~F+Y+A+T	370430	370718.2	557907	35054	281137	35021	49,61%	-185181.2
(B) S~F+Y+A+T+Y:T+Y:A	344840	345949.2	557907	35054	255353	34924	54,23%	-172289.1
GLMM								
(C)S~F+Y+A+T+(1|Y:T)+(1|Y:A)	345475.5	345780.3	557907	35054	345403.5	35019	38,09%	-172701.8

(A) Model with year as fixed effect only; (B) Model with year as a fixed effect and also in fixed-effect interaction; (C) Model with year as fixed effect and also in random-effect iteration

Table 3 - Analysis of deviance of the fitted models fitted. Approaches: (A) year as fixed effect only; (B) year as a fixed effect and also in fixed-effect interaction; (C) year as fixed effect and also in random-effect iteration. Likelihood Ratio tests (LR) were used for approaches A and B, and Wald test was used for approach C. The factors were: F –flag; Y – year; A –area; T – quarter. DL – degrees of freedom; Res. Dev. – Residual Deviance.

GLM			
	LR Chisq	GL	Pr(>Chisq)
Approach A			
F	145146	6	< 2.2e-16
Y	31531	22	< 2.2e-16
A	3347	2	< 2.2e-16
T	996	3	< 2.2e-16
Approach B			
F	105419	6	< 2.2e-16
Y	31531	22	< 2.2e-16
A	4196	2	< 2.2e-16
T	874	3	< 2.2e-16
Y:T	13050	62	< 2.2e-16
Y:A	10414	35	< 2.2e-16
GLMM			
	Chisq	GL	Pr(>Chisq)
Approach C			
F	87404.58	6	< 2.2e-16
Y	41.13	22	0.007955
A	19.61	2	5.508e-05
T	5.09	3	0.165611

Table 4 – Estimations of parameters (approach A). F – flag; Y – year; A –area; T – quarter.

Coeficientes	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-4.93279	0.02223	-221.895	< 2e-16 ***
(F) BRA-ESP	0.468031	0.00565	82.84	< 2e-16 ***
(F) BRA-HND	-0.26306	0.012406	-21.205	< 2e-16 ***
(F) BRA-JPN	-0.98615	0.01777	-55.496	< 2e-16 ***
(F) BRA-PAN	-2.53471	0.017957	-141.156	< 2e-16 ***
(F) BRA-TAI	-1.27608	0.009427	-135.37	< 2e-16 ***
(F) BRA-VCT	-1.862	0.009529	-195.412	< 2e-16 ***
(Y) 1991	0.417579	0.026293	15.882	< 2e-16 ***
(Y) 1992	-0.39842	0.027894	-14.283	< 2e-16 ***
(Y) 1993	-0.27658	0.04566	-6.057	1.38e-09 ***
(Y) 1994	-0.26339	0.029646	-8.884	< 2e-16 ***
(Y) 1995	-0.01304	0.026167	-0.498	0.61815
(Y) 1996	-0.22848	0.036622	-6.239	4.41e-10 ***
(Y) 1997	-0.42632	0.025906	-16.456	< 2e-16 ***
(Y) 1998	0.605429	0.023392	25.882	< 2e-16 ***
(Y) 1999	-0.25526	0.02357	-10.83	< 2e-16 ***
(Y) 2000	-0.15588	0.022958	-6.79	1.12e-11 ***
(Y) 2001	-0.62953	0.023439	-26.858	< 2e-16 ***
(Y) 2002	-0.35116	0.025379	-13.837	< 2e-16 ***
(Y) 2003	0.145584	0.027749	5.246	1.55e-07 ***
(Y) 2004	0.203979	0.023816	8.565	< 2e-16 ***
(Y) 2005	0.361746	0.023088	15.668	< 2e-16 ***
(Y) 2006	0.392119	0.023479	16.701	< 2e-16 ***
(Y) 2007	0.322761	0.025364	12.725	< 2e-16 ***
(Y) 2008	0.552303	0.028472	19.398	< 2e-16 ***
(Y) 2009	0.250013	0.023146	10.802	< 2e-16 ***
(Y) 2010	0.401379	0.025724	15.603	< 2e-16 ***
(Y) 2011	0.471368	0.024095	19.563	< 2e-16 ***
(Y) 2012	0.587533	0.051834	11.335	< 2e-16 ***
(A) C	-0.08724	0.005018	-17.385	< 2e-16 ***
(A) S	-0.39617	0.00705	-56.19	< 2e-16 ***
(T) 2	-0.01517	0.005616	-2.701	0.00691 **
(T) 3	-0.07196	0.0063	-11.423	< 2e-16 ***
(T) 4	-0.17505	0.006045	-28.958	< 2e-16 ***

