

Nathielly Lima do Rêgo

**Análise da estrutura do comprimento da sardinha-laje
(*Opisthonema oglinum*) capturada no Canal de Santa
Cruz-PE a partir dos modelos aditivos generalizados para
locação, escala e forma**

Recife

30 de agosto de 2021



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

**Análise da estrutura do comprimento da sardinha-laje (*Opisthonema oglinum*)
capturada no Canal de Santa Cruz-PE a partir dos modelos aditivos generalizados
para locação, escala e forma**

Dissertação julgada adequada para obtenção do título de Mestre em Biometria e Estatística Aplicada, defendida e aprovada por unanimidade em 30/08/2021 pela comissão examinadora

Área de concentração: Biometria e Estatística Aplicada

Orientador: Dr. Paulo José Duarte Neto
Coorientador: Dr. Humber Agrelli de Andrade

Recife

30 de agosto de 2021

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

- R343a Rêgo, Nathielly Lima do
Análise da estrutura do comprimento da sardinha-laje (*Opisthonema oglinum*) capturada no Canal de Santa Cruz-PE a partir dos modelos aditivos generalizados para locação, escala e forma / Nathielly Lima do Rêgo. - 2021.
82 f. : il.
- Orientador: Paulo Jose Duarte Neto.
Coorientador: Humber Agrelli de Andrade.
Inclui referências e apêndice(s).
- Dissertação (Mestrado) - Universidade Federal Rural de Pernambuco, Programa de Pós-Graduação em Biometria e Estatística Aplicada, Recife, 2021.
1. *Opisthonema oglinum*. 2. Canal de Santa Cruz. 3. GAMLSS. I. Neto, Paulo Jose Duarte, orient. II. Andrade, Humber Agrelli de, coorient. III. Título

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

**Análise da estrutura do comprimento da sardinha-laje (*Opisthonema oglinum*)
capturada no Canal de Santa Cruz-PE a partir dos modelos aditivos generalizados
para locação, escala e forma**

Nathielly Lima do Rêgo

Dissertação julgada adequada para obtenção
do título de Mestre em Biometria e Esta-
tística Aplicada, defendida e aprovada por
unanimidade em 30/08/2021 pela comissão
examinadora

Orientador:

Dr. Paulo José Duarte Neto
Orientador

Banca examinadora:

Dr. Francisco Marcante Santana
Universidade Federal Rural de
Pernambuco

Dr. Kleber Napoleão Nunes de
Oliveira Barros
Universidade Federal Rural de
Pernambuco

*Dedico este trabalho primeiramente a Deus,
por ser essencial em minha vida,
e a minha filha.*

Agradecimentos

À Deus agradeço pelo dom da vida, pelo seu infinito amor para comigo e por todas as bênçãos ofertadas, pois sem Ele nada sou.

A minha mãe e avó materna por serem meus maiores exemplos de mulher, mãe, obrigada pelas orações em meu favor, incentivos incondicionais, orientações e preocupações para que eu pudesse sempre percorrer todo caminho tranquila, pelo amor a mim e a minha filha dedicado que foi imprescindível neste percurso. As palavras não expressam a minha gratidão por terem feito do meu sonho o de vocês.

Ao meu pai por todo o esforço investido na minha educação e pelo apoio que sempre me deu durante toda a vida.

As minhas irmãs por todo amor e carinho que me dedica e por sempre se preocupar e acreditar nos meus sonhos.

As minhas tias Verônica, Ozita e Odaiza por toda ajuda foi muito valiosa neste percurso, obrigada por sempre serem presentes em minha vida.

Ao meu orientador Prof. Paulo José Duarte Neto, pelo esforço, os ensinamentos, a confiança e disponibilidade durante a orientação.

Ao meu coorientador Prof. Humber Agrelli de Andrade, pelo esforço, os ensinamentos a me transmitido de forma tão generosa, por toda disponibilidade, paciência, incentivos, pelo imenso apoio e colaboração para que eu chegasse até aqui.

As minhas amigas Herica, Kerolly, Evellyn, Lêda, pelos conselhos, disponibilidades, pela amizade foi um prazer imenso conviver com vocês.

Aos colegas acadêmicos Rayanne, Ikaro, Albaro, Rodrigo, Edy, Fábio, Samuel, Jonas, Isabelly, Filipe, Neidinha.

À CAPES, pela concessão de bolsa de estudos, à UFRPE, ao DEINFO, a todo o corpo docente pela oportunidade de se desenvolver academicamente, ao secretário do programa Marco.

A todos os nominados e não nominados que de alguma forma contribuíram para realização deste trabalho, meu muito obrigado.

Resumo

A manjuba (*Opisthonema oglinum*) é a sardinha de maior ocorrência nos desembarques do Canal de Santa Cruz (CSC) localizado no litoral norte do estado de Pernambuco. A espécie é classificada como de origem marinha visitante, visto que a mesma passa parte do seu ciclo de vida no mar e penetra nos estuários em períodos estratégicos. A problemática inicial desta pesquisa se concentra em especular algumas hipóteses referentes à captura excessiva de juvenis da *O. oglinum*, que sejam relevantes para profissionais da área pesqueira para tomadas de decisão. A alta exploração da espécie e sua relevância para a economia nos municípios que cercam o CSC provocaram-nos reflexões acerca da necessidade de investigações que promovam a pesca sustentável da *O. oglinum* e que, conseqüentemente, amplie a compreensão á respeito da espécie no CSC, uma vez que compreender a estrutura de comprimento nas capturas é válida não só para expandir o conhecimento, como também para levantar hipóteses a respeito da pressão pesqueira sobre o recurso. A investigação foi realizada através de estimativas do comprimento padrão dos exemplares, com a utilização de modelos lineares generalizados (GLM) e modelos aditivos generalizados para locação, escala e forma (GAMLSS). Após a realização da modelagem, observamos que os modelos estimados via GAMLSS viabilizaram um ajuste superior para descrever o comprimento padrão da *O. oglinum* em comparação aos obtidos via GLM, segundo o critério GAIC e as análises dos resíduos. A distribuição Box Cox t Original foi escolhida para ajustar a variável resposta com a modelagem dos parâmetros de posição e escala. Diante disto, notamos a existência de sazonalidade no comprimento padrão da *O. oglinum* capturada no Canal de Santa Cruz, com a presença de indivíduos maiores no período chuvoso, e maior concentração na região centro-sul do canal. Em contra partida as sardinhas de menor tamanho concentraram-se próximo à desembocadura do rio Botafogo, área de canais secundários e com maior cobertura de manguezal. Encontramos indícios de que poucos adultos ingressam no canal e a grande maioria dos exemplares capturados no CSC foram jovens. Desta forma, seria pertinente a exploração de exemplares que já completaram ao menos a maturidade sexual, uma vez que a captura de indivíduos jovens pode provocar desequilíbrio na sustentabilidade da espécie.

Palavras-chaves: *Opisthonema oglinum*. Canal de Santa Cruz. GAMLSS.

Abstract

Manjuba (*Opisthonema oglinum*) is the most common sardine in Santa Cruz Channel's (CSC) landings, located on the northern coast of Pernambuco state. The species is classified as visiting marine origin, as it spends part of its life cycle at the sea and enters estuaries at strategic times. The initial research's issue focuses on speculating some hypotheses related to the juveniles' over-catching from *O. oglinum* species, that are relevant to fisheries professionals on decision-making. The high exploitation of the species and its importance to the economy in the municipalities surrounding the CSC led us to reflect on the need for investigations to promote the sustainable fishing of *O. oglinum* and, consequently, expand the understanding of the species in the CSC, once the comprehension of the length structure in catches is valid not only to expand knowledge but also to raise hypotheses about the fishing pressure on the resource. The investigation was performed by estimating the standard length of the specimens, using generalized linear models (GLM) and generalized additive models for location, scale, and shape (GAMLSS). After performing the modeling, we observed that the models estimated via GAMLSS enabled a better fit to describe the standard length of *O. oglinum* compared to those obtained via GLM, according to the GAIC criterion and residue analysis. The Box Cox t Original distribution was chosen to adjust the response variable with the modeling of position and scale parameters. Therefore, we realized the existence of seasonality in the standard length of *O. oglinum* captured in the Santa Cruz Channel, with the presence of larger individuals in the rainy season, and greater concentration in the center-south region of the channel. On the other hand, the smaller sardines were concentrated near the outflow of the Botafogo River, an area of secondary channels with greater mangrove coverage. We have found evidence that few adults enter the channel and the vast majority of specimens captured in the CSC were young. Thus, it would be pertinent to explore specimens that have completed at least the sexual maturity, since the capture of young individuals can cause an imbalance in the sustainability of the species.

Key-words: *Opisthonema oglinum*. Santa Cruz Channel. GAMLSS.

Sumário

1	Introdução	1
2	Revisão de Literatura	3
3	Objetivos	10
3.1	Objetivo Geral	10
3.2	Objetivos Específicos	10
4	Metodologia	11
4.1	Material	11
4.2	Introdução a modelagem	13
4.3	Modelos Lineares Generalizados	15
4.3.1	Estimação	16
4.3.2	Medidas de qualidade de Ajuste (Função desvio)	17
4.3.3	Critério de seleção dos modelos	18
4.3.4	Análise de resíduos	19
4.4	Modelos Aditivos Generalizados	21
4.5	Modelo Aditivo Generalizado para Locação, Escala e Forma	22
4.5.1	Estimação	23
4.5.2	Critério de seleção dos modelos	26
4.5.3	Análise de resíduos	26
4.6	Teste de homocedasticidade - <i>Breusch-Pagan</i>	29
4.7	Teste de normalidade - <i>Shapiro-Wilk</i>	29
5	Resultados	31
5.1	Análise exploratória	31
5.2	Modelo Linear Generalizado	33
5.3	Modelos Aditivos Generalizados para Locação, Escala e Forma	36
6	Discussão	44
7	Conclusão	47
	Referências Bibliográficas	48

Apêndices	56
APÊNDICE A Código em R utilizado na aplicação	57
APÊNDICE B Distribuições citadas no trabalho	67
B.1 Box-Cox t , $BCT(\mu, \sigma, \nu, \tau)$	67
B.2 Box-Cox t original, $BCTo(\mu, \sigma, \nu, \tau)$	67
B.3 Beta generalizada tipo II, $GB2(\mu, \sigma, \nu, \tau)$	68
B.4 Johnson SU, $JSU(\mu, \sigma, \nu, \tau)$	68
B.5 Skew t tipo 1, $ST1(\mu, \sigma, \nu, \tau)$	68
B.6 Skew t tipo 2, $ST2(\mu, \sigma, \nu, \tau)$	69
B.7 Skew t tipo 3, $ST3(\mu, \sigma, \nu, \tau)$	69
B.8 Skew t tipo 4, $ST4(\mu, \sigma, \nu, \tau)$	69
B.9 Skew t tipo 5, $ST5(\mu, \sigma, \nu, \tau)$	70
B.10 Skew t de <i>Student</i> , $SST(\mu, \sigma, \nu, \tau)$	70
B.11 Box-Cox exponential potência, $BCPE(\mu, \sigma, \nu, \tau)$	70
B.12 Box-Cox exponential potência original, $BCPEo(\mu, \sigma, \nu, \tau)$	71
B.13 Exponencial Gaussiana, $exGAUS(\mu, \sigma, \nu)$	71

1 Introdução

A sardinha-laje (*Opisthonema oglinum*), também conhecida por manjuba, é uma espécie de pequeno clupeídeo pelágico que habita em camadas superiores das águas costeiras e estuárias (WHITEHEAD, 1985). A espécie está distribuída em toda a costa atlântica americana, desde a Nova Inglaterra até a Argentina (FIGUEREIDO & MENEZES, 1978). No Brasil, a *O. oglinum* é considerada a principal alternativa para manter o fornecimento de matéria prima para o setor pesqueiro industrial, frente ao declínio nas capturas da sardinha-verdadeira (*Sardinella brasiliensis*), cuja exploração entrou em colapso devido à sobrepesca no Sudeste brasileiro (NÓBREGA et al., 2009).

Em alguns municípios nordestinos, a captura da *O. oglinum* é realizada através da pesca artesanal, caracterizada como uma das mais tradicionais e de alto destaque socioeconômico para a região (LESSA et al., 2004). No litoral norte do estado de Pernambuco, a pesca artesanal com uso da rede de emalhar é o principal meio de sustento para diversas famílias de forma direta ou indireta, realizada em grande maioria por pescadores de comunidades litorâneas. Moura (2009) aponta que em Itapissuma, cerca de 70% da população é dependente ou estaria diretamente envolvida com as pescarias de pequenos peixes pelágicos. A fauna potencialmente econômica do canal pode chegar a subsidiar cerca de cinco mil pescadores e famílias envolvidas no processo (LIMA & QUINAMO, 2000).

De acordo com IBAMA/CEPENE (2008) a *Opisthonema oglinum* está entre os principais recursos pesqueiros capturados em Pernambuco, com 12,2% do volume desembarcado em todo estado, cuja contribuição do município de Itapissuma foi de aproximadamente 33,3% em 2007. A alta demanda na captura da sardinha-laje nos municípios litorâneos do estado, pode ocasionar a degradação do recurso. Lima (2015) apontou relatos de pescadores da região mencionando que os comprimentos mínimo e máximo da *O. oglinum* nas capturas realizadas no CSC vêm diminuindo ao longo dos anos. Diminuições nos comprimentos médios estão usualmente associados à mortalidade por pesca (BEVERTON & HOLT, 1957), e, eventualmente, podem ser usados como indicadores de sobrepesca.

A captura de *O. oglinum* abaixo do comprimento médio de seleção deve ser investigada para a gestão eficaz da pesca, visto que recursos pesqueiros atrativos economicamente e de importância social quando mau geridos tendem a colapsar (LIMA, 2015). Uma alternativa para gerar informações para profissionais na área de gestão de recurso pes-

queiro, seriam estimativas do comprimento médio da *O. oglinum* no CSC, pois permitiria verificar a estrutura de comprimento da espécie e observar se está havendo diminuição no comprimento mínimo do recurso. Logo, poderíamos levantar hipóteses sobre processos ecológicos que sejam relevantes para profissionais da área de pesca, através de ajustes de modelos para estimar o comprimento da *O. oglinum*.

Entre as técnicas de modelagem, a análise de regressão tem sido muito utilizada para modelar fenômenos biológicos. Técnica esta, cujo pressuposto de maior rigor é de que os resíduos sigam a distribuição normal. Um pressuposto extremamente necessário para o uso correto da análise de regressão. Porém, em algumas situações do cotidiano da pesquisa, existem fenômenos de maior complexidade no momento do ajuste, onde os resíduos não seguem a distribuição normal. Assim, os modelos clássicos de regressão são, de certa forma, inadequados.

De posse disto, Nelder e Wedderburn (1972) sugeriram os Modelos Lineares Generalizados (*Generalized Linear Models - GLM*), em que a ideia principal é expandir as opções de distribuição de probabilidade dos erros, associados ao modelo de regressão, e, conseqüentemente da variável resposta, de forma a abranger aquelas que pertençam a família exponencial, além de flexibilizar a relação funcional entre a variável resposta e as variáveis explicativas do modelo (PAULA, 2013). No entanto, ainda que os GLMs possibilitem o uso de distribuições com assimetria, muitas vezes, os dados apresentam uma variabilidade distinta para diferentes unidades amostrais, de forma não proporcional a uma função da média, ocasionando a necessidade de modelagem do parâmetro de escala do modelo.

Então, no intuito de viabilizar modelos com pressupostos mais flexíveis na especificação da distribuição da variável resposta, Rigby e Stasinopoulos (2005) introduziram os Modelos Aditivos Generalizados para Localização, Escala e Forma (*Generalized Additive Models for Location, Scale and Shape - GAMLSS*). Os GAMLSS proporcionam o ajuste de modelos que admitem inúmeras distribuições para a variável resposta sem pertencer obrigatoriamente à família exponencial. Sendo assim, modelos capazes de integrar as funções paramétricas e/ou não paramétrica, para a modelagem de qualquer parâmetro da distribuição assumida.

Em vista disso, propomos estruturar modelos no intuito de obter estimativas sobre o comprimento padrão da *Opisthonema oglinum*, em função das variáveis mês, ano, setor do CSC, maré no horário do lançamento da rede e, precipitação média acumulada em 30 dias; com o propósito de especular hipóteses que auxiliem no processo de tomada de decisão e implementação de regras para manter o recurso em níveis razoáveis para a sobrevivência da atividade pesqueira no Canal de Santa Cruz.

2 Revisão de Literatura

Litoral brasileiro

Constituído por uma área de 8.515.759,090 km^2 (IBGE, 2017), o Brasil é o país de maior extensão territorial da América do Sul, cujo litoral é banhado pelo Oceano Atlântico, com extensão de aproximadamente 8.500 km (MMA, 2020). A costa brasileira é caracterizada por correntes quentes que contribuem para a pluralidade de *habitats* que, por consequência, favorecem ampla diversidade de ecossistemas costeiros. As praias arenosas, por exemplo, são formadas por um sistema de baixa diversidade, devido à escassez de alimentos. Em contra partida, as lagoas costeiras e os estuários¹, constituem sistemas férteis que servem de abrigo e criadouro para uma vasta variedade de espécies (MMA, 2020).

Os estuários possuem uma ictiofauna constituída por espécies de origem marinha visitante e de água doce (ROZAS & ZIMMERMAN 2000; VIDY 2000; PAIVA, 2009). Algumas espécies migram para regiões mais próximas à nascente ou para a foz, por existir maior variação da salinidade (BLABER & BLABER 1980, BLABER et al. 1989); algumas deslocam-se de águas rasas para águas de maiores profundidades por mudanças na temperatura, ou para o mar, cujas condições hidrológicas são mais estáveis (BLABER et al. 1990, ALBERT & BERGSTAD 1993, LEKVE et al. 1999; PAIVA, 2009).

As oscilações sazonais e os parâmetros físico-químicos nas águas estuarinas podem determinar a permanência ou não de algumas espécies em certa época do seu ciclo de vida (CHAVES & OTTO 1999; PAIVA, 2009). A distribuição espacial das comunidades de peixes em cada estuário pode variar de acordo com as características do sedimento, além da presença de vegetação que influencia a disponibilidade de presas (BLABER & BLABER 1980, GARCIA & VIEIRA 1997, MARSHALL & ELLIOTT 1998, VIEIRA et al. 1998).

Desta forma, a abundância de peixes nos estuários deve-se principalmente à possibilidade de obtenção de alimentos, a partir da produção primária (ROBERTSON & BLABER, 1992; SIMONI, 2019); complexidade estrutural da vegetação de mangue, que propicia refúgio, especialmente para os peixes jovens; elevada turbidez da água; e reduzido número de peixes carnívoros de grande porte (ROBERTSON & BLABER 1992; MULLIN

¹ Os estuários são corpos de água costeiros, semi-fechados, caracterizado pelo encontro do rio com o mar, um ambiente de transição entre o oceano e o continente (MIRANDA et al. 2002; PAIVA, 2009)

1995; PAIVA, 2009).

Litoral do estado de Pernambuco

O litoral de Pernambuco possui aproximadamente 187 km de extensão e dispõe de 14 zonas estuarinas de grande relevância biológica por fertilizar as águas costeiras e contribuir para a produtividade da plataforma continental. O Canal de Santa Cruz (Figura 1), maior estuário do estado, possui área de aproximadamente 824 km^2 , seu canal estende-se por 22 km (MOURA, 2009) e, separa a Ilha de Itamaracá do município de Itapissuma no continente (CPRH, 2010).

O canal recebe a descarga líquida dos rios Igarassu, Botafogo, Arataca, Carrapicho e Catuama, onde Botafogo e Igarassu são os rios que mais contribuem em termos de volume de água. A água dos rios transportam grande quantidade de nutrientes e matéria orgânica para o estuário, que junto ao tempo de residência (ou retenção) da água, propicia condições para o desenvolvimento dos produtos primários e de uma cadeia alimentar ativa. Por isso, esse ambiente constitui um dos principais fornecedores de alimentos para organismos bentônicos, epi-bentônicos e pelágicos (SILVA, 2008).

Figura 1 – Mapa da localização do Canal de Santa Cruz.



Fonte: (PAIVA et al., 2005)

O Canal de Santa Cruz é também caracterizado por apresentar um clima pseudo-tropical, quente e úmido (GRIFFITHS, 1996), chove durante todo ano com totais anuais superiores a 1.500 mm, cujo período chuvoso é de março a agosto e o período seco de setembro a fevereiro (SANTOS et al., 2009). De acordo com Vasconcelos-Filho (1999) no período chuvoso, grande quantidade de água doce, proveniente dos rios adjacentes, penetram no Canal, alterando a salinidade, temperatura, transparência, oxigênio dissolvido e pH (para menos) e elevando os valores de sais nutrientes: nitritos, nitratos e fosfatos. A salinidade apresenta uma variação anual de 18⁰/₀₀ a 34⁰/₀₀ , mas nos meses de chuva é sempre inferior a 24⁰/₀₀ (VASCONCELOS-FILHO, 1999). A temperatura da água também apresenta variações sazonais, ficando entre 25° C e 31° C (MACÊDO et al, 1998).

Tais características levam ao desenvolvimento de um tipo especial de ecossistema no CSC, os chamados manguezais (RUSSELL, 1967). O seu desenvolvimento só ocorre em uma série de condições específicas associadas, tais como: clima quente úmido, solo areno-lamoso, salinidade variando entre 5⁰/₀₀ a 30⁰/₀₀, e precipitação pluviométrica acima de 1500 mm/ano (SILVA, 2008). O manguezal é uma zona ecológica de grande importância, por apresentar condições favoráveis para o abrigo e desenvolvimento da biota flúvio-marinha, que passa nele pelo menos uma parte do ciclo de sua vida. É um ambiente que serve como maternidade, berçário, área de alimentação, e, proporciona condições favoráveis para a procriação de diversas espécies aquáticas (SILVA, 2008).

Pesca no estado de Pernambuco

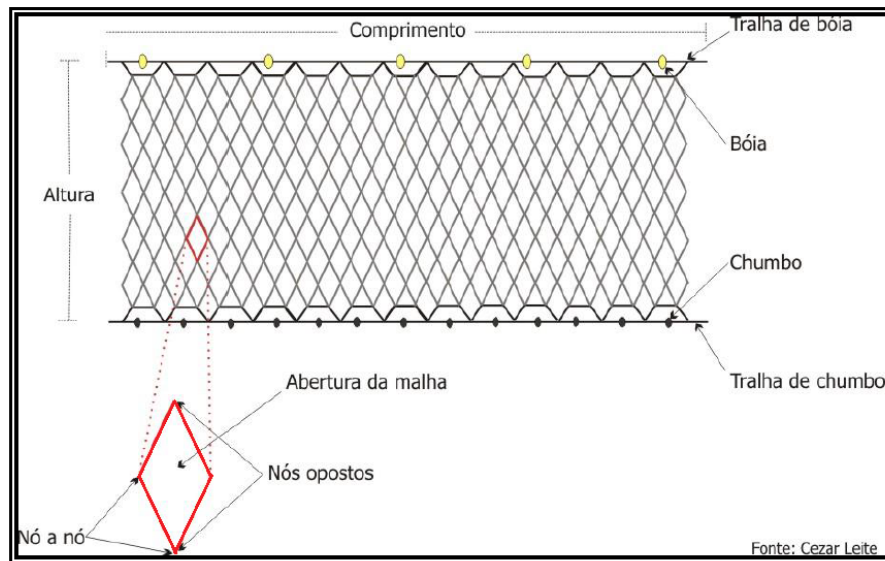
Parte da produtividade pesqueira em Pernambuco, é advinda da pesca artesanal², com aproximadamente 98,97% de contribuição em 2007 (IBAMA, 2008) se comparada às pescarias industriais. No município de Itapissuma a pesca artesanal representa uma importante fonte de subsistência, senão a única, para uma grande parcela da população marginal do CSC; cujos principais recursos são pequenos peixes pelágicos, especialmente manjubas, agulhinhas, saúnas e sardinhas, que constituem 61,2% dos desembarques (LESSA et al, 2011), capturados com rede de emalhe "redinha"³(Figura 2) ou com arrasto manual "mangote".

As redes mais utilizadas nas pescarias no CSC são as de tamanho de malha de 30 mm (entre nós opostos), conhecida como redinha Figura (2), a de 40 mm (entre nós opostos), conhecida como sauneiro e a de 50 mm (entre nós opostos), chamada na localidade de sauneiro-laço. No entanto, a rede de 30 mm é mais utilizada na localidade, direcionada à captura de um dos recursos mais importante do CSC, a manjuba (*O. oglinum*). Redes

² Pesca artesanal - caracteriza-se a favor da exploração de recursos pesqueiros utilizando tecnologia simples na captura do pescado, com produção em baixa escala

³ Redinha - é a designação frequentemente empregada pelos pescadores a rede de emalhar (entre nós opostos)

Figura 2 – Desenho esquemático da rede de emalhar.



Fonte: (DSPE, 2010)

de emalhar são também muito utilizadas em atividades de pesquisa, pois são úteis para monitorar a distribuição do comprimento de captura (HAMLEY, 1975), determinar a estrutura de tamanho da população, e avaliar o efeito da pesca em um estoque já explorado (MCAULEY et al., 2007).

Classificação, características e distribuição da manjuba (*Opisthonema oglinum*)

De acordo com Lesueur (1818), a *Opisthonema oglinum* (Figura 3) é classificada da seguinte maneira:

Reino Animalia

Filo Chordata

Classe Actinopterygii

Ordem Clupeiformes

Família Clupeidae

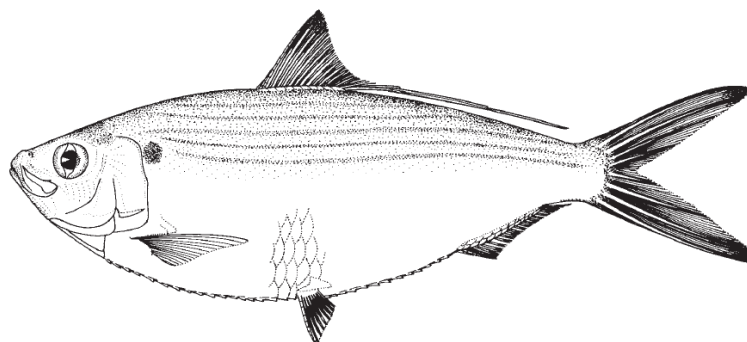
Genero *Opisthonema*

Espécie *O. oglinum*

A *Opisthonema oglinum* apresenta um corpo fusiforme⁴, moderadamente comprimido (CARPENTER, 2002). O último raio da nadadeira dorsal é prolongado, alcança o

⁴ Fusiforme - alongado e cujo centro é mais espesso, com as extremidades mais estreitas

Figura 3 – Exemplar de uma *Opisthonema oglinum*.



Fonte: (CARPENTER, 2002)

pedúnculo caudal, característica que distingue esta sardinha das demais (FIGUEIREDO & MENEZES, 1978). As nadadeiras pélvicas são pequenas e estão localizadas na metade do comprimento padrão⁵ (FIGUEIREDO & MENEZES, 1978). Seu corpo tem coloração azul esverdeada dorsalmente e algumas vezes, apresenta linhas escuras horizontalmente, coloração prateada na lateral inferior e no abdômen com uma mancha negra atrás da margem superior do opérculo⁶ (CARPENTER, 2002).

Segundo Couto & Vasconcelos-Filho (1986) a *O. oglinum* é uma espécie filtradora ou predadora, com a coleta por filtração muito difundida. O aparelho filtrador destes peixes, é constituído por raios espinhosos sustentados pelos arcos branquiais, que formam uma verdadeira rede de filtração. Logo, os organismos planctônicos, constituem o alimento essencial na dieta deste clupeídeo, constatando-se que a espécie é planctófaga. Apesar de ser essencialmente planctófaga, encontra-se alimentos de hábito Nectônico e Bentônico na dieta de exemplares adultos. (VASCONCELOS-FILHO, 1979)

De hábito pelágico costeiro, a sardinha-laje habita em águas com profundidade entre 3m e 35m (KINNEAR & FUSS, 1971; LESSA et al., 2004) cuja preferência é por regiões de águas com temperaturas relativamente elevadas, entre 26°C e 29°C, estando ausente em águas com temperaturas inferiores a 17°C (CARPENTER, 2002). A espécie está presente em toda a costa atlântica americana, desde a Nova Inglaterra até a Argentina (FIGUEREIDO & MENEZES, 1978) e é única espécie representante do gênero *Opisthonema* nesta costa. No Nordeste brasileiro distribui-se do Nordeste da foz do rio Parnaíba, Piauí até as proximidades de Salvador, Bahia e apresenta maior abundância no estado do Ceará, onde são registrados os desembarques mais expressivos (LESSA et al., 2004).

⁵ Comprimento padrão - distância da ponta do focinho à base da nadadeira caudal (extremidade da coluna vertebral, evidenciada por uma dobra formada ao se curvar a nadadeira caudal para os lados)

⁶ Estrutura presente nos peixes ósseos responsável pela proteção das brânquias.

Em Pernambuco, mais precisamente no Canal de Santa Cruz, a *Opisthonema oglinum* foi o clupeídeo mais abundante em 2006, totalizou 49,4% de todo o pescado capturado no município de Itapissuma (ANDRADE & SILVA, 2013). Segundo Lino (2003) os exemplares capturados no CSC menores de 20 cm são destinados a salga e conhecidos por manjuba, enquanto os exemplares maiores são chamados de sardinha-laje e vendidos frescos ou resfriados. Além de constituir-se como fonte de alimento abundante para os pescadores da região, a espécie sustenta a cadeia trófica do estuário, sendo também uma fonte de alimento para muitas espécies. De acordo com a classificação ecológica proposta por Lino (2013), a *O. oglinum* no CSC, foi incluída como pertencente a uma espécie de origem marinha visitante, ou seja, peixes que passam parte do seu ciclo de vida no mar, e penetram em estuários em períodos regulares ou ocasionalmente. Geralmente as espécies marinhas visitantes são encontradas em estuários em estágios jovens, o que torna o canal importante como área de berçário (VASCONCELOS-FILHO & OLIVEIRA, 1999).

A *O. oglinum* é uma espécie de pequeno porte, com tamanho semelhante ao das sardinhas comuns. O comprimento máximo encontrado foi de 380 milímetros (CARPENTER, 2002) nos EUA, 227 mm no México (MEXICANO-CÍNTORA et.al, 1996), 298 mm na Venezuela (GONZÁLEZ-CABELLOS & MENQUAL-IZQUIERDO, 1995), 310 mm no Nordeste brasileiro (LESSA et.al, 2004), 220 mm em Itapissuma - Pernambuco (LINO, 2013). Segundo Frehse (2006) a espécie foi encontrada no complexo estuarino de Paranaguá quase que exclusivamente na fase juvenil, uma vez que todos os indivíduos analisados mostraram-se sexualmente imaturos. Lino (2003) notou que 92,7% dos exemplares capturados na região de Itapissuma estavam abaixo do comprimento médio de primeira maturação⁷ $L_{pm} = 117mm$.

A época de atividade reprodutiva da *O. oglinum* na plataforma Sul do México segundo García-Abad et al. (1998) foi entre maio e outubro, com dois picos de desova, em maio e agosto, com recrutamento dos juvenis principalmente no mês de junho, sendo observado um segundo pico em outubro. Para os indivíduos na região Sudeste e Sul do Brasil Feltrim & Schwingel (2005) concluíram que houve um aumento da atividade reprodutiva a partir de novembro (primavera), estendendo-se durante o verão. Em relação ao período de recrutamento da espécie no Nordeste, Lessa et al. (2004), encontraram um período de recrutamento entre junho e fevereiro, com maior intensidade nos meses de dezembro e janeiro. Já Simoni (2019) afirma que a sardinha apresenta desova parcelada em todo ano no litoral norte de Pernambuco, mas com picos pontuais após o período chuvoso. Fator que pode estar relacionado com a variação da salinidade na zona costeira, que junto a temperatura, são fatores que têm papel fundamental no período de incubação

⁷ Comprimento médio de primeira maturação L_{pm} - indica o comprimento em que os indivíduos estão aptos a reproduzir.

e eclosão de ovos, estando diretamente relacionado a desova de muitas espécies (RIBEIRO & MOREIRA, 2012; SIMONI, 2019).

Sobrepesca

Devido ao declínio nos estoques da sardinha-verdadeira, *Sardinella brasiliensis* exploradas no Nordeste, a pesca de *Opisthonema oglinum* tem sido uma alternativa importante em algumas localidades, uma vez que apresenta qualidades nutricionais, gustativas e visuais muito semelhantes às da sardinha-verdadeira (FELTRIM & SCHWINGEL, 2005). Há indícios de que estoques da manjuba estejam em níveis de exploração acima do limite máximo sustentável (LESSA et al., 2004). No Brasil a espécie apresentou um declínio, nos últimos anos, passando de 13.252 toneladas capturadas em 2007 para 8.810 t em 2011 (MPA, 2012). Além disso, a espécie apresentou uma evolução na lista vermelha da IUCN, passando de dados deficientes para pouco preocupante, chegando perto de se tornar uma espécie ameaçada pela sobreexploração (SIMONI, 2019)

Diante da alta captura de peixes, observa-se que a maioria dos mananciais que entraram em colapso em décadas recentes, sofreram devido á sobrepesca. A sobrepesca é expressa como a situação pela qual a atividade pesqueira de uma espécie deixa de ser “sustentável”. Segundo Palazzo e Carbogim (2012) o conceito de sustentabilidade na atividade pesqueira é definido pelo esforço de captura que deve ser limitado pela capacidade de suporte do ambiente. A sobrepesca contínua pode levar a extinção de exemplares adultos de uma população ou estoque. Se houver exploração sem limite dos adultos reprodutores (desovantes), não haverá novos descendentes e, conseqüentemente, não será mantida a taxa de renovação da população. Neste caso, ocorre a chamada sobrepesca de recrutamento. Da mesma forma, se houver pesca de jovens (imaturos), não será possível que eles atinjam o tamanho no qual o rendimento em peso seja máximo, ocasionando a sobrepesca de crescimento (PALAZZO & CARBOGIM, 2012). Em contra partida, se a sobrepesca for controlada e os recursos pesqueiros forem manejados de forma sustentável, o comprimento médio aumentará, as áreas pesqueiras se tornarão mais produtivas, o preço final do recurso cairá e a captura aumentará de maneira sustentável.

3 Objetivos

3.1 Objetivo Geral

Avaliar a existência de variação espacial e sazonal no comprimento da *Opisthonema oglinum* no Canal de Santa Cruz através dos modelos de regressão no intuito de auxiliar em investigações mais aprofundadas para que haja gestão sustentável do recurso.

3.2 Objetivos Específicos

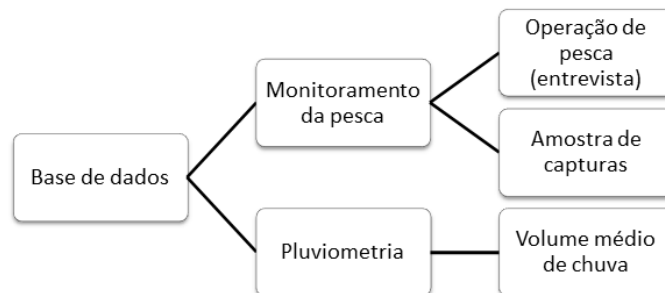
- Ajustar Modelos Lineares Generalizados (GLM) e Modelos Aditivos Generalizados para Localização, Escala e Forma (GAMLSS) para a análise do comprimento padrão da *Opisthonema oglinum*
- Selecionar o modelo que apresenta melhor desempenho quanto a qualidade de ajuste e conformidade dos resíduos
- Interpretar o significado das estimativas obtidas e levantar hipóteses sobre processos ecológicos que sejam de utilidade para profissionais na área da pesca para tomadas de decisões

4 Metodologia

4.1 Material

A base de dados utilizada nesta pesquisa é proveniente da junção de dois bancos de dados (Figura 5). O primeiro referente à informações sobre a pluviometria disponível na APAC – Agência Pernambucana de Águas e Clima, e o segundo está relacionado ao monitoramento da pesca no CSC entre os anos de 2013 à 2015, organizados no Laboratório de Modelagem Estatística Aplicada (MOE) do Departamento de Pesca e Aquicultura da URFPE, Recife-PE.

Figura 4 – Diagrama da junção de bancos de dados utilizados no trabalho.

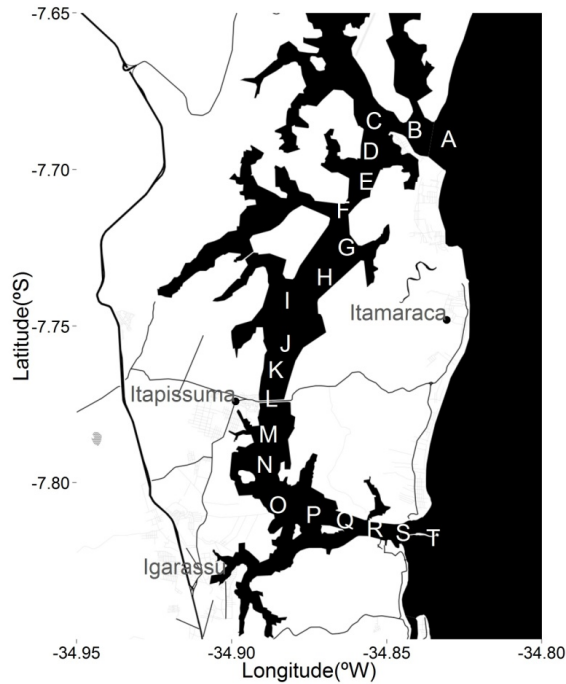


Fonte: Elaborada pelo Autor

O monitoramento da pesca foi realizado por meio de duas abordagens principais: a coleta de informações sobre a dinâmica da operação de pesca realizada a partir de visitas semanais ao CSC, em que, ao desembarcarem, os pescadores eram entrevistados; e a obtenção de amostras das capturas com o uso de rede com malha de 30 mm (entre nós opostos). Com o intuito de ajudar na identificação das posições aproximadas da operação de pesca para registrar nos questionários, o CSC foi subdividido em setores desde a desembocadura norte (setor A) até a desembocadura sul (setor T) (Figura 6). Com o auxílio de um *Global Positioning System* (GPS), as amostras foram coletadas com maior precisão e, logo em seguida, transportadas para o Laboratório de Modelagem Estatística Aplicada (MOE), onde foram congeladas e, posteriormente, identificadas segundo chaves taxonômicas como as publicadas por Figueiredo e Menezes (1978) e Carpenter (2002). Para cada exemplar ($n = 10.394$), foi ainda registrado o comprimento padrão com precisão

de 0,01 mm.

Figura 5 – Canal de Santa Cruz com a indicação da setorização espacial utilizada nas análises.

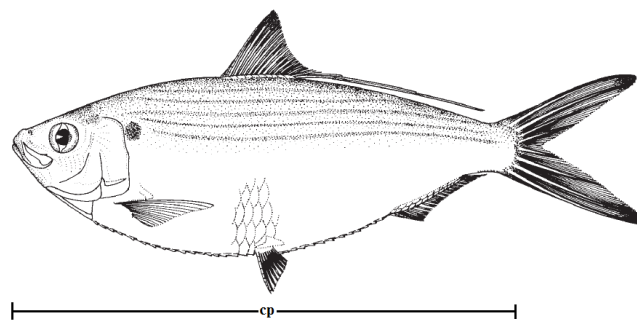


Fonte: Elaborado pelo autor

Resumidamente, as variáveis abordadas na pesquisa são:

- **Comprimento padrão (cp)** - Variável contínua caracterizada pela distância da ponta do focinho à base da nadadeira caudal (Figura 6).

Figura 6 – Descrição do comprimento padrão (cp) da *Ophistonema oglinum*.



Fonte: (FONTELES FILHO, 2011)

- **Mês** - Variável qualitativa que identifica o mês em que os dados foram coletados. O fator mês foi escolhido com o propósito de avaliar efeitos a curto prazo da variável resposta (janeiro a dezembro).
- **Ano** - Variável qualitativa aponta o ano em que os dados foram coletados. O fator ano foi optado com a intenção de avaliar efeitos a longo prazo da variável resposta (2013 a 2015).
- **Setor** - Variável qualitativa remete à setorização do CSC, desde a desembocadura norte (setor A) a desembocadura sul (setor T), conforme pode ser visto na Figura (5).
- **Maré de lançamento** - Variável qualitativa referente ao tipo da maré no lançamento da rede no momento da captura dos peixes. Os níveis do fator maré de lançamento foram: a) baixamar - nível mínimo atingido pelas águas após a vazante; b) enchente - período entre uma baixamar e uma preamar; c) preamar - o nível máximo atingido pelas águas de uma maré cheia; d) vazante - período entre uma preamar e uma baixamar, quando a altura das águas diminui.
- **Precipitação** - Variável contínua, representada pelo volume médio acumulado de chuva em 30 dias antes da coleta das amostras nas estações Itapissuma, Igarassu e Itamaracá.