Table 5 – Estimations of parameters (approach B). F – flag; Y – year; A –area; T – quarter.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-5.07446	0.067325	-75.373	< 2e-16 ***
(F) BRA-ESP	0.495597	0.006158	80.483	< 2e-16 ***
(F) BRA-HND	-0.41447	0.013503	-30.696	< 2e-16 ***
(F) BRA-JPN	-1.0409	0.021454	-48.517	< 2e-16 ***
(F) BRA-PAN	-2.39612	0.018921	-126.638	< 2e-16 ***
(F) BRA-TAI	-1.10499	0.010497	-105.269	< 2e-16 ***
(F) BRA-VCT	-1.7467	0.010065	-173.546	< 2e-16 ***
(Y) 1991	0.739832	0.074996	9.865	< 2e-16 ***
(Y) 1992	-0.7335	0.079626	-9.212	< 2e-16 ***
(Y) 1993	-0.95297	0.098095	-9.715	< 2e-16 ***
(Y) 1994	-1.59937	0.113136	-14.137	< 2e-16 ***
(Y) 1995	-0.15777	0.072983	-2.162	0.030637 *
(Y) 1996	-0.54676	0.16009	-3.415	0.000637 ***
(Y) 1997	-0.36711	0.077253	-4.752	2.01e-06 ***
(Y) 1998	1.238418	0.069746	17.756	< 2e-16 ***

(continuation)