4.2 Introdução a modelagem

Segundo Khuri (2010) os Modelos Normais Lineares têm a forma

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (4.1)$$

em que, \mathbf{Y} é um vetor de variáveis resposta conhecidas, \mathbf{X} é a matriz de variáveis explicativas conhecidas, $\boldsymbol{\beta}$ é um vetor de parâmetros desconhecido e $\boldsymbol{\epsilon}$ é um vetor de erros aleatórios do modelo. Para a utilização adequada de tais modelos, algumas pressuposições devem ser satisfeitas, como por exemplo, os erros devem ser normais, independentes e identicamente distribuídos de variância constante. Caso as pressuposições não sejam verificadas, existe a possibilidade de transformação de variáveis. O método de transformação de variáveis mais conhecido foi proposto por *Box e Cox* (1964), muito utilizado para que a distribuição da variável em questão se aproxime de uma normal e a variância se estabilize.

Para ajudar a solucionar problemas relacionados à violação de pressupostos, Nelder e Wedderburn (1972) desenvolveram a teoria dos modelos lineares generalizados, uma extensão dos modelos lineares clássicos. Semelhante aos modelos normais lineares, nos

Modelos Lineares Generalizados (*Generalized Linear Models* - GLM) as observações devem ser independentes, não existindo a necessidade de satisfazer a suposição de normalidade da variável resposta. Por utilizar os dados originais, a ideia básica dos modelos lineares generalizados consiste em abrir um leque de opções para a distribuição da variável resposta, permitindo que a mesma pertença à família exponencial de distribuições.

Família Exponencial

Seja $y = (y_1, \dots, y_n)'$ um vetor de observações referente às realizações das variáveis aleatórias $Y = (Y_1, \dots, Y_n)'$ cada uma com função densidade ou função de probabilidade pertencente à família exponencial deve ser expressa por

$$f_Y(y_i; \theta; \phi) = \exp\{\phi[y_i\theta_i - b(\theta_i)] + c(y_i, \phi)\}, \quad i = 1, \dots, n \quad (4.2)$$

em que, $E(Y_i) = \mu_i = b'(\theta_i)$ e $Var(Y_i) = \phi^{-1}V_i$, cujo $V_i = V(\mu_i) = \frac{d\mu_i}{d\theta_i}$, é a função de variância, $\phi^{-1} > 0$ é o parâmetro de dispersão e θ é o parâmetro canônico. Uma vez que se conhece a função de variância especificada como $V(\mu_i) = b''(\theta_i)$, é possível determinar qual será a classe de distribuição correspondente na família exponencial. Segundo Fisher (1934), a família exponencial de distribuição garante aos parâmetros estatísticas suficientes.

Mesmo assim, o uso dos GLM apresentam limitações, pois, afirma-se que a relação da média é linear às variáveis explicativas. Uma alternativa para resolver este problema é o uso de regressão não-paramétrica ou semi-paramétrica, abordagens muitas vezes superiores por sua flexibilidade, em detrimento às estruturas estritamente paramétricas (FLORENCIO, 2010). Dentre tais modelos, destacam-se os Modelos Aditivos Generalizados (*Generalized Additive Models* - GAM) introduzido por Hastie e Tibshirani (1990) que configuram uma extensão dos GLM com um preditor linear incluindo a soma de funções de suavização não-paramétrica das covariáveis configurando a possibilidade de os próprios dados guiarem sua relação com o preditor (η), que geralmente, decorre de maneira não-linear.

Com o intuito de viabilizar modelos com pressupostos ainda mais flexíveis na especificação da distribuição da variável resposta, Rigby e Stasinopoulos (2005) introduziram os Modelos Aditivos Generalizados para Localização, Escala e Forma, técnica essa que, proporciona o ajuste de modelos que admitem inúmeras distribuições para a variável resposta, sem pertencer obrigatoriamente à família exponencial. Em suma, os GAMLSS correspondem a uma modelagem apurada e flexível, propõem uma nova classe de modelos, que possuem como pressuposto que as observações sejam independentes e conseguem reunir um grupo de modelos lineares e não-lineares, modelos lineares generalizados (NELDER & WEDDERBURN, 1972), modelos aditivos generalizados (HASTIE & TIBSHIRANI, 1990) e modelos mistos.

4.3 Modelos Lineares Generalizados

Nos Modelos de Regressão Linear, a variável resposta é representada por uma combinação linear das suas covariáveis e um erro aleatório, em que tal erro, supomos que siga distribuição normal, logo, a variável resposta também deve seguir distribuição normal. No entanto, na prática isso muitas vezes não ocorre. Então, em 1972 uma extensão do modelo de regressão linear foi proposta por Nelder e Wedderburn, os Modelos Lineares Generalizados (*Generalized Linear Models* - GLM), que admitem a não normalidade nos erros e, portanto, na variável resposta. Estes modelos podem ser utilizados quando a distribuição da variável de interesse é qualquer distribuição pertencente a família exponencial. Tais modelos são compostos por dois componentes e uma relação funcional entre eles (CORDEIRO & DEMÉTRIO, 2013):

Componente aleatório: a variável resposta é representada por um conjunto de variáveis aleatórias independentes Y_1, \dots, Y_n obtidas de uma mesma distribuição de probabilidade pertencente a família exponencial de distribuições com médias μ_1, \dots, μ_n ,

$$E(Y_i) = \mu_i, \quad i = 1, 2, \dots, n. \quad (4.3)$$

Componente sistemático: representado pelas variáveis explicativas, que entram na forma de uma soma linear de seus efeitos

$$\eta_i = \sum_{j=1}^p x_{ij}\beta_j, \quad i = 1, 2, \dots, n, \quad (4.4)$$

ou, matricialmente,

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}, \quad (4.5)$$

em que, $\boldsymbol{\eta}$ é o vetor dos preditores lineares, \mathbf{X} a matriz de variáveis explicativas ($n \times p$) e $\boldsymbol{\beta}$ é o vetor ($p \times 1$) de parâmetros a serem estimados.

Função de ligação: uma função $g(\cdot)$, monótona e diferenciável, que relaciona o componente aleatório (μ_i) ao componente sistemático (η_i), ou melhor, vincula a média ao preditor linear,

$$g(\mu_i) = \eta_i. \quad (4.6)$$

Se a função de ligação é escolhida de modo que $g(\mu_i) = \theta_i = \eta_i = \sum_{j=1}^p x_{ij}\beta_j$, o preditor linear modela diretamente o parâmetro canônico θ_i , logo, tal função é denominada ligação canônica (Tabela 1). As funções de ligação canônicas produzem propriedades estatísticas de interesse para o modelo, tais como: suficiência, facilidade de cálculo, unicidade das estimativas de máxima verossimilhança e, em alguns casos, interpretação mais simples (CORDEIRO & DEMÉTRIO, 2013).

Tabela 1 – Funções de ligação canônica para as principais distribuições de probabilidade

Distribuição	Função de ligação canônica
Normal	Identidade: $\eta = \mu$
Poisson	Logarítmica: $\eta = \log(\mu)$
Binomial	Logística: $\eta = \log\left(\frac{\mu}{n-\mu}\right)$
Gama	Recíproca: $\eta = \frac{1}{\mu}$
Normal Inversa	Recíproca do quadrado: $\eta = \frac{1}{\mu^2}$

4.3.1 Estimação

Supondo ϕ conhecido, o logaritmo da função de verossimilhança de um MLG é dado por

$$L(\boldsymbol{\beta}) = \sum_{i=1}^n \phi[y_i \theta_i - b(\theta_i)] + \sum_{i=1}^n c(y_i, \phi).$$

Caso a função de ligação escolhida seja a canônica, temos $g(\mu_i) = \theta_i = \eta_i = \sum_{j=1}^p x_{ij} \beta_j$. Desta forma, $L(\boldsymbol{\beta})$ é representado por

$$L(\boldsymbol{\beta}) = \sum_{i=1}^n \phi\left[y_i \sum_{j=1}^p x_{ij} \beta_j - b\left(\sum_{j=1}^p x_{ij} \beta_j\right)\right] + \sum_{i=1}^n c(y_i, \phi).$$

Definindo a estatística $S_j = \phi \sum_{i=1}^n Y_i x_{ij}$, $L(\boldsymbol{\beta})$ fica na forma

$$L(\boldsymbol{\beta}) = \sum_{j=1}^p s_j \beta_j - \sum_{i=1}^n b\left(\sum_{j=1}^p x_{ij} \beta_j\right) + \sum_{i=1}^n c(y_i, \phi).$$

Então, a estatística $S = (S_1, \dots, S_p)$ é suficiente minimal para $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ pelo teorema da fatoração.

Algoritmo de estimação

Desenvolvido por Nelder e Wedderburn (1972) o algoritmo de estimação dos parâmetros β 's baseia-se em um método semelhante ao de Newton-Raphson, conhecido como Método Escore de *Fisher*. Para o desenvolvimento do método obtém-se a derivada de primeira ordem de $L(\boldsymbol{\beta})$ conhecido por função escore e calcula-se a informação de *Fisher* cujas expressões podem ser representadas respectivamente por

$$U(\boldsymbol{\beta}) = \frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \phi^{-1} \mathbf{X}' \mathbf{W}^{1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu}), \quad (4.7)$$

$$K(\boldsymbol{\beta}) = \left[-E \left(\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_s} \right) \right] = -E \left(\frac{\partial U(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right) = \phi^{-1} \mathbf{X}' \mathbf{W} \mathbf{X}. \quad (4.8)$$

em que, $\mathbf{W} = \text{diag}\{w_i\}$ é a matriz dos pesos, cujo i -ésimo peso é dado por $w_i = V(\mu_i)^{-1} g'(\mu_i)^{-2}$ e $\mathbf{V} = \text{diag}\{V_1, V_2, \dots, V_n\}$ a matriz das funções de variância.

Segundo Dobson (2002) para obter as estimativas de máxima verossimilhança de β por meio do processo iterativo de *Newton-Raphson*, realiza-se uma expansão na função escore $U(\beta)$ baseada na aproximação da série de *Taylor* em torno de um valor inicial $\beta^{(0)}$, dado por

$$\mathbf{U}(\beta) \cong \mathbf{U}(\beta^{(0)}) + \mathbf{U}'(\beta^{(0)})(\beta - \beta^{(0)}),$$

em que, $\mathbf{U}'(\beta^{(0)})$ é a derivada de primeira ordem da função escore em relação a β , no ponto $\beta^{(0)}$. Logo, utiliza-se a aproximação acima para especificar o processo iterativo em

$$\beta^{(m+1)} = \beta^{(m)} + [-\mathbf{U}'(\beta^{(m)})]^{-1}\mathbf{U}'(\beta^{(m)}), \quad m = 0, 1, \dots$$

em que, m é a m -ésima interação, $\beta^{(m)}$ é o vetor de parâmetros na interação e no passo m . Pelo fato da matriz $-\mathbf{U}'(\beta^{(m)})$ não ser positiva definida, por conveniência, substitui-se pelo correspondente valor esperado $\mathbf{K}(\beta)$, no ponto $\beta^{(0)}$. Então, resulta no seguinte processo iterativo

$$\beta^{(m+1)} = \beta^{(m)} + [\mathbf{K}(\beta^{(m)})]^{-1}\mathbf{U}'(\beta^{(m)}), \quad m = 0, 1, \dots \quad (4.9)$$

A partir do resultado da expressão (4.9), temos a versão final do algoritmo, então, o processo iterativo de mínimos quadrados ponderados é dado por

$$\beta^{(m+1)} = (\mathbf{X}'\mathbf{W}^{(m)}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{(m)}\mathbf{z}^{(m)}, \quad m = 0, 1, \dots \quad (4.10)$$

em que, $\mathbf{W}^{(m)} = \text{diag}\{w_1^{(m)}, w_2^{(m)}, \dots, w_n^{(m)}\}$ é a matriz de pesos que muda a cada passo do processo iterativo, cujos pesos são representados por $w_i^{(m)} = V(\mu_i^{(m)})^{-1}g'(\mu_i^{(m)})^{-2}$, $\mathbf{z}^{(m)}$ é o vetor da variável resposta modificada ajustada na m -ésima interação, onde \mathbf{z} é denotada por

$$\mathbf{z} = \boldsymbol{\eta} + \mathbf{W}^{(-1/2)}\mathbf{V}^{(-1/2)}(\mathbf{y} - \boldsymbol{\mu}). \quad m = 0, 1, \dots \quad (4.11)$$

De acordo com Paula (2013) a convergência de (4.10) ocorre em um número finito de passos, independente dos valores iniciais utilizados, entretanto, é usual iniciar o processo com $\boldsymbol{\eta}^{(0)} = \mathbf{g}(\mathbf{y})$

4.3.2 Medidas de qualidade de Ajuste (Função desvio)

Pode-se construir modelos com até n parâmetros, tal modelo é normalmente chamado de saturado e contém $p = n$ parâmetros, um para cada observação. Na prática, o modelo nulo ($p = 1 = \mu$) é muito simples, enquanto o modelo saturado é não-informativo. Entretanto, o modelo saturado é útil para medir a discrepância de um modelo intermediário (sob investigação) com ($p < n$) parâmetros.

A função desvio, comumente conhecida por *deviance* é uma das medidas mais usadas para avaliar a qualidade do ajuste de um modelo linear generalizado. Tal medida

equivale à diferença entre o logaritmo da função de verossimilhança dos modelos saturados e sob investigação. Segundo Paula (2013) considera-se ϕ conhecido, logo, o logaritmo da função de verossimilhança do modelo sob investigação é expresso por

$$L(\hat{\boldsymbol{\mu}}; \mathbf{y}) = \sum_{i=1}^n L(\mu_i; y_i) = \sum_{i=1}^n \phi [y_i \hat{\theta}_i - b(\hat{\theta}_i)] \sum_{i=1}^n c(y_i, \phi), \quad (4.12)$$

em que, $\hat{\theta}_i = \theta(\hat{\mu}_i)$ é a estimativa de máxima verossimilhança do parâmetro canônico do modelo sob investigação, $L(\hat{\boldsymbol{\mu}}; \mathbf{y})$ é a estimativa de $L(\boldsymbol{\mu}; \mathbf{y})$. Aqui, a estimativa de máxima verossimilhança de μ_i será dada por $\hat{\mu}_i = g^{-1}(\hat{\eta}_i)$ e $\hat{\eta}_i = \mathbf{x}_i' \hat{\boldsymbol{\beta}}$. Para o modelo saturado, temos o logaritmo da função de verossimilhança estimado por

$$L(\mathbf{y}; \mathbf{y}) = \sum_{i=1}^n L(y_i; y_i) = \sum_{i=1}^n \phi [y_i \tilde{\theta}_i - b(\tilde{\theta}_i)] \sum_{i=1}^n c(y_i, \phi), \quad (4.13)$$

ou seja, a estimativa de máxima verossimilhança de μ_i fica nesse caso dada por $\hat{\mu}_i = y_i$. Então temos,

$$\begin{aligned} D_{ev}(\mathbf{y}; \hat{\boldsymbol{\mu}}) &= \phi D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2[L(\mathbf{y}; \mathbf{y}) - L(\hat{\boldsymbol{\mu}}; \mathbf{y})] \\ &= 2 \sum_{i=1}^n (\tilde{\theta}_i - \hat{\theta}_i) + [b(\tilde{\theta}_i) - b(\hat{\theta}_i)], \end{aligned}$$

em que, $D_{ev}(\mathbf{y}; \hat{\boldsymbol{\mu}})$ é a função desvio, distância entre o logaritmo da função de verossimilhança do modelo saturado e o modelo sob investigação. Com a obtenção de um valor significativamente pequeno da *deviance*, tem-se evidência de que, para um número menor de parâmetro, obtém-se um ajuste tão bom quanto o ajuste com o modelo saturado (DOBSON, 2002; PAULA, 2013).

4.3.3 Critério de seleção dos modelos

A idéia básica é selecionar um modelo que esteja bem ajustado e tenha um número reduzido de parâmetros (princípio da parcimônia). Com o aumento do número de parâmetros no modelo, o logaritmo da função de verossimilhança cresce exponencialmente, logo, uma proposta razoável seria utilizar o Critério de Informação Akaike (AIC) (AKAIKE, 1974), definido por

$$AIC = D_{ev}(\mathbf{y}; \hat{\boldsymbol{\mu}}) + 2p, \quad (4.14)$$

em que, $D_{ev}(\mathbf{y}; \hat{\boldsymbol{\mu}})$ é o desvio do modelo e p o número de parâmetros. A utilização do critério consiste em encontrarmos o modelo tal que a quantidade acima seja o menor possível, isto é, um valor baixo para AIC é considerado como representativo de um melhor ajuste.

4.3.4 Análise de resíduos

Uma das etapas mais importante na modelagem estatística é a verificação de possíveis afastamentos das suposições feitas para o modelo, tanto para o componente aleatório, quanto para a parte sistemática do modelo, bem como a existência de observações discrepantes com alguma interferência desproporcional ou inferencial nos resultados do ajuste. Segundo Paula (2013), o processo da análise de resíduos serve para detectar a presença de pontos aberrantes e avaliar a adequação da distribuição proposta para a variável resposta.

Ponto de Alavanca

A idéia principal de ponto de alavanca (leverage) é de avaliar a influência de y_i sobre o próprio valor ajustado \hat{y}_i . A definição de leverage é construída por meio de uma analogia entre a solução de máxima verossimilhança para $\hat{\beta}$ num MLG e a solução de mínimos quadrados de regressão normal ponderada (PAULA, 2013). Diante disto, nota-se que na convergência do processo iterativo dado em (4.10), tem-se o seguinte

$$\hat{\beta} = (\mathbf{X}'\hat{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{W}}\hat{\mathbf{z}}, \quad (4.15)$$

em que, $\hat{\mathbf{z}} = \hat{\boldsymbol{\eta}} + \hat{\mathbf{W}}^{-1/2}\hat{\mathbf{V}}^{-1/2}(\mathbf{y})\hat{\boldsymbol{\mu}}$. Logo a matriz de projeção da solução de mínimos quadrados da regressão linear de \mathbf{z} contra \mathbf{X} com pesos \mathbf{W} proposta por Pregibon (1981) fica representada por

$$\mathbf{H} = \mathbf{W}^{1/2}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{1/2}, \quad (4.16)$$

o que é equivalente a substituir \mathbf{X} por $\mathbf{W}^{1/2}\mathbf{X}$. Nota-se que \mathbf{H} , depende das variáveis explicativas, da função de ligação e da função de variância, tornando mais difícil a interpretação da medida de leverage. Então utilizam-se os valores da diagonal principal da matriz de projeção (h_{ii}) como uma medida da influência da i -ésima observação sobre o próprio valor ajustado. Supondo que todos os pontos exerçam a mesma influência sobre os valores ajustados, pode-se esperar que (h_{ii}) esteja próximo de $tr(\mathbf{H})/n = p/n$. Segundo Paula (2013), convém investigar pontos tais que (h_{ii}) $> 2p/n$, conhecidos como pontos de alavanca, que geralmente localizam-se em regiões remotas no subespaço gerado pelas colunas da matriz \mathbf{X} . Como em geral (\hat{h}_{ii}) depende de $\hat{\mu}_{ii}$, então, sugere-se utilizar o gráfico de (\hat{h}_{ii}) contra valores ajustados para detectar (observar) possíveis pontos de alavanca.

Ponto aberrante

Os resíduos são importantes para detectar a presença de observações aberrantes que devem ser estudadas detalhadamente (PAULA, 2013). Apresentam-se, a seguir, os tipos de resíduos mais comuns nos MLG.

- **Resíduos ordinários** - O resíduo r_i expressa a discrepância (distância) entre a observação y_i e o seu valor ajustado $\hat{\mu}_i$

$$r_i = y_i - \hat{\mu}_i. \quad (4.17)$$

- **Resíduo de Pearson padornizado**

$$t_{P_i} = \frac{\phi^{1/2} r_i}{\sqrt{V(\hat{\mu}_i)(1 - \hat{h}_{ii})}}, \quad (4.18)$$

em que, $V(\hat{\mu}_i)$ é a função de variância da respectiva distribuição utilizada.

- **Resíduo com componente da função desvio** - Resíduo definido a partir dos componentes da função desvio, têm versão padronizada representada por

$$t_{D_i} = \frac{\phi^{1/2} d(y_i; \hat{\mu}_i)}{\sqrt{(1 - \hat{h}_{ii})}}, \quad (4.19)$$

em que, $d(y_i; \hat{\mu}_i) = \pm \sqrt{2\{y_i(\tilde{\theta}_i - \hat{\theta}_i) + [b(\hat{\theta}_i) - b(\tilde{\theta}_i)]\}}$. O sinal de $d(y_i; \hat{\mu}_i)$, é o mesmo de r_i .

Recomenda-se utilizar o gráfico de algum tipo de resíduo padronizado (t_{P_i} ou t_{D_i}) versus os valores ajustados de tal forma a se ter variância constante para a distribuição em uso. O padrão nulo desse gráfico é uma distribuição dos resíduos em torno de zero com amplitude constante. Desvios sistemáticos podem apresentar algum tipo de curvatura ou, mudança sistemática da amplitude com o valor ajustado (CORDEIRO & DEMÉTRIO, 2013).

Ponto influente

A medida que verifica a influência da retirada de i -ésima observação em β mais utilizada nos MLGs é conhecida por distância de Cook e é definida por

$$D_i \cong \left[\frac{\hat{h}_{ii}}{(1 - \hat{h}_{ii})} \right] \times t_{P_i}^2, \quad (4.20)$$

Recomenda-se plotar os valores de D_i versus valores ajustados para a visualização de pontos supostamente influentes. Segundo Lee (1996) considera-se ponto influente quando o ponto $D_i > p^{-1} \chi_{(p), \alpha}^2$.

Função de ligação

O método para verificar a adequação da função de ligação utilizada no ajuste é o gráfico da variável resposta ajustada estimada $\hat{\mathbf{z}}$ versus o preditor linear estimado $\hat{\boldsymbol{\eta}}$.

O padrão nulo é uma reta, isto é, uma tendência linear indica adequação da função de ligação (CORDEIRO & DEMÉTRIO, 2013; PAULA, 2013).

Gráfico normal e semi-normal de probabilidade

O gráfico normal e semi-normal de probabilidades destaca-se por dois aspectos: a identificação da distribuição originária dos dados e a identificação de valores que se destacam no conjunto (WEISBERG, 2005). McCullagh e Nelder (1989) sugerem o uso do gráfico normal de probabilidades para os resíduos e o gráfico semi-normal de probabilidades para medidas positivas como é o caso de h_{ii} e da distância de Cook.

No caso do gráfico normal de probabilidades para os resíduos, espera-se que na ausência de pontos discrepantes, o aspecto seja linear, mas não há razão para se esperar que o mesmo ocorra quando são usados h_{ii} ou a distância de Cook. Os valores extremos aparecerão nas extremidades do gráfico, possivelmente com valores que desviam da tendência indicada pelos demais. Então, de acordo com Cordeiro e Demétrio (2013) formatos aproximados comuns que indicam que a distribuição utilizada não se adequa bem aos dados são

- **S** - indica distribuições com caudas muito curtas, isto é, distribuições cujos valores estão muito próximos da média;
- **S invertido** - indica distribuições com caudas muito longas e, portanto, presença de muitos valores extremos;
- **J e J invertido** - indicam distribuições assimétricas, positivas e negativas, respectivamente.

Para auxiliar na interpretação do gráfico normal e semi-normal de probabilidades, Atkinson (1985) propôs a adição de um envelope simulado. Esse envelope é tal que, sob o modelo correto, as estatísticas (resíduos, *leverage*, distância de Cook, etc) obtidas a partir das observações ficam inseridas no envelope.

4.4 Modelos Aditivos Generalizados

Os Modelos Aditivos Generalizados (*Generalized Additive Models* - GAM) foram introduzido por Hastie e Tibshirani (1990), e são definidos como uma extensão aos Modelos Lineares Generalizados, basta considerar $\beta_j x_j = f_j(x_j)$. Em analogia aos GLM's temos a substituição do preditor linear

$$\eta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi},$$

por um preditor aditivo

$$\eta_i = \alpha + f_1(x_{1i}) + f_2(x_{2i}) + \dots + f_p(x_{pi}), \quad (4.21)$$

ou, matricialmente

$$\boldsymbol{\eta} = g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + f_1(\mathbf{x}_1) + f_2(\mathbf{x}_2) + \dots + f_p(\mathbf{x}_p),$$

em que, $f_j(x_j)$ são funções de suavização univariada da j -ésima variável explicativa, com $j = 1, 2, \dots, p$. Tais funções de suavização são aplicáveis apenas a variáveis explicativas contínuas (STASINOPOULOS et al., 2017).

Apesar dessas funções $f_j(x_j)$ serem funções suaves, elas podem assumir uma forma paramétrica mais rígida, expandindo esta classe de modelos ao caso paramétrico e semi-paramétrico. Em relação à interpretação dos termos $f_j(x_j)$ é análoga ao coeficiente β_j no caso dos modelos lineares, já que a estimativa $f_j(x_j)$ explica como a resposta muda de acordo com x_j (LIU, 2007).

Para termos não lineares a função $f_j(x_j)$ pode ser estimada por qualquer método de suavização, dentre eles podemos citar *P-splines*, *splines* cúbicas, *loess* e redes neurais (LIU, 2007). A spline é uma curva definida matematicamente por dois ou mais pontos de controle e é, dessa forma que os dados estabelecem a relação entre $\boldsymbol{\eta} = g(\boldsymbol{\mu})$ e as variáveis explicativas, que é o conceito principal para fazer uso de uma função de suavização (EILERS et.al, 2015; LEITE, 2019).

4.5 Modelo Aditivo Generalizado para Localização, Escala e Forma

Hastie e Tibshirani (1990) introduziram nas classes de modelos GLM os parâmetros de suavização, que resultou no termo GAM. Porém, nem todas as limitações foram resolvidas nos ajustes dos modelos GLM e GAM. Desta forma, Rigby e Stasinopoulos (2005) propuseram a classe modelos de regressão semi paramétrico denominada Modelos Aditivos Generalizados para Localização, Escala e Forma (*Generalized Additive Models for Location, Scale and Shape* – GAMLSS) com a vantagem de modelar os parâmetros de assimetria e curtose, que possibilitou uma expansão ao número de distribuições consideradas, pois agora as distribuições não necessitam fazer parte da família exponencial.

Como a parte sistemática dos modelos GAMLSS é expandida para possibilitar a modelagem não apenas da média, mas também dos parâmetros de escala e forma da distribuição da variável resposta, tais parâmetros são modelados em função das variáveis explicativas. Afirma-se que os GAMLSS são modelos de regressão semi-paramétricos. Paramétricos, por se exigir a suposição de distribuição paramétrica para a variável resposta

e ao mesmo tempo semi-paramétricos no sentido de que permitem a modelagem dos parâmetros da distribuição, como funções de variáveis explicativas, por meio de funções de suavização não-paramétricas (FLORENCIO, et al., 2012). Esta nova categoria de modelos (GAMLSS) dispõem dos modelos MRL, GLM e GAM como casos especiais. Neste sentido, para o uso dos GAMLSS impõe-se que as observações da variável resposta Y sejam independentes entre si.

Como a premissa de que a variável resposta deve pertencer à família exponencial nos modelos GAMLSS é relaxada, então, substituiu-se por uma família de distribuições mais geral, dada por

$$\mathbf{Y} \stackrel{ind}{\sim} D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau})$$

$$\begin{aligned} \boldsymbol{\eta}_1 &= g_1(\boldsymbol{\mu}) = \mathbf{X}_1\boldsymbol{\beta}_1 + f_{11}(\mathbf{x}_{11}) + \dots + f_{1J_1}(\mathbf{x}_{1J_1}) \\ \boldsymbol{\eta}_2 &= g_2(\boldsymbol{\sigma}) = \mathbf{X}_2\boldsymbol{\beta}_2 + f_{21}(\mathbf{x}_{21}) + \dots + f_{2J_2}(\mathbf{x}_{2J_2}) \\ \boldsymbol{\eta}_3 &= g_3(\boldsymbol{\nu}) = \mathbf{X}_3\boldsymbol{\beta}_3 + f_{31}(\mathbf{x}_{31}) + \dots + f_{3J_3}(\mathbf{x}_{3J_3}) \\ \boldsymbol{\eta}_4 &= g_4(\boldsymbol{\tau}) = \mathbf{X}_4\boldsymbol{\beta}_4 + f_{41}(\mathbf{x}_{41}) + \dots + f_{4J_4}(\mathbf{x}_{4J_4}) \end{aligned} \tag{4.22}$$

em que, $D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau})$ é uma distribuição de quatro parâmetros (pode ser qualquer distribuição, incluindo distribuições contínuas, com assimetria ou curtose acentuadas e distribuições discretas), $\boldsymbol{\mu}$ é um parâmetro de locação, $\boldsymbol{\sigma}$ é um parâmetro de escala, $\boldsymbol{\nu}$ e $\boldsymbol{\tau}$ são os parâmetros de forma da distribuição, comumente associados à assimetria e curtose, respectivamente. Já $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ e \mathbf{X}_4 são as matrizes que podem ou não coincidir, ou melhor, o preditor de cada parâmetro da distribuição recebe diferentes variáveis explicativas (RIGBY; STASINOPOULOS, 2005). Os β_p são os parâmetros lineares e $f_{pj}(x_{pj})$ representam diferentes funções de suavização para diferentes variáveis explicativas x_{pj} para $p = 1, 2, 3, 4$ e $j = 1, 2, \dots, J_p$.

Os modelos GAMLSS são adequados não apenas para modelagem da variável resposta que não segue uma distribuição da família exponencial (por exemplo, leptocúrtica ou platicúrtica e/ou com assimetria positiva ou negativa), como também em casos que observa-se heterocedasticidade dos resíduos (por exemplo, quando a escala ou a forma da distribuição da variável resposta mudam com as variáveis explicativas) ou esteja relacionado a dados de contagem com sobredispersão (FLORENCIO, et al., 2012).

4.5.1 Estimação

Segundo Stasinopoulos et al. (2017) os modelos GAMLSS são ditos não paramétrico quando são ajustados com funções de suavizações (efeitos aleatórios), e, ainda exige-se a estimação não apenas do $\boldsymbol{\beta}$, mas também, $\boldsymbol{\gamma}$ e $\boldsymbol{\lambda}$. Já no caso paramétrico não ajusta-se o

modelo com funções de suavizações, logo, exige-se apenas a estimação de β , neste caso o modelo resume-se à

$$\begin{aligned} \mathbf{Y} &\stackrel{ind}{\sim} D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau}) \\ \boldsymbol{\eta}_1 &= g_1(\boldsymbol{\mu}) = \mathbf{X}_1\boldsymbol{\beta}_1 \\ \boldsymbol{\eta}_2 &= g_2(\boldsymbol{\sigma}) = \mathbf{X}_2\boldsymbol{\beta}_2 \\ \boldsymbol{\eta}_3 &= g_3(\boldsymbol{\nu}) = \mathbf{X}_3\boldsymbol{\beta}_3 \\ \boldsymbol{\eta}_4 &= g_4(\boldsymbol{\tau}) = \mathbf{X}_4\boldsymbol{\beta}_4 \end{aligned} \quad (4.23)$$

em que, os modelos são ajustados pelas estimativas de máxima verossimilhança em relação à estimação de β . Onde, $\mathbf{Y} \stackrel{ind}{\sim} D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau})$ implica que o logaritmo da função de verossimilhança, determinada pela verossimilhança observada da amostra é,

$$L(\boldsymbol{\beta}) = \sum_{i=1}^n \ln[f(y_i|\mu_i, \sigma_i, \nu_i, \tau_i)] \quad (4.24)$$

de acordo com o pressuposto de independência das observações.

Segundo Stasinopoulos et al. (2017) para os modelos GAMLSS preliminarmente determinados na Equação (4.22), considera-se que a maioria das suavizações podem ser escritas na forma

$$f(\mathbf{x}) = \mathbf{Z}\boldsymbol{\gamma},$$

em que, \mathbf{Z} é a matriz base para a suavização e que depende da variável explicativa \mathbf{x} , $\boldsymbol{\gamma}$ é um vetor de parâmetros a serem estimados, submetidos à penalização $\boldsymbol{\lambda}\boldsymbol{\gamma}^T\mathbf{G}\boldsymbol{\gamma}$, indicando uma matriz conhecida $\mathbf{G} = \mathbf{D}^T\mathbf{D}$, cujo \mathbf{D} indica a matriz de diferenças e $\boldsymbol{\lambda}$ é um vetor ou escalar de hiperparâmetros que ajusta (ou controla) o grau de suavização necessária para o ajuste. Logo, o modelo (4.22) generalizado pode ser reescrito por

$$\begin{aligned} \mathbf{Y}|\boldsymbol{\gamma} &\stackrel{ind}{\sim} D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau}) \\ \boldsymbol{\eta}_1 &= g_1(\boldsymbol{\mu}) = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{Z}_{11}\boldsymbol{\gamma}_{11} + \dots + \mathbf{Z}_{1J_1}\boldsymbol{\gamma}_{1J_1} \\ \boldsymbol{\eta}_2 &= g_2(\boldsymbol{\sigma}) = \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{Z}_{21}\boldsymbol{\gamma}_{21} + \dots + \mathbf{Z}_{2J_2}\boldsymbol{\gamma}_{2J_2} \\ \boldsymbol{\eta}_3 &= g_3(\boldsymbol{\nu}) = \mathbf{X}_3\boldsymbol{\beta}_3 + \mathbf{Z}_{31}\boldsymbol{\gamma}_{31} + \dots + \mathbf{Z}_{3J_3}\boldsymbol{\gamma}_{3J_3} \\ \boldsymbol{\eta}_4 &= g_4(\boldsymbol{\tau}) = \mathbf{X}_4\boldsymbol{\beta}_4 + \mathbf{Z}_{41}\boldsymbol{\gamma}_{41} + \dots + \mathbf{Z}_{4J_4}\boldsymbol{\gamma}_{4J_4} \end{aligned} \quad (4.25)$$

em que, os β 's são os parâmetros de efeitos fixos, expressos por

$$\boldsymbol{\beta} = (\beta_1^T, \beta_2^T, \beta_3^T, \beta_4^T)^T$$

e definidos por Stasinopoulos et al. (2017), os parâmetros de efeitos aleatórios são independentes entre si e expresso por

$$\boldsymbol{\gamma} = (\gamma_{11}^T, \dots, \gamma_{1p_1}^T, \gamma_{21}^T, \dots, \gamma_{4p_4}^T),$$

no caso em que a distribuição possui até quatro parâmetros. Logo,

$$\gamma_{pj} \sim N(0, [\mathbf{G}_{pj}(\boldsymbol{\lambda}_{pj})]^{-1}),$$

em que, $[\mathbf{G}_{pj}(\boldsymbol{\lambda}_{pj})]^{-1}$ é a inversa (generalizada) de ordem $q_{jp} \times q_{jp}$ da matriz simétrica $\mathbf{G}_{pj}(\boldsymbol{\lambda}_{pj})$ que depende do vetor de hiperparâmetros $\boldsymbol{\lambda}_{pj}$. A matriz $\mathbf{G}_{pj}(\boldsymbol{\lambda}_{pj})$ é singular e γ_{pj} especifica uma função de densidade imprópria proporcional a $\exp\left(-\frac{1}{2}\boldsymbol{\gamma}_{pj}^T \mathbf{G}_{pj}(\boldsymbol{\lambda}_{pj})\boldsymbol{\gamma}_{pj}\right)$ (FLORENCIO, et al., 2012; STASINOPOULOS et al., 2017).

Desta forma, o modelo GAMLSS semi paramétrico com efeitos aleatórios é ajustado pelo método de estimação por máxima verossimilhança penalizada, em relação à estimação de $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ para um $\boldsymbol{\lambda}$ fixo. Representada pela expressão

$$L(\boldsymbol{\beta}_{\boldsymbol{\gamma}_p}) = L(\boldsymbol{\beta}) - \frac{1}{2} \sum_{p=1}^4 \sum_{j=1}^{J_p} \boldsymbol{\gamma}_{pj}^T \mathbf{G}_{pj}(\boldsymbol{\lambda}_{pj}) \boldsymbol{\gamma}_{pj}. \quad (4.26)$$

Algoritmos de maximização

Quando se trata de valores fixos de hiperparâmetros, a fim de maximizar a função de verossimilhança penalizada, recomenda-se dois algoritmos iterativos de estimação para o ajuste dos modelos GAMLSS (RIGBY & STASINOPOULOS, 2005). O primeiro, algoritmo **CG**, é uma generalização do algoritmo de *Cole & Green* (1992). Este algoritmo propõe o uso das primeiras derivadas e os valores esperados ou aproximados das derivadas de segunda ordem e derivadas cruzadas da função de verossimilhança no que se refere aos parâmetros da distribuição (por exemplo, $\boldsymbol{\beta} = (\mu, \sigma, \nu, \tau)$ para uma distribuição com quatro parâmetros). No entanto, para muitas funções de densidade de probabilidade, os parâmetros $\boldsymbol{\beta}$ são ortogonais, ou melhor, os valores esperados das derivadas cruzadas da função de log-verossimilhança são iguais a zero. Logo, neste caso, é utilizado um algoritmo mais simples e que não utiliza o valor esperado das derivadas cruzadas, conhecido como **RS**, que é uma generalização do algoritmo usado por *Rigby & Stasinopoulos* (2005) no ajuste da média e da dispersão de modelos aditivos.

A divergência fundamental entre os algoritmos é que; o algoritmo **RS** é maximizar a verossimilhança em cada parâmetro (μ, σ, ν, τ) até alcançar a convergência; já o algoritmo **CG**, a partir das derivadas cruzadas desde que não sejam nulas, irá atualizar os parâmetros em conjunto com cada interação (LEITE, 2019). No caso totalmente paramétrico, o algoritmo **CG** corresponde ao método score de *Fisher*, logo, os algoritmos maximizam a função de verossimilhança $L(\boldsymbol{\beta})$ e $L(\boldsymbol{\beta}_{\boldsymbol{\gamma}_p})$. Ressalta-se também, que em geral, o algoritmo **RS** é mais estável e mais rápido, quando comparado ao **CG** (STASINOPOULOS et al., 2017). Mais detalhes sobre os algoritmos **CG** e **RS** podem ser obtidos em *Rigby & Stasinopoulos* (2005).

4.5.2 Critério de seleção dos modelos

Uma importante medida para a seleção de modelos GAMLSS, é o critério de Akaike generalizado (GAIC), no mesmo contexto do critério de informação de Akaike (AIC) (AKAIKE, 1998). Definido por Voudouris et al. (2012), o GAIC, leva em consideração o número de parâmetros e de graus de liberdade utilizados no modelo para penalizar os modelos mais complexos e evitar sobreajustes aos dados em amostras de grandes tamanhos (PAIVA et.al, 2008). O GAIC é expresso por

$$GAIC(k) = -2L(\hat{\beta}) + (k \times gl),$$

em que L é o logaritmo da função verossimilhança e gl são os graus efetivos de liberdade do modelo ajustado, k é constante e é a penalidade para cada grau de liberdade utilizado. O $GAIC(k)$ penaliza modelos com muitos parâmetros, quanto menor o valor de $GAIC(k)$, melhor o modelo (STASINOPOULOS et al., 2017). O critério (AIC) (AKAIKE, 1998) e o critério de informação bayesiano (BIC) (SCHWARZ, 1978) são casos especiais do $GAIC(k)$, se $k = 2$ tem-se o (AIC) e se $k = \ln(n)$ recai no (BIC) (VOUDOURIS et al., 2012).

A seleção da distribuição apropriada pode ser realizada em dois estágios, que são: o estágio de ajuste e o estágio de diagnóstico. O estágio de ajuste envolve a comparação de diferentes modelos GAMLSS ajustados usando o (GAIC) por Voudouris et al. (2012). O estágio de diagnóstico é feito via análise gráfica dos resíduos e pelo gráfico de *worm plot*. A escolha da melhor função de ligação deve se realizada de forma a garantir que os parâmetros estejam sempre dentro de seus intervalos. A função de ligação é importante do ponto de vista de ajudar nas interpretações dos parâmetros. Diferentes funções de ligação podem ser comparadas usando o GAIC ou o gráfico *worm plot* (STASINOPOULOS et al., 2017). Os procedimentos usuais para seleção de variáveis via critério de informação são os mesmos de um modelo de regressão, sejam eles os algoritmos *forward*, *backward* e *stepwise*, que podem ser aplicados para cada parâmetro da distribuição. Já a seleção dos hiperparâmetros, um método padrão para definir os parâmetros de suavização, sugerido por Hastie e Tibshirani (1990), é fixá-los de acordo com os graus efetivos de liberdade para suavização.

4.5.3 Análise de resíduos

Na etapa de análise de resíduos podemos extrair informações que permitem diagnosticar a qualidade do ajuste de um modelo, e averiguar se as suposições do modelo foram satisfeitas. Para qualquer que seja a distribuição da variável resposta, os resíduos verdadeiros sempre terão uma distribuição normal padrão quando o modelo assumido

está correto. A averiguação de pressupostos por meio da normalidade dos resíduos já é bem estabelecido na literatura estatística. Dunn e Smyth (1996) propuseram os resíduos quantílicos (aleatorizados) normalizados, resíduo este que utilizaremos nos diagnósticos de modelos GAMLSS. Tais resíduos nos fornecem uma maneira familiar de se verificar a adequação de um modelo (STASINOPOULOS et al., 2017).