	Estimate	Std. Error	z value	Pr(>|z|)
(Y) 1999	0.004599	0.069262	0.066	0.947063
(Y) 2000	0.066304	0.068257	0.971	0.331356
(Y) 2001	-0.2138	0.068512	-3.121	0.001805 **
(Y) 2002	-0.63882	0.070853	-9.016	< 2e-16 ***
(Y) 2003	0.474803	0.078353	6.06	1.36e-09 ***
(Y) 2004	-0.28181	0.074582	-3.779	0.000158 ***
(Y) 2005	0.249724	0.069142	3.612	0.000304 ***
(Y) 2006	0.354296	0.069592	5.091	3.56e-07 ***
(Y) 2007	0.308924	0.072751	4.246	2.17e-05 ***
(Y) 2008	0.775437	0.071138	10.9	< 2e-16 ***
(Y) 2009	0.231912	0.068931	3.364	0.000767 ***
(Y) 2010	0.471997	0.070043	6.739	1.60e-11 ***
(Y) 2011	0.433417	0.071371	6.073	1.26e-09 ***
(Y) 2012	0.450219	0.084977	5.298	1.17e-07 ***
(A) C	0.169907	0.02065	8.228	< 2e-16 ***
(A) S	0.147509	0.040588	3.634	0.000279 ***
(T) 2	-0.61476	0.095608	-6.43	1.28e-10 ***
(T) 3	0.126615	0.077236	1.639	0.101143
(T) 4	0.252503	0.074502	3.389	0.000701 ***
(Y) 1991: (T) 2	0.006294	0.105712	0.06	0.952523
(Y) 1992: (T) 2	1.191529	0.11196	10.642	< 2e-16 ***
(Y) 1993: (T) 2	1.836874	0.137105	13.398	< 2e-16 ***
(Y) 1994: (T) 2	2.903251	0.137937	21.048	< 2e-16 ***
(Y) 1995: (T) 2	0.634585	0.105977	5.988	2.12e-09 ***
(Y) 1996: (T) 2	1.414172	0.17905	7.898	2.83e-15 ***
(Y) 1997: (T) 2	0.260871	0.103626	2.517	0.011822 *
(Y) 1998: (T) 2	0.642135	0.097701	6.572	4.95e-11 ***
(Y) 1999: (T) 2	0.124562	0.098761	1.261	0.207221
(Y) 2000: (T) 2	0.102722	0.097085	1.058	0.290025
(Y) 2001: (T) 2	0.708585	0.097005	7.305	2.78e-13 ***
(Y) 2002: (T) 2	0.661004	0.102271	6.463	1.02e-10 ***
(Y) 2003: (T) 2	0.745126	0.106863	6.973	3.11e-12 ***
(Y) 2004: (T) 2	1.047115	0.102909	10.175	< 2e-16 ***
(Y) 2005: (T) 2	0.708868	0.097113	7.299	2.89e-13 ***
(Y) 2006: (T) 2	0.732643	0.097512	7.513	5.76e-14 ***
(Y) 2007: (T) 2	1.190184	0.10061	11.83	< 2e-16 ***
(Y) 2008: (T) 2	0.350577	0.101759	3.445	0.000571 ***
(Y) 2009: (T) 2	0.570721	0.097157	5.874	4.25e-09 ***
(Y) 2010: (T) 2	0.937175	0.101254	9.256	< 2e-16 ***
(Y) 2011: (T) 2	0.64712	0.098574	6.565	5.21e-11 ***
(Y) 1991: (T) 3	-0.35786	0.089278	-4.008	6.11e-05 ***
(Y) 1992: (T) 3	1.019362	0.09705	10.503	< 2e-16 ***
(Y) 1993: (T) 3	1.261743	0.213396	5.913	3.37e-09 ***
(Y) 1994: (T) 3	2.218941	0.130417	17.014	< 2e-16 ***
(Y) 1995: (T) 3	0.100161	0.088765	1.128	0.259158
(Y) 1996: (T) 3	-0.2016	0.183653	-1.098	0.272335
(Y) 1997: (T) 3	0.593599	0.087251	6.803	1.02e-11 ***
(Y) 1998: (T) 3	-0.35888	0.08069	-4.448	8.68e-06 ***
(Y) 1999: (T) 3	-0.21516	0.081594	-2.637	0.008365 **
(Y) 2000: (T) 3	-0.59231	0.079006	-7.497	6.53e-14 ***
(Y) 2001: (T) 3	-0.56826	0.081442	-6.978	3.00e-12 ***
(Y) 2002: (T) 3	0.446394	0.083826	5.325	1.01e-07 ***
(Y) 2003: (T) 3	-0.94327	0.099504	-9.48	< 2e-16 ***
(Y) 2004: (T) 3	0.403911	0.084897	4.758	1.96e-06 ***
(Y) 2005: (T) 3	-0.22516	0.079413	-2.835	0.004578 **
(Y) 2006: (T) 3	-0.05411	0.080623	-0.671	0.502089
(Y) 2007: (T) 3	0.054233	0.087124	0.622	0.533628
(Y) 2009: (T) 3	-0.20368	0.080103	-2.543	0.010999 *

(continuation)