Segundo Dunn e Smyth (1996) o resíduo do quantil aleatório normalizado é definido por,

$$r_i = \Phi^{-1} \left\{ F \left(y_i; \hat{\beta} \right) \right\},$$

em que, Φ^{-1} representa a inversa da função de distribuição acumulada de uma normal padrão, $F(\cdot)$ é a função de distribuição acumulada adequada aos dados e $\hat{\beta}$ é o vetor de parâmetros. Logo, o modelo é dito adequado quando seus resíduos r_i seguem a distribuição Normal Padrão.

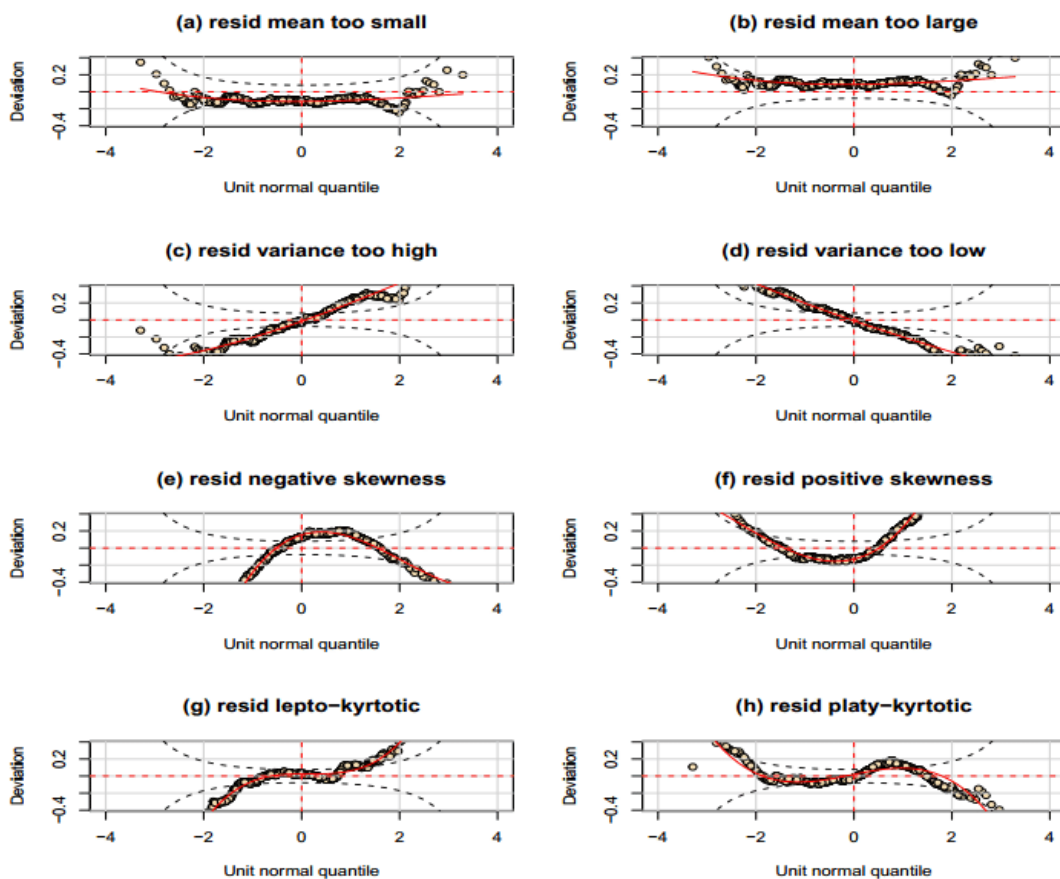
O gráfico *worm plot*, mais conhecido por gráfico de minhoca é uma das técnicas mais relevante aplicada para analisar os resíduos de um GAMLSS, pois ao aplicá-lo, permite-se detectar inadequações no modelo globalmente ou dentro de um intervalo de uma variável explicativa (BUUREN & FREDRIKS, 2001; BUUREN, 2007). O *worm plot* exerce uma função no mesmo sentido do gráfico normal quantil-quantil (Q-Q), sem tendência e com o intuito de evidenciar alterações locais no domínio de uma dada covariável.

Na utilização do *worm plot* é permitida a identificação de regiões em que o modelo não está bem ajustado aos dados, pois o eixo vertical descreve para cada observação a diferença entre a sua localização nas distribuições teórica e empírica. A forma gráfica de uma minhoca que os pontos formam em conjunto, aponta como os dados se distanciam da distribuição assumida e, quando tomadas em conjunto, sugere modificações úteis no modelo. Na tabela (2) pode ser observado as diferentes interpretações para os possíveis formatos dos pontos abrangendo desde variações na média (alta e baixa), variância (alta e baixa), assimetria (positiva e negativa) e curtose (caldas leves e pesadas). Estas interpretações auxiliam na tomada de decisão quanto ao ajuste do modelo aos dados, indicando se há necessidade de se escolher outro modelo probabilístico e/ou a inclusão de algum termo aos parâmetros de locação, escala e forma (STASINOPOULOS et al., 2017). Graficamente este comportamento pode ser visto na Figura (7).

Tabela 2 – Interpretação de diferentes formatos do worm plot (STASINOPOULOS et al., 2017).

Gráfico	Forma	Resíduos	Parâmetro ajustado
(a)	curva acima da origem	média muito alta	locação subestimada
(b)	curva abaixo da origem	média muito baixa	locação superestimada
(c)	curva inclinação positiva	variância muito alta	escala subestimada
(d)	curva inclinação negativa	variância muito baixa	escala superestimada
(e)	curva U	assimetria positiva	assimetria subestimada
(f)	curva U invertido	assimetria negativa	assimetria superestimada
(g)	S com curva esquerda pra baixo	leptocurtose	calda muito leve
(h)	S com curva esquerda pra cima	platycurtose	calda muito pesada

Figura 7 – Diferentes problemas que podem ser diagnosticados pelo *worm plot*.



Fonte: (STASINOPOULOS et al., 2017).

Os diagnósticos por meio de gráficos na Figura (7) apontam como os dados se distanciam da distribuição assumida e, a interpretação pode ser combinada com as informações da Tabela (2). O *worm plot* é uma ferramenta bastante útil para verificar adequabilidade e qualidade de ajuste.

4.6 Teste de homocedasticidade - *Breusch-Pagan*

Homocedasticidade é o termo para designar variância constante dos erros experimentais. Baseado no teste multiplicador de *Lagrange*, o teste de *Breusch-Pagan* é bastante utilizado para testar a hipótese nula de que as variâncias dos erros são iguais (homoscedasticidade) versus a hipótese alternativa de que existe pelo menos uma variância diferente. Então, a hipótese a ser testada é dada por

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 \\ H_1 : \text{pelo menos um dos } \sigma_i^2 \text{ é diferente, } i = 1, \dots, n. \end{cases}$$

O procedimento para realização do teste é dividido em cinco etapas

- Ao ajustar o modelo, calcula-se o resíduo $\hat{r}_1, \hat{r}_2, \dots, \hat{r}_n$;
- Obtêm-se o estimador de máxima verossimilhança de σ^2 , dado por $\hat{\sigma}^2 = \sum \hat{r}_i^2 / n$;
- Constrói-se uma nova variável v_i , definida por $v_i = \hat{r}_i^2 / \hat{\sigma}^2$;
- Realiza-se uma nova regressão para a variável v_i construídas sobre as variáveis explicativas;
- Calcula-se a soma de quadrados explicados pela regressão (SQReg) e define-se a estatística de teste por:

$$BP = \frac{SQReg}{2}.$$

Pressupondo que r_i se distribui normalmente, se há homocedasticidade e se o tamanho da amostra n aumenta indefinidamente, então, $BP \sim \chi_{n-1}^2$. 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 (OLIVEIRA, 2018).

4.7 Teste de normalidade - *Shapiro-Wilk*

Os testes de normalidade são utilizados para verificar se a distribuição de probabilidade associada a um conjunto de dados pode ser aproximada pela distribuição normal. Em alguns casos, assumir a normalidade dos resíduos é o primeiro passo que tomamos para simplificar nossas análises. Para dar suporte a esta suposição, consideramos, o teste de *Shapiro-Wilk* (1965), que pode ser utilizado para avaliar as hipóteses

$$\begin{cases} H_0 : \text{Os dados seguem uma distribuição normal} \\ H_1 : \text{Os dados não seguem uma distribuição normal.} \end{cases}$$

Dada uma amostra aleatória ordenada, $y_1 < y_2 < \dots < y_n$ a estatística de teste original *Shapiro-Wilk* é definido por

$$W = \frac{(\sum_{i=1}^n a_i y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (4.27)$$

em que, $a_i = \frac{m'V^{-1}}{m'V^{-1}V^{-1}m}^{1/2}$, $m = (m_1, \dots, m_n)'$ são os valores esperados das estatísticas das variáveis aleatórias distribuídas de forma independente e idênticamente distribuídas e V é a covariância matriz dessas estatísticas (RAZALI & WAH, 2011).

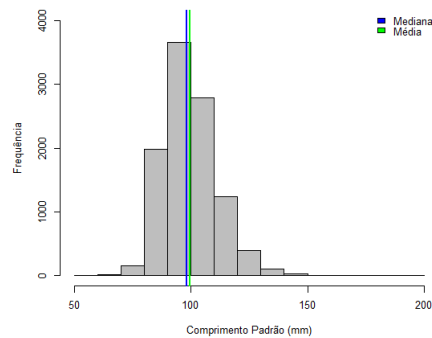
O valor de W está entre zero e um, onde pequenos valores de W levam à rejeição da normalidade, enquanto um valor de um indica normalidade dos dados. Se W é maior que o valor crítico, rejeitamos a hipótese de normalidade dos dados com $(1 - \alpha)100\%$ de confiança. Caso contrário, não rejeitamos a hipótese de normalidade (RAZALI & WAH, 2011).

5 Resultados

5.1 Análise exploratória

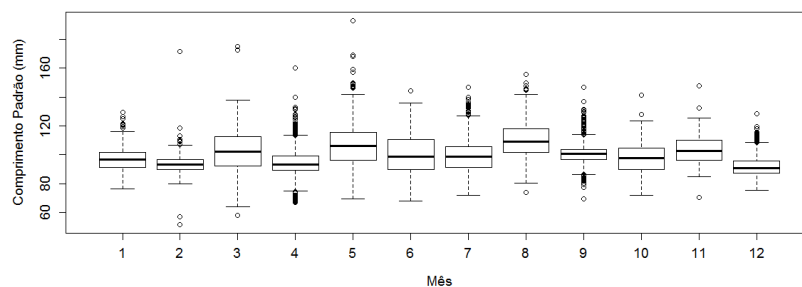
Comprimento padrão (cp) - (variável resposta) contínua por definição e a representação gráfica da distribuição de frequência foi realizada por meio do histograma (Figura 8). Os cp's variaram de 30,14 mm a 193 mm, com mediana de 97,24 mm e média 99,54 mm.

Figura 8 – Distribuição de frequência do comprimento padrão da *Opisthonema oglinum*.



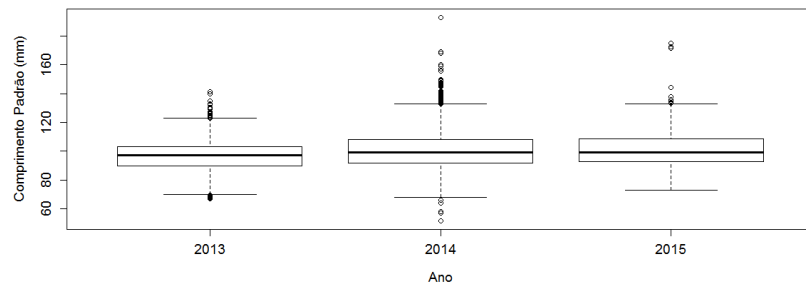
Mês - Não houve distinção muito relevante entre os comprimentos padrões observados e os níveis do fator Figura (9). Porém, os peixes menores foram capturados nos meses de fevereiro e março e houve menor variabilidade dos valores de comprimento padrão nos meses de fevereiro, abril, setembro e dezembro.

Figura 9 – Boxplots do comprimento padrão da *Opisthonema oglinum* segundo os meses de estudo.



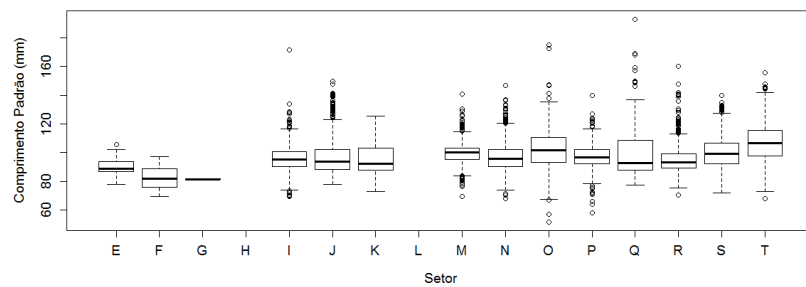
Ano - O fator ano foi inserido ao modelo por possibilitar a avaliação dos efeitos de longo prazo na variável resposta. Observou-se na Figura (10) que houve um pequeno aumento no comprimento padrão mediano de 2013 para 2014 e leve declínio de 2014 para 2015.

Figura 10 – Boxplots do comprimento padrão da *Opisthonema oglinum* segundo os anos de estudo.



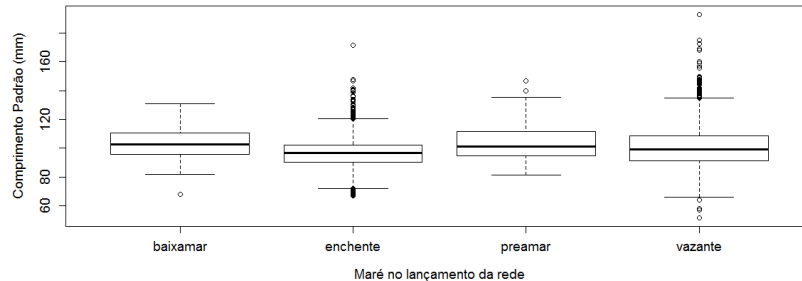
Setor - Notou-se na Figura (11) a falta de observação nos setores H e L, onde a área L está localizada na ponte Getúlio Vargas, logo não existiu possibilidade de coleta dos exemplares. Houve maior concentração de peixes menores nos setores E, F e G.

Figura 11 – Boxplots do comprimento padrão da *Opisthonema oglinum* segundo os setores do CSC.



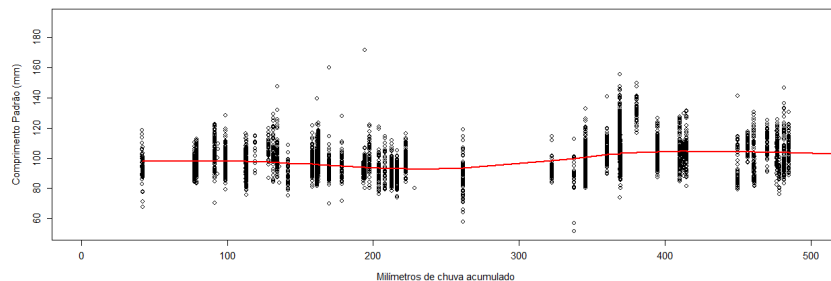
Maré de lançamento - Observou-se na Figura (12) que não houve distinção relevante entre os comprimentos padrões observados e a maré no momento do lançamento da rede.

Figura 12 – Boxplots do comprimento padrão da *Opisthonema oglinum* segundo a maré de lançamento.



Precipitação - Houve um comportamento aproximadamente constante na Figura (13), e, de acordo com a Agência Pernambucana de Águas e Clima (APAC) os maiores níveis pluviométricos na região estudada são entre os meses de abril a setembro.

Figura 13 – Diagrama de dispersão do comprimento padrão da *Opisthonema oglinum* segundo volume médio acumulado de chuva de trinta dias antes da coleta dos exemplares, nas estações Itapissuma, Igarassu e Itamaracá.



5.2 Modelo Linear Generalizado

Seleção do modelo

Para escolher a ordem em que as variáveis explicativas foram incluídas na estrutura do modelo, inicialmente, ajustou-se modelos considerando cada variável explicativa isoladamente e, então, calculou-se a *deviance*. Diante do ajuste que resultou na menor *deviance*, a variável explicativa então foi selecionada para o primeiro posto na estrutura do modelo. Repetiu-se o procedimento com a adição de cada uma das demais variáveis por vez, até que todas as variáveis fossem adicionadas ou até que a adição de uma nova variável não acarretasse no decréscimo da *deviance*. Para selecionar as variáveis explicativas e interações que deveriam ser mantidas nos modelos, utilizou-se a função `stepAIC()`

que leva em consideração o método de seleção *stepwise* juntamente com o Critério de Informação Akaike (AIC) (AKAIKE, 1974).

Diante do modelo estruturado, a função densidade que melhor adequou-se aos dados foi a Normal Inversa. Y_{ijklmn} é o comprimento padrão observado do i -ésimo peixe ($i = 1, 2, \dots, 10394$), capturado no j -ésimo mês ($j = 1, 2, \dots, 12$), no k -ésimo setor ($k = 1, 2, \dots, 14$), no l -ésimo ano ($l = 1, 2, 3$), na m -ésima maré ($m = 1, 2, 3, 4$), e com volume médio acumulado de chuva e o efeito da n -ésima interação do entre mês e ano. Então, admite-se que o comprimento padrão Y_{ijklmn} segue uma distribuição Normal Inversa de média μ_{ijklmn} e parâmetro de dispersão constante ϕ^{-1} , isto é, $Y_{ijklmn} \sim NI(\mu_{ijklmn}, \phi)$. A função de ligação considerada foi a identidade, cujo o preditor linear é expresso por

$$\eta_{ijklmn} = \mu_{ijklmn} = \alpha + \beta_j + \lambda_k + \theta_l + \delta_m + \varpi + \nu_n + \varepsilon_{ijklmn} \quad (5.1)$$

em que, $\beta_j, \lambda_k, \dots, \nu_n$ denotam, respectivamente, os efeitos do mês, setor, ano, maré, interação mês \times ano e ε_{ijklmn} o efeito aleatório atribuído a outros fatores não controláveis.

Análise de resíduo

De posse do modelo ajustado iniciou-se a análise gráfica de resíduos, não houve ponto de alavanca Figura 14 (a), nem indicação de ponto influente Figura 14 (b). No entanto, algumas observações apareceram em destaque no gráfico de pontos aberrantes Figura 14 (c) e os pontos não apresentaram tendência linear Figura 14 (d). Aplicou-se o teste de *Breusch-Pagan* ($p < 0,05$) e confirmou-se a heterocedasticidade dos resíduos no modelo ajustado. A eliminação de alguns pontos extremos individualmente e, em conjunto, altera algumas estimativas, mas, todos os efeitos continuam marginalmente significativos ao nível de 5%. Pôde-se observar por meio do gráfico semi-normal de probabilidades com envelope gerado (Figura 15) que os resíduos não possuem uma distribuição aproximadamente normal, houve violação na cauda a direita, indícios de que a distribuição normal inversa seja inadequada para explicar o comprimento padrão da *O. oglinum* no Canal de Santa Cruz.

A Figura 16 corrobora para os resultados encontrados nas Figuras (14 e 15), pois, por meio do *worm plot* percebeu-se que os dados não foram bem modelados pela distribuição Normal Inversa, muitos pontos ultrapassaram as bandas de confiança, evento que sugere falta de ajuste do modelo. Notou-se ainda, através da Figura 16 que os pontos formam um S com a curva esquerda para baixo, o que pode ser interpretado como leptocurtose.

Figura 14 – Análise gráfica dos resíduos para o modelo com distribuição Normal Inversa.

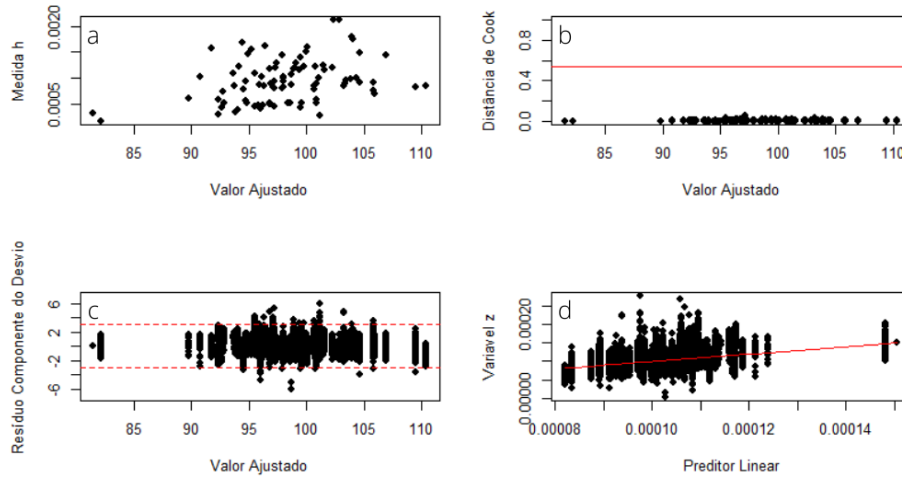


Figura 15 – Gráfico Semi-normal de probabilidade para o ajuste do modelo com distribuição Normal Inversa.

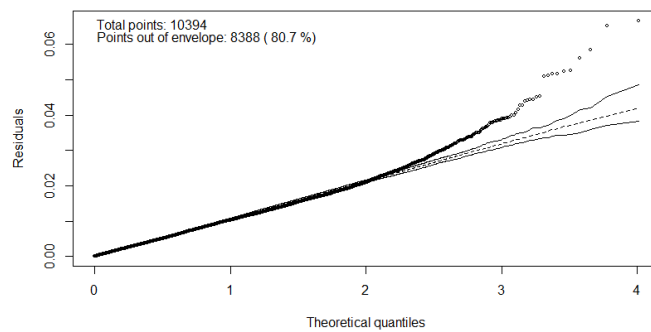
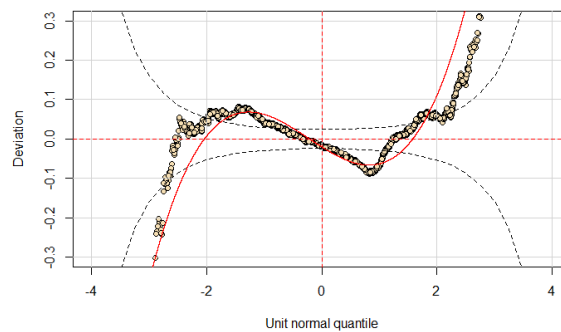


Figura 16 – Worm plot do ajuste do modelo com distribuição Normal Inversa.



Estes resultados indicam que os pressupostos da análise de variância não foram atendidos, o que sugere que outras abordagens devem ser utilizadas para obtenção de estimativas e predições confiáveis, para o levantamento de hipóteses sobre os processos ecológicos na captura da *Opisthonema oglinum*.

5.3 Modelos Aditivos Generalizados para Localização, Escala e Forma

Seleção do modelo

O processo de seleção do modelo GAMLSS foi realizado por meio de comparação entre diferentes distribuições atribuída a variável resposta, com a mesma estrutura utilizada em MLG, em que, diversos modelos concorrentes, com diferentes combinações de componentes foram utilizadas. Na Tabela (3), estão descritas as dez melhores possíveis candidatas a distribuição, de acordo com os valores GAIC em ordem crescente.

Tabela 3 – Principais distribuições candidatas a modelar os dados de comprimento padrão da *Opisthonema oglinum* sem configurações paramétricas.

Distribuição	GAMLSS (nome)	GAIC
Box-Cox t original	(BCTo)	78136.73
Box-Cox t	(BCT)	78143.58
Beta Generalizada tipo II	(GB2)	78148.14
Johnson SU	(JSU)	78171.34
Skew t tipo 3 repar.	(SST)	78177.84
Skew t tipo 3	(ST3)	78179.61
Box-Cox Exponencial potência original	(BCPEo)	78191.81
Box-Cox Exponencial potência	(BCPE)	78198.04
Exponencial Gaussian	(exGAUS)	78208.05
Skew t tipo 5	(ST5)	78218.68

Dentre os resultados apresentados na Tabela 3, chama-se atenção ao fato de que a distribuição Normal Inversa, considerada a melhor para o ajuste do modelo aos dados em (MLG), não está entre as candidatas a modelar os dados do comprimento padrão da *Opisthonema oglinum*, quando se leva em consideração os valores do GAIC ordenados do menor ao maior valor de ajustes realizados sem a inclusão dos parâmetros na configuração do modelo.

Diante disso, os principais modelos probabilísticos com respectivas configurações paramétricas, adotadas nas distribuições consideradas em função dos resultados encontrados na Tabela (3), são apresentados na Tabela (4). Foram listados os dez melhores modelos que apresentaram também desempenho superior, considerando análise diagnóstica

e convergência. O número de distribuições com combinações de parâmetros (μ, σ, τ, ν) ajustados aos dados do comprimento padrão da manjuba superou três dezenas.

Tabela 4 – Descrição dos principais modelos ajustados ao comprimento padrão da *Ophis-tonema oglinum*, com marcação dos parâmetros considerados

Modelo ajustado	Parâmetros				AIC
	μ	σ	ν	τ	
M1:Box-Cox t orig.	x	x	-	-	73142.71
M2:Box-Cox t	x	x	-	x	73231.36
M3:Beta generalizada tipo 2	x	x	-	x	73237.21
M4:Skew t type 3 repar.	x	x	-	-	73250.19
M5:Skew t type 3 repar.	x	x	x	x	73299.36
M6:Skew t type 3 repar.	x	-	-	x	73344.28
M7:Beta generalizada tipo 2	x	x	x	x	73371.34
M8:Johnson's SU repar.	x	x	x	-	73381.68
M9:Box-Cox t orig.	x	x	x	-	73394.42
M10:Box-Cox t	x	x	-	-	73399.73

x parâmetro considerado ao ajuste - parâmetro não considerado ao ajuste

Nota-se através da Tabela (4) que a distribuição *Box-Cox t orig.* com μ e σ na configuração paramétrica foi considerada a melhor candidata ao ajuste do modelo de acordo com o critério de GAIC. Então, a estrutura do modelo selecionado foi

$$\log(\hat{\mu}) = \text{Intercepto} + \text{mês} + \text{área} + \text{ano} + \text{maré} + \text{precipitação} + \text{mês} \times \text{ano}$$

$$\log(\hat{\sigma}) = \text{Intercepto} + \text{mês} + \text{área} + \text{ano} + \text{maré} + \text{precipitação} + \text{mês} \times \text{ano}$$

$$\hat{\nu} = \text{Intercepto}$$

$$\log(\hat{\tau}) = \text{Intercepto}$$

No modelo ajustado, as variáveis mês, área, ano, maré, precipitação e a interação (mês \times ano) foram relevantes na modelagem dos parâmetros μ (parâmetro de localização) e σ (parâmetro de escala). Entretanto, esse comportamento não se repete na modelagem do parâmetro ν e τ (parâmetros de forma). A adição das variáveis explicativas para a modelagem de tais parâmetros, provocou um aumento no valor do GAIC e ocasionou problemas de convergência ao modelo. Desta forma, ν e τ , configurados como constante, foi o mais adequado para a modelagem.

Análise de resíduo

Stasinopoulos e Rigby (2007) ao adequarem a função `plot()` para GAMLSS no *R*, substituíram os valores ajustados, pelos valores ajustados dos parâmetros de localização da distribuição de cada observação.

Figura 17 – Análise gráfica de resíduos do modelo ajustado com distribuição BCTo para o comprimento padrão da *Opisthonema Oglinum*.

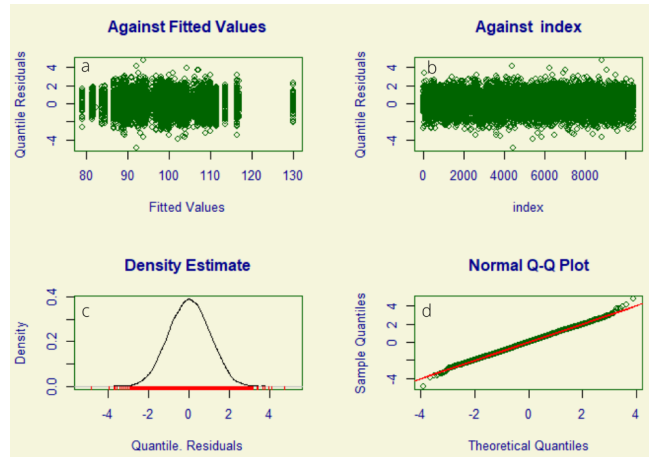
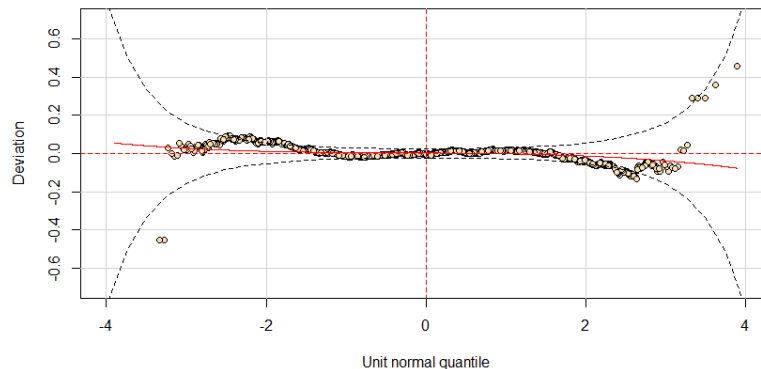


Figura 18 – Worm plot do modelo ajustado com distribuição BCTo para o comprimento padrão da *Opisthonema Oglinum*.



Por meio do gráfico de resíduos *versus* valores ajustados do parâmetro μ Figura (17 a) percebeu-se que os pontos distribuíram-se aleatoriamente em torno do zero. No gráfico de resíduos *versus* indexação da ordem das observações (17 b), observou-se uma nuvem de pontos completamente aleatória, resultado este, desejável, e que sugere homocedasticidade dos resíduos e pôde ser confirmada por meio do teste de *Breusch-Pagan* ($p < 0,05$). No gráfico da estimação da densidade *kernel* dos resíduos (17 c), notou-se um formato semelhante à um sino, o que nos leva a concluir que os resíduos são normalmente distribuídos, cuja hipótese foi confirmada por meio do teste de *Shapiro-Wilk* ($p < 0,05$). No gráfico normal Q-Q dos resíduos (17 d), poucos pontos ficaram fora da parte inferior e superior da distribuição e encontraram-se linearmente distribuídos. Na Figura (18) *worm plot*, alguns pontos próximos às caudas de confiança chamaram atenção, porém não corresponderam a 5% dos pontos que poderiam estar fora do intervalo com 95% de confiança, além de não

apresentarem qualquer padrão (linear, quadrático e/ou cúbico).

Foram obtidas também algumas estatísticas por meio do ajuste do modelo, média dos resíduos quantílicos aleatorizados próxima a zero (0,0011), variância próxima a 1 (0,9990), aproximando-se a uma distribuição normal padrão, assimetria próxima a zero (-0,0047) e curtose próxima a 3 (3,0015). Assim, o modelo final é apresentado nas equações abaixo

$$\mathbf{CP} \stackrel{ind}{\sim} BCTo(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau})$$

$$\boldsymbol{\eta}_1 = \log(\boldsymbol{\mu}) = \beta_{01} + \text{mês} + \text{área} + \text{ano} + \text{maré} + \text{precipitação} + \text{mês} \times \text{ano}$$

$$\boldsymbol{\eta}_2 = \log(\boldsymbol{\sigma}) = \beta_{02} + \text{mês} + \text{área} + \text{ano} + \text{maré} + \text{precipitação} + \text{mês} \times \text{ano}$$

$$\boldsymbol{\eta}_3 = \boldsymbol{\nu} = \beta_{03}$$

$$\boldsymbol{\eta}_4 = \log(\boldsymbol{\tau}) = \beta_{04}$$

Isolando os parâmetros μ , σ , ν e τ das equações acima, teremos as equações para substituir os coeficientes estimados da Tabela (5), em que, podemos quantificar os efeitos das covariáveis nos parâmetros, dadas por:

$$\mu = \epsilon^{\beta_{01}} \times \epsilon^{\text{mês}} \times \epsilon^{\text{área}} \times \epsilon^{\text{ano}} \times \epsilon^{\text{maré}} \times \epsilon^{\text{precipitação}} \times \epsilon^{\text{mês} \times \text{ano}}$$

$$\sigma = \epsilon^{\beta_{02}} \times \epsilon^{\text{mês}} \times \epsilon^{\text{área}} \times \epsilon^{\text{ano}} \times \epsilon^{\text{maré}} \times \epsilon^{\text{precipitação}} \times \epsilon^{\text{mês} \times \text{ano}}$$

$$\nu = \beta_{03}$$

$$\tau = \epsilon^{\beta_{04}}$$

Tabela 5 – Coeficientes estimados para μ por meio do ajuste do modelo com distribuição Box cox t original aos dados de comprimento padrão da *Ophistonema oglinum*.

Parâmetros - μ	Estimativa	Erro Padrão	t valor	valor p
Intercepto	4.733e+00	2.470e-02	191.607	< 2e-16
Fevereiro	-9.351e-02	7.308e-03	-12.795	< 2e-16
Março	4.582e-02	7.996e-03	5.730	1.03e-08
Abril	-2.933e-01	1.185e-02	-24.738	< 2e-16
Mai	-1.183e-01	1.481e-02	-7.989	1.51e-15
Junho	-3.016e-01	1.134e-02	-26.596	< 2e-16
Julho	-2.285e-01	1.303e-02	-17.540	< 2e-16
Setembro	-5.040e-02	7.990e-03	-6.307	2.96e-10
Outubro	-2.217e-01	1.375e-02	-16.122	< 2e-16
Novembro	1.286e-01	1.289e-02	9.973	< 2e-16
Dezembro	-1.113e-01	2.411e-02	-4.616	3.95e-06
Área F	-2.225e-01	3.090e-02	-7.200	6.42e-13
Área G	-1.296e-01	2.214e-02	-5.855	4.91e-09
Área I	7.311e-02	2.153e-02	3.396	0.000687
Área J	4.054e-02	2.097e-02	1.933	0.053252
Área K	9.806e-02	2.166e-02	4.527	6.04e-06
Área M	6.365e-02	2.127e-02	-2.992	0.002774
Área S	5.545e-02	2.188e-02	2.534	0.011293
2014	-2.628e-01	1.503e-02	-17.489	< 2e-16
2015	-1.559e-01	8.949e-03	-17.417	< 2e-16
Maré enchente	-1.366e-02	7.532e-03	-1.813	0.069869
Maré preamar	4.830e-02	9.930e-03	4.864	1.17e-06
Maré vazante	2.568e-02	7.507e-03	3.421	0.000626
Precipitação	7.024e-05	9.609e-06	7.310	2.88e-13
Fevereiro:2014	9.097e-02	1.494e-02	6.088	1.18e-09
Março:2014	-1.440e-01	1.918e-02	-7.509	6.45e-14
Abril:2014	3.522e-01	1.602e-02	21.984	< 2e-16
Mai:2014	2.952e-01	1.851e-02	15.946	< 2e-16
Junho:2014	4.292e-01	1.653e-02	25.957	< 2e-16
Julho:2014	4.012e-01	1.704e-02	23.543	< 2e-16
Agosto:2014	2.520e-01	1.379e-02	18.275	< 2e-16
Setembro:2014	2.225e-01	1.425e-02	15.610	< 2e-16
Outubro:2014	2.898e-01	1.827e-02	15.867	< 2e-16
Dezembro:2014	1.220e-01	2.622e-02	4.655	3.29e-06
Junho:2015	4.333e-01	1.265e-02	4.257	< 2e-16
Outubro:2015	2.836e-01	2.199e-02	12.897	< 2e-16

Na tabela (5 e 6) foram incluídos os efeitos principais mês, setor, ano, maré, precipitação e a interação mês \times ano, onde, contemplaram-se ao ajuste a modelagem dos parâmetros de posição e escala μ e σ , com ν e τ constantes. A variável resposta (CP) segue distribuição (BCTo) e as funções de ligação utilizadas para modelagem dos parâmetros foram logarítmica e identidade. No ajuste foram apresentadas apenas as estimativas

Tabela 6 – Coeficientes estimados para σ por meio do ajuste do modelo com distribuição Box cox t original aos dados de comprimento padrão da *Ophistonema oglinum*.

Parâmetros - σ	Estimativa	Erro Padrão	t valor	valor p
Intercepto	-3.169e+00	2.410e-01	-13.151	< 2e-16
Março	8.859e-01	6.997e-02	12.662	< 2e-16
Abril	1.373e+00	1.238e-01	11.091	< 2e-16
Mai	5.169e-01	1.430e-01	3.613	0.000304
Junho	-2.528e-01	1.274e-01	-1.985	0.047192
Julho	6.375e-01	1.360e-01	4.688	2.80e-06
Agosto	1.903e-01	8.474e-02	2.245	0.024762
Outubro	6.216e-01	1.353e-01	4.593	4.43e-06
Área G	-3.217e+00	5.740e-01	-5.604	2.15e-08
Área J	3.624e-01	1.986e-01	1.825	0.067993
Área N	4.103e-01	1.977e-01	2.076	0.037930
2014	4.430e-01	1.576e-01	2.810	0.004961
Maré preamar	1.970e-01	8.846e-02	2.227	0.025981
Maré vazante	1.765e-01	6.960e-02	2.535	0.011251
Março:2014	-6.558e-01	1.648e-01	-3.980	6.95e-05
Abril:2014	-1.484e+00	1.660e-01	-8.943	< 2e-16
Julho:2014	-7.122e-01	1.764e-01	-4.039	5.41e-05
Agosto:2014	-2.719e-01	1.373e-01	-1.980	0.047761
Outubro:2014	-6.639e-01	1.803e-01	-3.682	0.000232
Dezembro:2014	-8.040e-01	2.581e-01	-3.116	0.001841
Junho:2015	6.588e-01	1.367e-01	4.819	1.46e-06
Julho:2015	-8.551e-01	1.516e-01	-5.642	1.72e-08

estatisticamente significativas, com valor- $p < 0,05$. Assumimos a parametrização casela de referência, onde, esta restrição considera o primeiro nível de cada fator nulo, e então, são apresentados os efeitos dos fatores estimáveis. Assumimos a parametrização casela de referência, onde, esta restrição considera o primeiro nível de cada fator nulo, e então, são apresentados os efeitos dos fatores estimáveis.

Interpretação dos coeficientes estimados em relação a μ

- A maioria das estimativas dos níveis do fator mês foram de maneira isolada negativa, exceto os meses março e novembro. Os coeficientes das estimativas positivos em isolado indicam que o comprimento padrão médio da *O. oglinum* foi, no geral, maior nesses meses se comparado ao mês de janeiro.
- Os coeficiente estimados da área F e G foram negativos, afirmamos então que o comprimento padrão médio dos exemplares coletados próximo a desembocadura do rio Botafogo foram menores se comparados à área E. Já o sinal positivo nas demais áreas indicam que os peixes maiores encontraram-se na parte central e desembocadura sul do CSC.

- De forma isolada, o sinal dos coeficientes das estimativas dos níveis do fator ano foram negativos, apontou-se então, que o comprimento padrão médio da *O. oglinum* foi menor em 2014 e 2015 se comparado ao ano de 2013.
- Os coeficientes das estimativas de maré no momento do lançamento da rede quando em preamar e vazante indicaram que o comprimento padrão médio da *O. oglinum* foi maior nos referidos níveis que no valor de referência.
- O sinal positivo do coeficiente de precipitação indica que houve um aumento significativo no comprimento padrão médio da *O. oglinum* em épocas mais chuvosas.
- Nas estimativas da interação entre mês e ano observamos que os meses de fevereiro, abril, maio, junho, julho, setembro, outubro e dezembro, isoladamente, apresentaram contribuição negativa e particularmente no ano de 2014, as perspectivas negativas não se mantiveram. Da mesma forma, o mês de março dispôs de estimativa positiva e no ano de 2014 obtive estimativa negativa. Analogamente as interações anteriores, os meses de junho e outubro, e o ano de 2015 mostravam coeficientes negativos em separado, entretanto, na interação entre esses níveis a perspectiva não confirmou-se.

Interpretação dos coeficientes estimados em relação a σ

- O sinal positivo dos coeficientes estimados de maneira isolada dos meses de março, abril, maio, julho agosto e outubro indicaram que a dispersão do comprimento padrão da *Opisthonema* no CSC foi maior do outono para o inverno. Já no mês em que o sinal foi negativo, existe menor dispersão do comprimento padrão da espécie.
- O coeficiente estimado da área G foi negativo, então, pôde-se afirmar que a variabilidade no comprimento padrão da manjuba foi menor que na área E, já o sinal positivo do coeficiente estimado das áreas J e N, pôde-se dizer que a variabilidade foi maior que na área E.
- Notou-se por meio do sinal positivo do coeficiente estimado do ano 2014 de forma isolada da interação que a variabilidade no comprimento padrão da espécie no ano de 2013 foi menor que no referido ano.
- Com relação ao coeficiente estimado positivo da maré em preamar e vazante no momento do lançamento da rede mostrou que a dispersão no comprimento padrão da *Opisthonema oglinum* quando capturada na baixamar foi inferior ao preamar e vazante.