	Estimate	Std. Error	z value	Pr(>|z|)
(Y) 2010: (T) 3	-0.63599	0.089105	-7.138	9.50e-13 ***
(Y) 2011: (T) 3	-0.04494	0.083777	-0.536	0.591702
(Y) 1991: (T) 4	-1.48312	0.09294	-15.958	< 2e-16 ***
(Y) 1992: (T) 4	0.009259	0.088688	0.104	0.916855
(Y) 1993: (T) 4	1.355852	0.127614	10.625	< 2e-16 ***
(Y) 1994: (T) 4	1.697438	0.120745	14.058	< 2e-16 ***
(Y) 1995: (T) 4	-0.44521	0.083964	-5.302	1.14e-07 ***
(Y) 1996: (T) 4	0.000357	0.169779	0.002	0.998321
(Y) 1997: (T) 4	-0.38193	0.087575	-4.361	1.29e-05 ***
(Y) 1998: (T) 4	-1.21245	0.077829	-15.578	< 2e-16 ***
(Y) 1999: (T) 4	-0.27606	0.076873	-3.591	0.000329 ***
(Y) 2000: (T) 4	-0.19313	0.07579	-2.548	0.010827 *
(Y) 2001: (T) 4	-0.93902	0.077027	-12.191	< 2e-16 ***
(Y) 2002: (T) 4	-0.24593	0.084955	-2.895	0.003793 **
(Y) 2003: (T) 4	-1.235	0.093732	-13.176	< 2e-16 ***
(Y) 2004: (T) 4	0.25338	0.082434	3.074	0.002114 **
(Y) 2005: (T) 4	-0.47888	0.07684	-6.232	4.60e-10 ***
(Y) 2006: (T) 4	-0.22813	0.078056	-2.923	0.003471 **
(Y) 2007: (T) 4	0.166102	0.088312	1.881	0.059991 .
(Y) 2009: (T) 4	-0.26786	0.078389	-3.417	0.000633 ***
(Y) 2010: (T) 4	-0.85737	0.099533	-8.614	< 2e-16 ***
(Y) 2011: (T) 4	-0.45732	0.080658	-5.67	1.43e-08 ***
(Y) 2012: (T) 4	-0.60168	0.281427	-2.138	0.032519 *
(Y) 1991: (A) C	-1.39639	0.085839	-16.268	< 2e-16 ***
(Y) 1992: (A) C	-0.51586	0.046116	-11.186	< 2e-16 ***
(Y) 1993: (A) C	-0.53303	0.710934	-0.75	0.453398
(Y) 1994: (A) C	0.423759	0.110096	3.849	0.000119 ***
(Y) 1995: (A) C	-0.54847	0.080967	-6.774	1.25e-11 ***
(Y) 1996: (A) C	-1.29186	0.294635	-4.385	1.16e-05 ***
(Y) 1997: (A) C	-0.31845	0.038172	-8.343	< 2e-16 ***
(Y) 1998: (A) C	-0.65821	0.025866	-25.447	< 2e-16 ***
(Y) 1999: (A) C	-0.68169	0.028176	-24.194	< 2e-16 ***
(Y) 2000: (A) C	-0.32939	0.024254	-13.581	< 2e-16 ***
(Y) 2001: (A) C	-0.37514	0.025209	-14.881	< 2e-16 ***
(Y) 2002: (A) C	-0.80864	0.048273	-16.752	< 2e-16 ***
(Y) 2003: (A) C	-0.14876	0.05832	-2.551	0.010749 *
(Y) 2004: (A) C	0.037776	0.034401	1.098	0.272165
(Y) 2005: (A) C	0.129323	0.025398	5.092	3.55e-07 ***
(Y) 2006: (A) C	-0.06438	0.027035	-2.381	0.017252 *
(Y) 2007: (A) C	-0.61795	0.034766	-17.774	< 2e-16 ***
(Y) 2009: (A) C	0.060129	0.025897	2.322	0.020240 *
(Y) 2010: (A) C	0.323123	0.049652	6.508	7.63e-11 ***
(Y) 1991: (A) S	0.061088	0.056534	1.081	0.279893
(Y) 1992: (A) S	-0.86657	0.063473	-13.653	< 2e-16 ***
(Y) 1993: (A) S	-2.24325	0.300093	-7.475	7.71e-14 ***
(Y) 1994: (A) S	-1.97839	0.069934	-28.29	< 2e-16 ***
(Y) 1995: (A) S	-0.18796	0.055675	-3.376	0.000735 ***
(Y) 1997: (A) S	-0.60899	0.061674	-9.874	< 2e-16 ***
(Y) 1998: (A) S	-1.66477	0.050135	-33.206	< 2e-16 ***
(Y) 1999: (A) S	-0.72812	0.049503	-14.709	< 2e-16 ***
(Y) 2000: (A) S	-0.32238	0.044426	-7.257	3.97e-13 ***
(Y) 2001: (A) S	-1.81747	0.049512	-36.708	< 2e-16 ***
(Y) 2002: (A) S	0.140314	0.050246	2.793	0.005229 **
(Y) 2003: (A) S	-0.48253	0.089635	-5.383	7.32e-08 ***
(Y) 2004: (A) S	-0.09558	0.051027	-1.873	0.061043 .
(Y) 2005: (A) S	-0.18621	0.045338	-4.107	4.01e-05 ***
(Y) 2006: (A) S	-0.62916	0.047026	-13.379	< 2e-16 ***
(Y) 2007: (A) S	-1.67698	0.116964	-14.338	< 2e-16 ***

Table 6 – Estimations of parameters (approach C). F – flag; Y – year; A – area; T – quarter.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercepto)	-5.21345	0.371228	-14.04	< 2e-16 ***
(F) BRA-ESP	0.495026	0.006154	80.44	< 2e-16 ***
(F) BRA-HND	-0.414317	0.013491	-30.71	< 2e-16 ***
(F) BRA-JPN	-1.045821	0.021419	-48.83	< 2e-16 ***
(F) BRA-PAN	-2.396636	0.018916	-126.7	< 2e-16 ***
(F) BRA-TAI	-1.107157	0.010492	-105.52	< 2e-16 ***
(F) BRA-VCT	-1.747708	0.010061	-173.71	< 2e-16 ***
(Y) 1991	0.224517	0.484605	0.46	0.643
(Y) 1992	-0.260225	0.483128	-0.54	0.59
(Y) 1993	-0.422948	0.513381	-0.82	0.41
(Y) 1994	-0.049023	0.479228	-0.1	0.919
(Y) 1995	0.0512	0.48339	0.11	0.916
(Y) 1996	-0.629945	0.524024	-1.2	0.229
(Y) 1997	-0.180397	0.481345	-0.37	0.708
(Y) 1998	0.610795	0.480592	1.27	0.204
(Y) 1999	-0.178489	0.476038	-0.37	0.708
(Y) 2000	0.056957	0.479051	0.12	0.905
(Y) 2001	-0.764767	0.479427	-1.6	0.111
(Y) 2002	-0.266229	0.465471	-0.57	0.567
(Y) 2003	0.285172	0.484657	0.59	0.556
(Y) 2004	0.502767	0.482079	1.04	0.297
(Y) 2005	0.6099	0.467124	1.31	0.192
(Y) 2006	0.613721	0.487479	1.26	0.208
(Y) 2007	0.283913	0.481852	0.59	0.556
(Y) 2008	0.700381	0.621968	1.13	0.26
(Y) 2009	0.653628	0.481106	1.36	0.174
(Y) 2010	0.684309	0.50858	1.35	0.178
(Y) 2011	0.660301	0.51162	1.29	0.197
(Y) 2012	0.827209	0.639339	1.29	0.196
(A) C	-0.210572	0.130841	-1.61	0.108
(A) S	-0.604828	0.137089	-4.41	1.02e-05 ***
(T) 2	0.164778	0.115397	1.43	0.153
(T) 3	0.182008	0.117553	1.55	0.122
(T) 4	-0.027673	0.11568	-0.24	0.811

Table 7 – Estimation of variance components for random effects (year x quarter interaction and (Y) x area interaction). Y – year; A – area; T – quarter.

Group	Variance	Std. Dev.
Y:T	0.1506	0.3881
Y:A	0.1709	0.4134

Number of observations: 35055; Groups: Y: T = 88; Y: A = 60.