- Já nas estimativas dos coeficientes da interação entre mês e ano observamos que os meses de março, abril, julho e outubro, isoladamente, apresentaram contribuição positiva e particularmente em 2014, as perspectivas positivas não se mantiveram, então afirmou-se que a variabilidade no comprimento padrão da manjuba foi menor nos meses citados do ano de 2014. Analogamente as interações anteriores, os meses de junho e julho mostraram coeficientes negativo e positivo em separado respectivamente, entretanto, na interação com o ano 2015 as perspectivas não confirmaram-se.

6 Discussão

Houve predominância de indivíduos maiores nos meses de Abril a Novembro de 2014, fato que deve estar relacionado ao regime pluviométrico e ao ciclo de vida. A estação chuvosa na região do entorno do CSC foi registrada entre os meses de Abril a Setembro. Por tratar-se de um período de elevadas descargas fluviais, ocasiona concentrações mais elevadas de biomassa primária (PAIVA et al. 2009; SIMONI, 2019; LIMA et al. 2020), o que torna o CSC um local favorável para algumas espécies filtradoras que tem parte importante da dieta sobre fito e zooplâncton. Em adição o ciclo de vida deve estar sincronizado com esse regime de variação ambiental. Onde as formas jovens disponham desse momento de maior disponibilidade de alimento para crescer usufruindo da biomassa primária maior nesse período chuvoso.

De acordo com os resultados observados, notamos que o comprimento padrão da *O. oglinum* no CSC foi em média, menor entre as áreas E e G, próximo a desembocadura do rio Botafogo. Este fato pode está relacionado à menor largura do canal no local e à maior área marginal com cobertura de manguezal. A vasta área de canais secundários e rasos propicia uma ampla região de abrigo, favorecendo o uso da área como berçário inicial, abrigando grandes quantidades das formas mais jovens e de menor tamanho. A hipótese é de que à medida que se desenvolvem e crescem a preferência começa a ser maior por áreas de canais largos, até mais expostos a predadores, mas que propicia alimento mais adequando para as formas já um pouco maiores. Assim o comprimento padrão médio da manjuba foi maior na área central do CSC, onde o substrato é areno-lamoso e lamo-arenoso especificamente nesta região (SILVA, 2008; PAIVA et al. 2009), o que possibilita o crescimento de constituintes da meiofauna e macrofauna, com organismos como copépodes, nematóides, poliquetas e crustáceos (SANTOS, 2009; PAIVA et al. 2009). Desta forma, podemos supor que a *O. oglinum* faz uso da porção central do canal por esta área disponibilizar diversos recursos alimentares, já em uma fase mais avançada do desenvolvimento. Importante notar também que os comprimentos maiores tendem a estar mais concentradas além da porção central, já próximo da desembocadura sul do estuário. Assim a hipótese ecológica quanto ao uso do espaço do canal ao longo do tempo desde as fases iniciais são: a) formas bastante jovens, exemplares pequenos, mais concentrados em áreas com maior margem de manguezal e canais mais estreitos. b) exemplares um pouco maiores já mais concentrados na porção central do canal; c) formas ainda maiores já concentrados perto da desembocadura do canal, já com mais fácil acesso à parte externa e região marinha adjacente.

A *Opisthonema oglinum* é classificada como de origem marinha visitante, passa boa parte do seu ciclo de vida no mar, e, em período regular penetra no estuário (VASCONCELOS-FILHO et al. 2009). Mas encontramos indícios neste trabalho de que poucos adultos ingressam no canal e a grande maioria dos exemplares capturados no CSC são jovens, que é portanto a fase que mais faz uso da área estuarina. Mais do que isso, à medida que se desenvolvem, ganham em idade e tamanho, os exemplares se distribuem cada vez mais perto da desembocadura sul do canal. O que indica claramente a região interna como importante área de berçário e crescimento inicial no ciclo de vida da sardinha. Entendimento esse que tem suporte em resultados obtidos em outras áreas, incluindo regiões de clima menos equatorial/tropical. Frehse (2006), por exemplo, observou que no estuário de Paranaguá, no Paraná, a espécie foi encontrada quase que exclusivamente na fase juvenil, já que todos os indivíduos analisados mostraram-se sexualmente imaturos. De acordo com Simoni (2019), na costa do litoral norte do estado de Pernambuco, no ambiente marinho, cerca de 90% dos exemplares capturados tinham comprimento total acima de $L_{pm} = 114mm$, que seria o tamanho que já se observa avanço da maturidade. No entanto, Lino (2003) encontrou que 92,7% dos exemplares capturados na região de Itapissuma (área interna do estuário) estavam abaixo do que o autor coloca como comprimento médio de primeira maturação ($L_{pm} = 117mm$). Ao levar em consideração os resultados que encontramos e os extraídos da literatura e mencionados acima, há fortalecimento da hipótese de que as formas jovens ingressam na área interna do estuário do CSC, que funciona como berçário, e a medida que se desenvolvem avançam para as proximidades da desembocadura, enquanto que as frações de pré adultos e adultos do estoque são encontradas na área marinha adjacente. Dada a dimensão do CSC, esta região de berçário é possivelmente a mais importante para o estoque que é acessível no litoral de Pernambuco, e talvez seja também o berçário mais importante da margem oriental do nordeste.

Por ser uma espécie marinha visitante, a manjuba na fase adulta ocuparia preponderantemente a região marinha adjacente à costa, mas entraria no estuário em períodos regulares. No caso do estuário de Paranaguá, no Paraná, Frehse (2006) afirma que o ingresso na região interna se daria logo após o período de desova. Lino (2003) sugere que no caso do CSC, em Pernambuco, as formas adultas se afastam do entorno do canal para desova nos meses de março, e especialmente de maio a outubro, o que indicaria haver um momento de maior intensidade de desova. Porém Simoni (2019) obteve evidências de que a sardinha laje apresenta desova parcelada durante todo o ano nesta região tropical no litoral norte de Pernambuco, incluindo eventos tanto em períodos de chuva e seca, mas com picos pontuais de atividade reprodutiva em meses de seca. Os resultados que obtivemos, indicam tendência geral de menores tamanhos de Dezembro a Fevereiro, e também comprimentos reduzidos em Abril. Esse período corresponde à estiagem. É uma evidência de que os

exemplares de maior tamanho, já mais próximos da reprodução, teriam uma tendência a estarem proporcionalmente menos concentrados na área interna, e estariam portanto fora do canal. Isto é indício a favor da hipótese de que ainda que parcelada ao longo de todo o ano, haja maior concentração de atividade desova na região marinha adjacente na época de seca.

Outro aspecto a ser levado em consideração é a variação do comprimento ao longo dos anos. LIMA (2015) indica que há indícios de tendência de que os comprimentos da sardinha capturada no CSC, estariam diminuindo ao longo dos anos. Se compararmos os comprimentos mínimo (30,14 mm) e máximo (193,00 mm) observados neste com os obtidos no passado por Lino (2003) – mínimo de 52,50 mm e máximo de 209,00 mm – haveria nova evidência favorável a hipótese de tendência de decréscimo. A comparação tem que ser vista com cautela por não ser possível comparar as seletividades dos aparelhos de pesca utilizados nos diferentes trabalhos. Porém, ainda assim, já é um alerta visto que tendências de diminuições nos comprimentos capturados estão classicamente associadas a aumento do coeficiente de mortalidade por pesca, e conseqüentemente do coeficiente de mortalidade total (BEVERTON & HOLT, 1957). Cenários que podem conduzir à sobrepesca de crescimento (PALAZZO & CARBOGIM, 2012). Verificar se isto ocorreu e ainda ocorre no caso da pesca da sardinha laje na margem oriental do Nordeste é de suma importância no futuro, e deve gerar informações valiosas para tomadas de decisão e manejo do estoque explorado. Esse trabalho é no entanto desafiador, pois deve-se levar e conta o mosaico espacial e temporal marcado pelo ciclo de vida complexo, em que diferentes parcelas da população (jovens e adultos) fazem uso de diferentes partes do ambiente (estuário interno, margem externa do estuário, e região marinha). Portanto, não somente os níveis de mortalidade natural devem ser diferentes para as diversas frações da população, mas também, há que se incorporar na análise níveis de mortalidade por pesca diferentes.

Em virtude disso, seria adequado o desenvolvimento de investigações mais aprofundadas para o levantamento de novas hipóteses e, até a implementação de algumas medidas de gestão caso necessária, como a limitação da captura a nível sustentável (FRID et al., 2003), o estabelecimento de tamanhos mínimos de captura (STERGIOU et al., 2009), a mudança do tamanho de malha (FERRO et al. 2008; LIMA, 2020) em períodos e locais específicos do Canal de Santa Cruz, para manter o recurso em níveis razoáveis para a sobrevivência da atividade pesqueira no Canal de Santa Cruz.

7 Conclusão

Ao longo do desenvolvimento desta pesquisa foram apresentadas as características e propriedades da classe de modelos de regressão, mais precisamente os modelos lineares generalizados (GLM) e os modelos aditivos generalizados para localização, escala e forma (GAMLSS). As análises realizadas revelaram que os modelos estimados via GAMLSS viabilizaram um ajuste superior aos obtidos via GLM, segundo o critério GAIC e as análises dos resíduos, apontando que a classe dos modelos GAMLSS aparenta ser mais apropriada que as tradicionais modelagens via GLM para descrever o comprimento padrão da *Opisthonema oglinum*. Uma vez que a natureza dos dados analisados tenha sugerido a distribuição Box Cox t Original para ajuste da variável resposta e que tenhamos modelado apenas os parâmetros de posição e escala, os modelos GAMLSS possibilitam o ajuste de uma ampla família de distribuições que podem fornecer informações adicionais sobre a assimetria e a curtose, o que não é permitido na modelagem via GLM.

Assim sendo, através do emprego do modelo GAMLSS ajustado com a distribuição BCTo, foi possível verificar que existiu sazonalidade no comprimento padrão da *O. oglinum* capturada no Canal de Santa Cruz com predominância de indivíduos maiores no período chuvoso, e com maior concentração na região centro-sul do canal. Embora ocorra a captura de indivíduos um pouco maiores, há, ao mesmo tempo, uma frequência atuante de *O. oglinum* muito jovens na porção norte do canal, o ideal seria que a pesca explorasse exemplares que já completaram pelo menos a maturidade sexual, pois a captura de indivíduos tão jovens, pode levar a um desequilíbrio na sustentabilidade da espécie, mesmo a *O. oglinum* sendo uma espécie com altos níveis reprodutivos merece cautela. Logo, é fundamental que existam investigações mais aprofundadas para o levante de novas hipóteses e até mesmo para que sejam implementadas medidas de gestão caso necessário, para evitar possível sobrepesca da *Opisthonema oglinum*.

Referências Bibliográficas

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

AKAIKE, H. Information theory and an extension of the maximum likelihood principle. Selected Papers of Hirotugu Akaike. **Springer**, p. 199-213, 1998.

ALBERT, O.T.; BERGSTAD, O.A. Temporal and spatial variation in the species composition of trawl samples from a demersal fish community. **Journal of Fish Biology**, v. 43, p. 209-222, 1993.

ANDRADE, H. A.; SILVA, R.M.M. Dinâmica das frotas de pesca de emalhe e linha de mão de Itapissuma-PE no Canal de Santa Cruz. **Boletim Técnico Científico CEPENE**, Tamandaré, v. 19, p. 83-91, 2013.

ATKINSON, A. C. Transformations and Regression. **Oxford University Press**, Oxford. 1985.

BEVERTON, R. J. H.; HOLT, S. J. On the dynamics of exploited fish populations: **Fishery Investigations**, London, Series II, XIX, p. 533, 1957.

BLABER, S.S.J.; D.T. BREWER; J.P. SALINI & J. KERR. Biomass, catch rates and abundances of demersal fishes, particularly predators of prawns, in a tropical bay in the Gulf of Carpentaria, Australia. **Marine Biology**, v. 107, p. 397-408, 1990.

BLABER, S.M.J.; T.G. BLABER. Factors affecting the distribution of juvenile estuarine and inshore fish. **Journal of Fish Biology**, v. 17, p. 143-162, 1980.

BLABER, S.M.J.; D.T. BREWER & J.P. SALINI. Species composition and biomass of fishes in different habitats of a tropical norther Australia estuary: their occurrence in the adjoining sea and estuarine dependence. **Estuarine and Coastal Shelf Science**, v. 29, p. 509 - 531, 1989.

BOX, G. E. P.; COX, D. R. An Analysis of Transformations. **Journal of the Royal Statistical Society**. Series B (Methodological), v. 26, p. 211 - 252, 1964.

BUUREN, S.V. Worm Plot to Diagnose Fit in Quantile Regression. **Statistical Modelling**. SAGE Publications, v. 7, p. 363-376, 2007.

BUUREN, S. V.; FREDRIKS, M. Worm plot: a simple diagnostic device for

modelling growth reference curves. **Statistics in medicine**, Wiley Online Library, v. 20, p. 1259 - 1277, 2001.

CARPENTER, K. E. The living marine resources of the Western Central Atlantic. Volume 2: Bony fishes part 1 (Acipenseridae to Grammatidae). **FAO Species Identification Guide for Fishery Purposes and American Society of Ichthyologists and Herpetologists Special Publication**. Rome, v.5, p. 601 - 1374, 2002.

CHAVES, P.T.C.; G. OTTO. The mangrove as a temporary habitat for fish: the *Eucinostomus* species at Guaratuba Bay, Brazil (25°52'S; 48°39'W). **Brazilian Archives of Biology and Technology**, v 42, p. 61 - 68, 1999.

COLE, T. J.; GREEN, P. J. Smoothing reference centile curves: the lms method and penalized likelihood. **Statistics in medicine**, Wiley Online Library, v. 11, p. 1305 - 1319, 1992.

CONOVER, W.J. Practical Nonparametric Statistic. Third Edition, **John Wiley & Sons**, Inc. New York, p. 428 - 433, 1999.

CORDEIRO, G. M.; DEMÉTRIO, C. G. B.; **Modelos Lineares Generalizados e Extensões**. Piracicaba: ESALQ/USP, 2013.

COUTO, L.M.M.R; VASCONCELOS-FILHO, A.L. Sobre a biologia da Sardinha-Bandeira *Opisthonema oglinum* (LESUEUR, 1818), no Canal de Santa Cruz, PE. **Ser. Ci. Aquát.**, Recife, 1986.

CPRH (Companhia Pernambucana de Controle da Poluição Ambiental e de Administração de Recursos Hídricos) 2010. **Diagnóstico Sócioambiental da Área de Proteção Ambiental de Santa Cruz**. Disponível em: <http://www.cprh.pe.gov.br/ARQUIVOS/ANEXO/diagnosticosantacruz;2243;20120906.pdf>. Acessado em: 14/02/2020.

DOBSON, A.J. **An Introduction to Generalized Linear Models**. 2 Edição. Chapman & Hall, Boca Raton, p. 225, 2002.

DSPE, Diagnóstico socioeconômico da pesca artesanal do litoral de Pernambuco. **Instituto Oceanário de Pernambuco**: Departamento de Pesca e Aquicultura da UFRPE, Recife, PE. 2010.

DUNN, P. K.; SMYTH, G. K. Randomized Quantile Residuals. **Journal of Computational and Graphical Statistics**, JSTOR, v. 5, p. 236, 1996.

EILERS, P. H.; MARX, B. D.; DURBÁN, M. Twenty years of p-splines. **SORT: statistics and operations research transactions**, v. 39, p. 149-186, 2015.

FELTRIM, M.C.; SCHWINGEL, P.R. *Opisthonema oglinum* (Lesueur, 1818). In:

Cergole, M.C., Ávila-da-Silva, A. O. & Rossi-Wongtschowky, C. L. B. (eds) Análise das principais pescarias comerciais da região Sudeste-Sul do Brasil: dinâmica populacional das espécies em exploração. Série Doc. **Revizee/Score Sul**, p. 112 - 115, 2005.

FERRO, R. S. T.; OZBILGIN, H. ; BREN, M. The potential for optimizing yield from a haddock trawl fishery using seasonal changes in selectivity, population structure and fish condition. **Fisheries Research**, v. 94, p. 15 - 159, 2008.

FIGUEIREDO, J. L.; MENEZES, N. A. Manual de peixes marinhos do sudeste do Brasil. II. **Teleostei**. Universidade de São Paulo, São Paulo. p. 110, 1978.

FISHER, R.A. Two new properties of mathematical likelihood. **Philosophical Transactions of the Royal Society A**, Nottingham, v. 144, p. 285-307, 1934.

FLORENCIO, L.A. **Engenharia de avaliações com base em modelos GAMLSS**. 2010 Dissertação de (mestrado) - Universidade Federal de Pernambuco. Departamento de Estatística, Recife, 2010.

FLORENCIO. L.A., CRIBARI-NETO. F., AND MARTÍNEZ. R.O., Real estate appraisal of land lots using GAMLSS models Chilean **Journal of Statistics**, v. 3, 2010.

FREHSE, F.A. **Estrutura da população de três espécie da família Clupeidae: Harengull clupeola, Opisthonema oglinum e Sardinella brasiliensis, no Complexo Estuarino de Paranaguá, Paraná**. Monografia (Graduação) - Universidade Federal do Paraná - Curso de Ciências Biológicas, Curitiba, 2006.

FONTELES FILHO, A. A. **Oceanografia, biologia e dinâmica populacional de recursos pesqueiros**. Expressão Gráfica e Editora. Fortaleza, CE. p. 464, 2011.

FRID, C.; HAMMER, C.; LAW, R.; LOENG, H.; PAWLAK, J.F.; REID, P.C.; TASKER, M. Environmental Status of the European Seas. **ICES**. 2003.

GARCIA, A.M.; VIEIRA, J.P. Abundância e diversidade da assembléia de peixes dentro e fora de uma pradaria de *Ruppia marítima* L., no estuário da lagoa dos Patos (RS - Brasil). **Atlântica**, v. 19, p. 161 - 181, 1997.

GARCÍA-ABAD, M.C.; YÁNEZ-ARANCIBIA, A.; SANCHEZ-GI P.; TAPIA-GARCÍA, M. Distribución, abundancia y reproducción de *Opisthonema oglinum* (Pisces: Clupeidae) en la plataforma continental del sur del Golfo de México. *Rev. Biol. Trop.*, v. 46(2), p. 257-266, 1998.

GRIFFITHS, J. F. Applied climatology. New York: **Oxford University Press**, p. 118, 1996.

GONZÁLEZ-CABELLOS, L. W.; MENQUAL-IZQUIERDO, A. Edad y crecimiento

del machuelo, *Opisthonema oglinum* (Le Seuer, 1818) (Teleostei - Clupeidae), de la Isla de Margarita, Venezuela. Instituto de Investigaciones Científicas Universidad de Oriente, Núcleo de Nueva Esparta. v. 2, p. 387 - 399, 1995.

HAMLEY, J.M., Reviews of gillnet selectivity. **Journal of the Fisheries Research Board of Canada**, v.32, p. 1943 - 1969, 1975.

HASTIE, T.; TIBSHIRANI, R. **Generalized additive models**. [S.l.]: Wiley Online Library, 1990.

IBAMA/CEPENE. Boletim estatístico da pesca marítima estuarina do nordeste do Brasil. Tamandaré: IBAMA/CEPENE. p. 385, 2008.

IBGE, INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Área Territorial Brasileira** - DOU 2017. Disponível em: <https://ww2.ibge.gov.br/home/geociencias/cartografia/default-territ-area.shtm>>. Acesso em 02 Agosto 2017.

KHURI, A. I. **Linear Model Methodology**. Taylor and Francis Group, LLC. London, New York, 2010.

KINNEAR, B. S.; FUSS, C. M. Sthread herring distribution off Floridas West Coast. **COMMERICAL FISHERIES REVIEW**. v. 33, p. 27, 1971.

LEE, Y.; NELDER, J. A. Hierarchical generalized linear models. **Journal of the Royal Statistical Society**, p. 619 - 678, 1996.

LEITE, R.S. Aplicação de modelos aditivos generalizados para locação, escala e forma em clones de *Eucalyptus* spp., no polo gesseiro do Araripe- PE - Dissertação (Mestrado) – Universidade Federal Rural de Pernambuco, Programa de Pós-Graduação em Biometria e Estatística Aplicada, Recife, BR-PE, 2019.

LEKVE, K.; N.C. STENSETH; J.G. JFSÆTER; J.M. FROMENTIN & J. GRAY. Spatiotemporal patterns in diversity of a fish assemblage along the Norwegian Kagerrak coast. **Marine Ecology Progress Series** v. 178, p. 17 - 27, 1999.