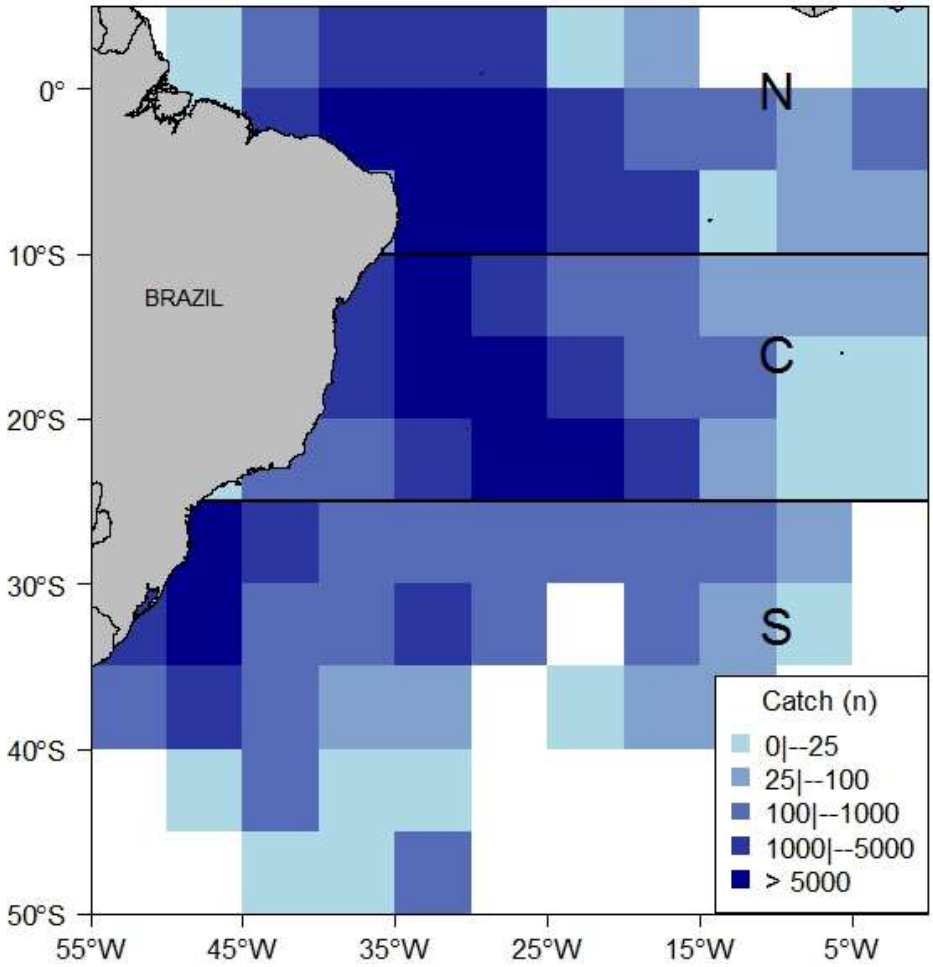


Figure 1 – Distribution of catches of *Xiphias gladius* from 1990 to 2012. Fishing areas: N - North; C - Central and S - South.

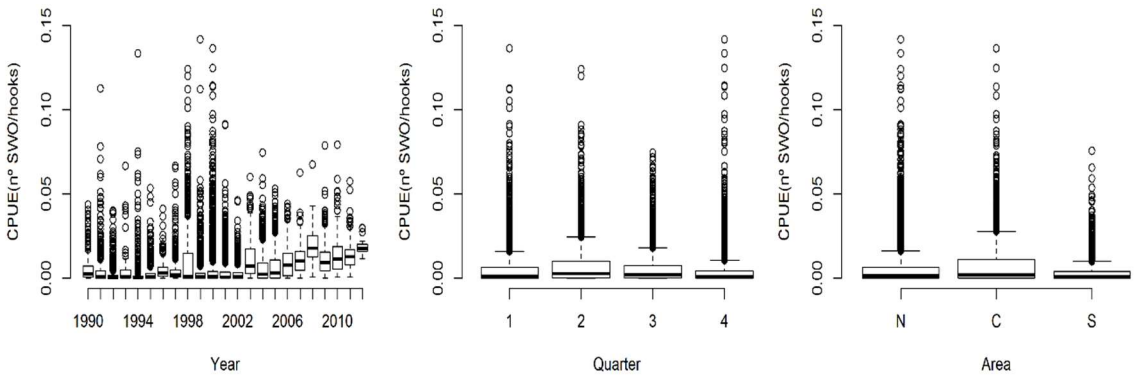


Figure 2 – Catch per unit effort (CPUE) by year, quarter and area (N-north, C-central, S-south).

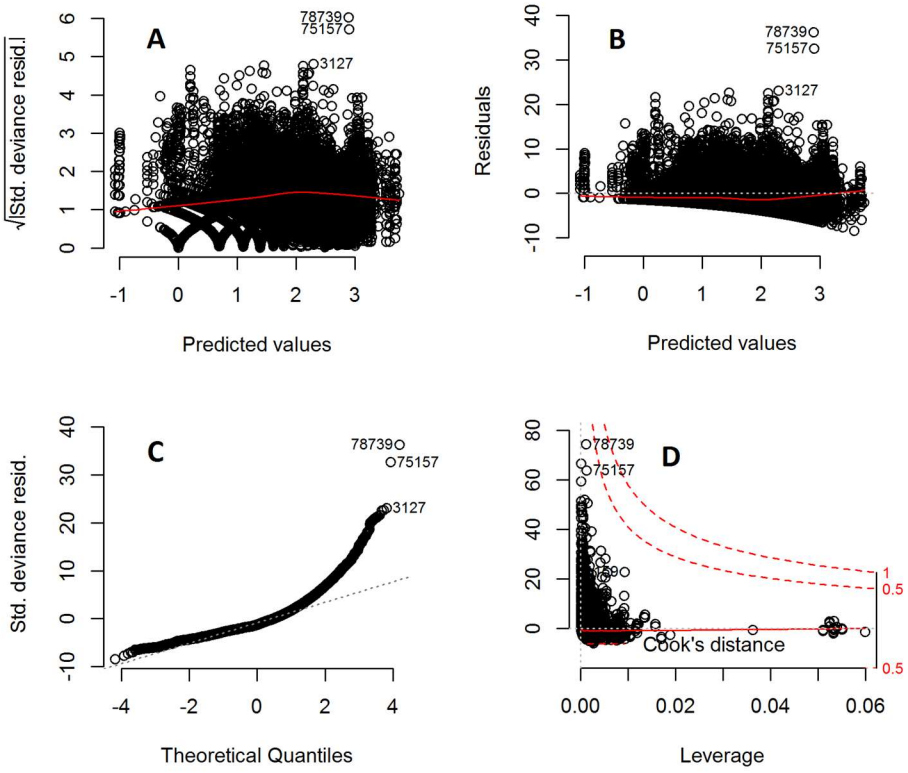


Figure 3 – Diagnostic of residuals (approach A).

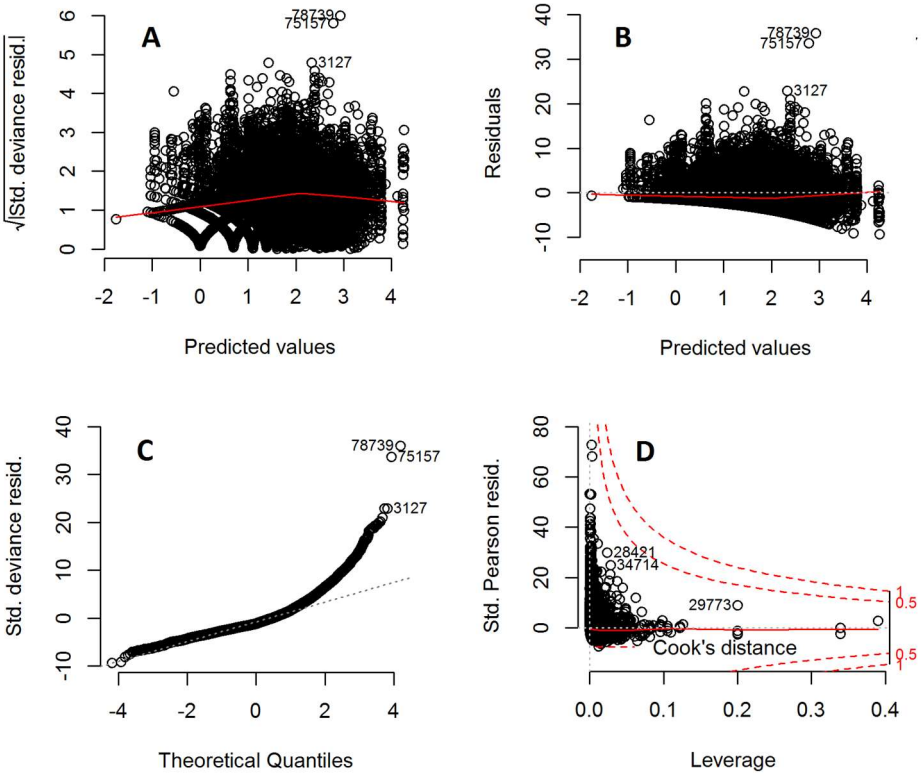


Figure 4 - Diagnostic of residuals (approach B).

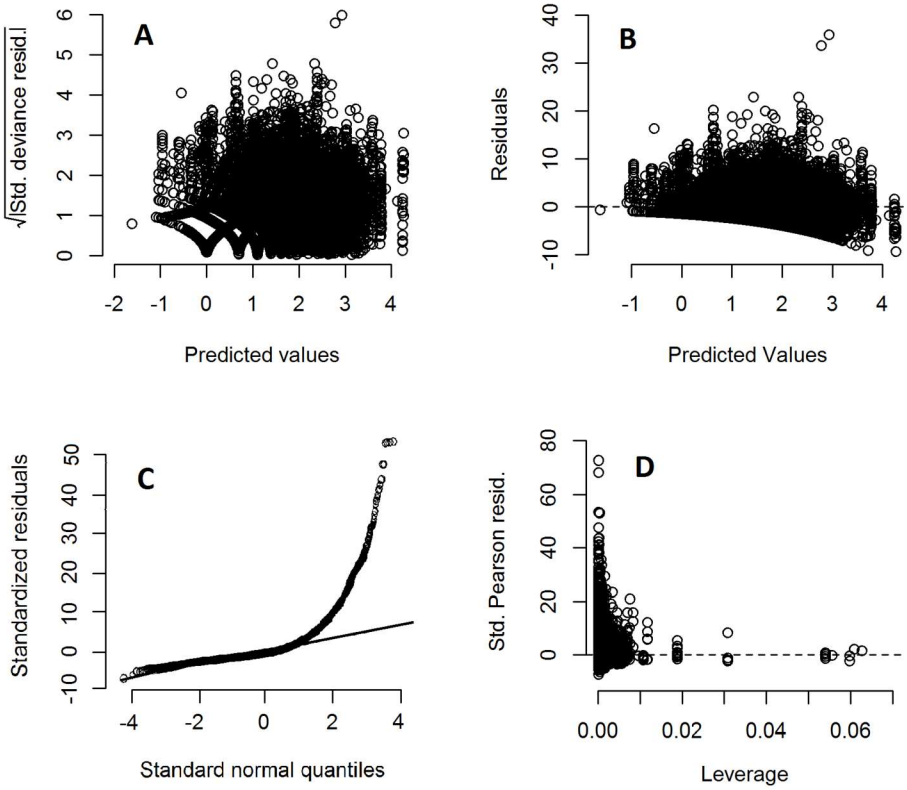


Figure 5 - Diagnostic of residuals (approach C).

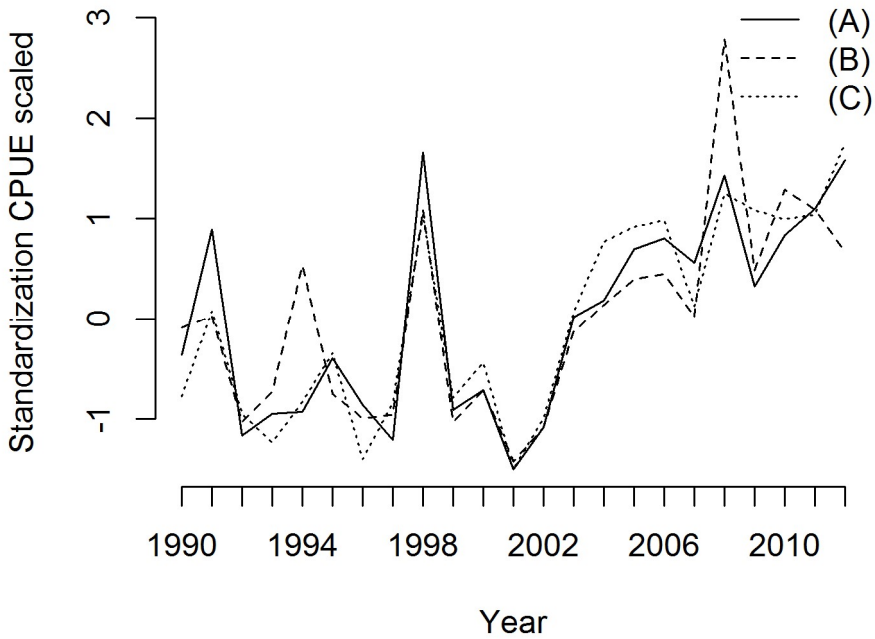


Figure 6 - Standardized CPUE as calculated following the approaches A, B and C. Estimations were scaled by calculating the difference between the original value and the mean of the series, divided by the standard deviation of the series.