LESSA, R. P.; MONTEIRO, A.; DUARTE-NETO, P. J. e VIEIRA, A. C. **Análise Multidimensional dos Sistemas de Produção Pesqueira do Estado de Pernambuco, Brasil**. In: Haimovici, M. (Ed). *Sistemas Pesqueiros Marinhos e Estuarinos do Brasil*. Rio grande: Editora da FURG, p. 41 - 54. 2011.

LESSA, R. P.; NÓBREGA, M. F.; JUNIOR, J. L. B. **Dinâmica de populações e avaliação de estoques dos recursos pesqueiros da região nordeste. Programa de Avaliação do Potencial Sustentável dos Recursos Vivos da Zona Econômica Exclusiva SCORE-NE**. DIMAR, Departamento de Pesca-Universidade Federal Rural de Pernambuco, Recife-Brazil, v.2, 2004.

LESUEUR, C. A. Description of several new species of North American fishes. *J. Acad. Nat. Sci. Philad.* n.1, v.2 , p. 222 - 235, 1818.

LIMA, S.A.O. Sazonalidade e seletividade nas pescarias de emalhe de pequenos peixes pelágicos no Canal de Santa Cruz - PE. Dissertação de (mestrado). Pós-graduação em Recursos Pesqueiros e Aquicultura. DEPAq-UFRPE. 2015.

LIMA, S. A. O., ANDRADE, H. A., SOUSA, R. G. C. Dinâmica da pesca de peixes pelágicos de pequeno porte no estuário do canal de Santa Cruz (Pernambuco – Brasil) - **Brazilian Journal of Development** - Curitiba, v. 6, n. 9, p. 73423 - 73435, 2020.

LIMA, T; QUINAMO, T. **Características sócio-econômicas. Gerenciamento Participativo de estuários e manguezais**, Recife Ed. Universitária UFPE, p. 181-224, 2000.

LINO, M. A. S. Estudo biológico-pesqueiro da manjuba *Opisthonema oglinum* (Lesueur, 1818) na região de Itapissuma, Pernambuco. Dissertação de (mestrado). Pós-Graduação em Recursos Pesqueiros e Aquicultura. DEPAq-UFRPE, p. 46, 2013.

LIU, W.; CELA, J. Improving Credit Scoring by Generalized Additive Model. **SAS Institute Inc**, 2007.

MACÊDO, S. J.; FLORES MONTES, M. J. ; KOENING, M. L. ; LINS, I. C. Phytoplankton Nyctemeral Variation at the Santa Cruz Channel, Itamaracá- Pernambuco, Brazil. **Anais Acad. Bras.** Rio de Janeiro, p. 857 - 871, 1998.

MARSHALL, S.; ELLIOTT, M. Environmental influences on the fish assemblage of the Humber Estuary, U.K. **Estuarine & Coastal Shelf Science**, v. 46, p. 175 - 184, 1998.

MCAULEY, R.B.; SIMPFENDORFER, C.A.; WRIGHT, I.W. Gillnet mesh selectivity of the sandbar shark (*Carcharhinus plumbeus*): implications for fisheries management. **ICES Journal of Marine Science**, v. 64, p. 1702 - 1709, 2007.

MCCULLAGH, P.; NELDER, J. A. **Generalized linear models**. Chapman and Hall, London, 1989.

MEXICANO-CÍNTORA, G.; SALAS, S.; CABRERA, M. A. **ANÁLISIS DE LA PESQUERÍA Y DINÁMICA POBLACIONAL DE LA VIVITA DE HEBRA (*Opisthonema oglinum*) EN YUCATÁN, MÉXICO**. Centro de Investigación y de Estudios Avanzados del IPN, Unidad Mérida Universidad Autónoma de Baja California, 1996.

MIRANDA, L.B.; CASTRO, B.M.; KJERFVE, B. **Princípios de Oceanografia Física de Estuários**. São Paulo: Edusp - Editora da Universidade de São Paulo, São

Paulo, p. 411, 2002.

MMA, GERÊNCIA DE BIODIVERSIDADE AQUÁTICA E RECURSOS PESQUEIROS. Panorama da conservação dos ecossistemas costeiros e marinhos no Brasil. Brasília. 2020. <https://www.mma.gov.br/biodiversidade/biodiversidade-aquatica/zona-costeira-e-marinha>. Acesso em 02 Agosto 2017.

MOURA, R.T. Aspectos Gerais da Hidrobiologia do Litoral Norte de Pernambuco - Brasil. Brasília, DF: **Ibama**. p. 138, 2009.

MPA. Ministério da Pesca e Aquicultura. **Boletim Estatístico da Pesca e Aquicultura**. Brasil, Brasília, 2012

MULLIN S.J. Estuarine fish populations among red mangrove prop roots of small overwash islands. **Wetlands** **15**: , p. 324 – 329, 1995.

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

NÓBREGA, M.F.; LESSA, R.; SANTANA, F.M. Peixes marinhos da região Nordeste do Brasil. **Programa Revizee – Score Nordeste**. Editora Martins e Cordeiro. Fortaleza, 208p. 2009.

OLIVEIRA, E.S.C. 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 (Doutorado) – Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, Recife, 2018.

PAIVA, A.C.G. Ecologia de peixes estuarinos-recifais e caracterização ambiental dos estuários de Pernambuco, (Tese de doutorado) - Universidade Federal de Pernambuco. Departamento de Oceanografia. 2009.

PAIVA, A.C.G.; CHAVES, P.T.C.; ARAÚJO, M.E. Estrutura e organização trófica da ictiofauna de águas rasas em um estuário tropical. **Revista Brasileira de Zoologia**, v. 25, p. 647 - 661, 2008.

PAIVA, A.C.G.; COELHO, P.A.; TORRES, M.F.A. INFLUÊNCIA DOS FATORES ABIÓTICOS SOBRE A MACROFAUNA DE SUBSTRATOS INCONSOLIDADOS DA ZONA ENTRE-MARÉS NO CANAL DE SANTA CRUZ, PERNAMBUCO, BRASIL. **Arq. Ciên. Mar**, Fortaleza, v. 38, p. 85 - 92, 2005.

PAIVA, C. S. M; FREIRE, D. M. C; CECATTI, J. G. Modelos Aditivos Generalizados para Posição, Escala e Forma (GAMLSS) na Modelagem de Curvas de Referência. **Revista Brasileira de Ciências da Saúde**, v. 12, n. 3, p. 302, 2008.

PALAZZO, J.T.;CARBOGIM, J.B.P. **Conservação da natureza e eu com isso**. Fundação Brasil Cidadão. Fortaleza. 1a Edição. 2012.

PAULA, G. A. **Modelos de regressão com apoio computacional**. São Paulo: IME/USP, p. 428, 2013.

PREGIBON, D. Logistic regression diagnostics. **Annals of Statistics**. v. 9, p. 705-724, 1981.

RAZALI, N.M. WAH, Y.B. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests, **Jornal of Statistical Modeling and Analytics**. v. 2, p.21 - 33, 2011.

RIBEIRO, C.S.; MOREIRA, R. G. Fatores ambientais e reprodução dos peixes. **Revista da Biologia**, v. 8(1), p. 58-61, 2012.

RIGBY, R. A.; STASINOPOULOS, D. M. Generalized additive models for location, scale and shape. **Journal of the Royal Statistical Society**. Series C (Applied Statistics), Wiley Online Library, v. 54, n. 3, p. 507 - 554, 2005.

ROBERTSON, A.I.; S.J.M. BLABER. **Plankton, epibenthos and fish communities**. Tropical Mangrove Ecosystems (Coastal and Estuarine Studies). Washinhton, DC, American Geophysical Union, p. 236, 1992.

ROZAS, L.P.; ZIMMERMAN, R.J. Small-scale patterns of nekton use among marsh and adjacent shallow nonvegetated areas of the Galveston Bay Estuary, Texas (USA), **Marine Ecology Progress Series**, v. 193, p. 217-239, 2000.

RUSSELL, R. J. Origins of estuaries. **Estuaries**, Washington, v. 83, p. 93-99, 1967.

SANTOS, T.G.; GUSMÃO, L.M.O, LEITÃO-NEUMANN, S. CUNHA, A.G. ZOOPLÂNCTON COMO INDICADOR BIOLÓGICO DA QUALIDADE AMBIENTAL NOS ESTUÁRIOS DOS RIOS CARRAPICHO E BOTAFOGO, ITAMARACÁ - PE. **Rev. Bras. Enga. Pesca**, 2009

SHAPIRO, S.S.; WILK, M.B. An Analysis of Variance Test for Normality (Complete Samples). **Biometrika**, v. 52, n. 3/4, p. 591-611, 1965.

SCHWARZ, B. G. Estimating the dimension of a model. The annals of statistics, **Institute of Mathematical Statistics**, v. 6, n. 2, p. 461-464, 1978.

SILVA, L.A. Estudos sedimentológico, morfológico e hidrodinâmico do Canal de Santa Cruz - PE, 2008 Tese de (doutorado) - Universidade Federal de Pernambuco. Programa de Pós-Graduação em Geociência, 2008.

SIMONI, M.E.R. Dinâmica reprodutiva da SARDINHA-LAJE *Opisthonema oglinum*, Lesueur, 1818 capturada no litoral norte de Pernambuco, Brasil. Tese (Doutorado) – Universidade Federal Rural de Pernambuco, Programa de Pós-Graduação em Recursos Pesqueiros e Aqüicultura, Recife, BR-PE, 2019.

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

STERGIOU, K.I.; MOUTOPOULOS, D.K.; ARMENIS, G. Perish legally and ecologically: the ineffectiveness of the minimum landing sizes in the Mediterranean Sea. *Fisheries Management and Ecology*, v. 16, p. 368 - 375, 2009.

VASCONCELOS-FILHO, A.L. Alimentação da Sardinha Bandeira, *Opisthonema oglinum* (Le Sueur, 1817) no Canal de Santa Cruz. **Tropical Oceanography**. Recife, v. 14, p. 105-116, 1979.

VASCONCELOS-FILHO, A. L.; OLIVEIRA, A. M. E. Composição de ecologia da ictiofauna do Canal de Santa Cruz (Itamaracá – PE, Brasil). **Trab. Oceanog.** UFPE, Recife, v. 27, p. 101-113, 1999.

VASCONCELOS FILHO, A.L.; NEUMANN-LEITÃO, S.; ESKINAZI-LEÇA, E.; OLIVEIRA, A. M. E.. Hábitos alimentares de peixes consumidores primários do sistema estuarino de Itamaracá (Pernambuco, Brasil). **Revista Brasileira de Engenharia de Pesca**. Recife, v. 4, n. 1. p. 20-30, 2009.

VIDY, G. Estuarine and mangrove systems and the nursery concept: which is which? The case of the Sine-Saloum system (Senegal). **Wetlands Ecology and Management**, v. 8, p. 37-51, 2000.

VIEIRA, J.P.; J.P. CASTELLO & L.E. PEREIRA. Ictiofauna. Os ecossistemas costeiro e marinho do extremo sul do Brasil. **Ecoscientia**, Rio Grande, p.60-68, 1998.

VOUDOURIS, V.; GILCHRIST, R.; RIGBY, R.; SEDGWICK, J.; STASINOPOULOS, D. Modelling skewness and kurtosis with the BCPE density in GAMLSS. **Journal of Applied Statistics**, **Taylor & Francis**, v. 39, n. 6, p. 1279-1293, 2012.

WEISBERG, S. **Applied linear regression**. John Wiley, New York, third edition. 2005.

WHITEHEAD, P.J.P. Clupeoid fishes of the world (Suborder Clupeoidei). An annotated and illustrated catalogue of the herings, sardines, pilchards, sprats, anchovies, and wolf-herings. Part I: Chirocentridae, Clupeidae and Pristigasteridae. **FAO Fish Synopsis**, v. 7, p. 1-103, 1985.

Apêndices

APÊNDICE A – Código em R utilizado na aplicação

```

#-----
#-----
# Script da dissertacao
#-----
#-----

require(graphics)
require(gamlss)

dados <- read.csv('~ / biometria.csv')
opog<-subset(dados, especie=="opog")

#-----
# Graficos de dispersao

hist(opog$cp, ylab="Frequencia", xlab="Comprimento Padrao (mm)",
main="", col = "gray", ylim=c(0,4000))
plot(opog$mes.a, opog$cp, xlab="Mês", ylab="Comprimento Padrao (mm)",
cex.lab=1.2, cex.axis=1.2)
plot(opog$ano.a, opog$cp, xlab="Ano", ylab="Comprimento Padrao (mm)",
cex.lab=1.2, cex.axis=1.2)
plot(opog$area, opog$cp, xlab="Setor", ylab="Comprimento Padrao (mm)",
cex.lab=1.2, cex.axis=1.2)
plot(opog$mare.lanc, opog$cp, xlab="Mare no lanamento da rede",
ylab="Comprimento Padrao (mm)", cex.lab=1.2, cex.axis=1.2)
plot(opog$pre_m_mes, opog$cp, xlab="Milímetros de chuva acumulado",
ylab="Comprimento Padrao (mm)", cex.lab=1.2, cex.axis=1.2, xlim=c(0,500))

```

```

#-----
#-----
# GLM
#-----
#-----
#Modelo Normal Inversa

mod.mlg<- glm(cp~mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
family = inverse.gaussian(link = "identity"), data=opog)

summary(mod.mlg)

#-----
#Diagnostico

X <- model.matrix(mod.mlg)
n <- nrow(X)
p <- ncol(X)
w <- fit.model$weights
W <- diag(w)
H <- ginv(t(X)%*%W%*%X)
H <- sqrt(W)%*%X%*%H%*%t(X)%*%sqrt(W)
h <- diag(H)
soma <- resid(mod.mlg, type="pearson")
soma <- sum(soma^2)
fi <- (n-p)/soma
ts <- resid(mod.mlg, type="pearson")*sqrt(fi/(1-h))
td <- resid(mod.mlg, type="deviance")*sqrt(fi/(1-h))
par(mfrow=c(2,2))
di <- (h/(1-h))*(ts^2)
a <- max(td)
b <- min(td)
k<-qr(X)$rank
#
Leverage=2*length(mod.mlg$coefficients)/nrow(opog)      #2p/n
#
cook<-(qchisq(.05, df=k))/k

```

```

par(mfrow=c(2,2))
#
plot(fitted(mod.mlg),h,xlab="Valor Ajustado", ylab="Medida h",
pch=16)

abline(h=Leverage, col="red",lwd=1.3)
#identify(fitted(mod.mlg), h, n=1)
#
plot(fitted(mod.mlg), di,xlab="Valor Ajustado", ylab = "Distancia
de Cook",pch=16, ylim=c(0,1))

abline(h=cook, col="red",lwd=1.3)
#identify(fitted(mod.mlg),di, n=1)
#
plot(fitted(mod.mlg),td,xlab="Valor Ajustado", ylab="Residuo
Componente do Desvio",ylim=c(b-1,a+1),pch=16)

abline(3,0,lty=2,col="red")
abline(-3,0,lty=2,col="red")
#identify(fitted(mod.mlg),td, n=1)
#
eta <- predict(mod.mlg)
z <- eta + resid(mod.mlg, type="pearson")/sqrt(w)
plot(predict(mod.mlg),z,xlab="Preditor Linear",
ylab="Variavel z", pch=16)
lines(smooth.spline(predict(mod.mlg), z, df=2), col = "red")

#-----
#-----
# GAMLSS
#-----
#-----

#-----
#Modelo Box-Cox t Original

```

```
#-----  
  
mod.BCTO<- gamlss(cp~mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,  
family=BCTo, data=opog, trace=FALSE)  
  
summary(mod.BCTO)  
plot(mod.BCTO)  
wp(mod.BCTO)  
  
#-----  
#Modelo Box-Cox t  
#-----  
  
mod.BCT<- gamlss(cp~mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,  
family=BCT, data=opog, trace=FALSE)  
  
summary(mod.BCT)  
plot(mod.BCT)  
wp(mod.BCT)  
  
#-----  
#Modelo Beta Generalizada tipo II  
#-----  
  
mod.GB2<- gamlss(cp~mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,  
family=GB2, data=opog, trace=FALSE)  
  
summary(mod.GB2)  
plot(mod.GB2)  
wp(mod.GB2)  
  
#-----  
#Modelo Johnson SU  
#-----  
  
mod.JSU<- gamlss(cp~mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,  
family=JSU, data=opog, trace=FALSE)
```

```
summary(mod.JSU)
plot(mod.JSU)
wp(mod.JSU)
```

```
#-----
#Modelo Skew t tipo 3 repar.
#-----
```

```
mod.SST<- gamlss(cp~mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
family=SST, data=opog, trace=FALSE)
```

```
summary(mod.SST)
plot(mod.SST)
wp(mod.SST)
```

```
#-----
#Modelo Skew t tipo 3
#-----
```

```
mod.ST3<- gamlss(cp~mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
family=ST3, data=opog, trace=FALSE)
```

```
summary(mod.ST3)
plot(mod.ST3)
wp(mod.ST3)
```

```
#-----
#Modelo Box-Cox Exponencial potencia original
#-----
```

```
mod.BCPEo<- gamlss(cp~mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
family=BCPEo, data=opog, trace=FALSE)
```

```
summary(mod.BCPEo)
plot(mod.BCPEo)
wp(mod.BCPEo)
```

```

#-----
#Modelo Box-Cox Exponencial potencia
#-----

mod.BCPE<- gamlss(cp~mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
family=BCPE, data=opog, trace=FALSE)

summary(mod.BCPE)
plot(mod.BCPE)
wp(mod.BCPE)

#-----
#Modelo Exponencial Gaussian
#-----

mod.exGAUS<- gamlss(cp~mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
family=exGAUS, data=opog, trace=FALSE)

summary(mod.exGAUS)
plot(mod.exGAUS)
wp(mod.exGAUS)

#-----
#Modelo Skew $t$ tipo 5
#-----

mod.ST5<- gamlss(cp~mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
family=ST5, data=opog, trace=FALSE)

summary(mod.ST5)
plot(mod.ST5)
wp(mod.ST5)

#-----
# Modelos com configuracoes parametricas

```



```

#-----

#-----
#Modelo Box-Cox t Original (mu e sigma)
#-----

mod.BCTO <- gamlss(cp~mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
mu.fo = ~ mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
sigma.fo = ~ mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
family=BCTo, data=opog, trace=FALSE)

summary(mod.BCTO)
plot(mod.BCTO)
wp(mod.BCTO)

#-----

#Modelo Box-Cox t (mu, sigma e tau)
#-----

mod.BCT <- gamlss(cp~mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
mu.fo = ~ mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
sigma.fo = ~ mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
tau.fo = ~ mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
family=BCT, data=opog, trace=FALSE)

summary(mod.BCT)
plot(mod.BCT)
wp(mod.BCT)

#-----

#Modelo Beta Generalizada tipo II (mu, sigma e tau)
#-----

mod.GB2 <- gamlss(cp~mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
mu.fo = ~ mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
sigma.fo = ~ mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,

```

```
tau.fo = ~ mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
family=GB2, data=opog, trace=FALSE)
```

```
summary(mod.GB2)
plot(mod.GB2)
wp(mod.GB2)
```

```
#
#Modelo Skew t tipo 3 (mu e sigma)
#
```

```
mod.ST3 <- gamlss(cp~mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
mu.fo = ~ mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
sigma.fo = ~ mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
family=ST3, data=opog, trace=FALSE)
```

```
summary(mod.ST3)
plot(mod.ST3)
wp(mod.ST3)
```

```
#
#Modelo Skew t tipo 3 (mu, sigma, nu e tau)
#
```

```
mod.ST3 <- gamlss(cp~mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
mu.fo = ~ mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
sigma.fo = ~ mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
nu.fo = ~ mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
tau.fo = ~ mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
family=ST3, data=opog, trace=FALSE)
```

```
summary(mod.ST3)
plot(mod.ST3)
wp(mod.ST3)
```

```
#
#Modelo Skew t tipo 3 (mu e tau)
```

```

#-----
mod.ST3 <- gamlss(cp~mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
mu.fo = ~ mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
tau.fo = ~ mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
family=ST3, data=opog, trace=FALSE)

summary(mod.ST3)
plot(mod.ST3)
wp(mod.ST3)

#-----
#Modelo Beta Generalizada tipo II (mu, sigma, nu e tau)
#-----

mod.GB2 <- gamlss(cp~mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
mu.fo = ~ mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
sigma.fo = ~ mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
nu.fo = ~ mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
tau.fo = ~ mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
family=GB2, data=opog, trace=FALSE)

summary(mod.GB2)
plot(mod.GB2)
wp(mod.GB2)

#-----
#Modelo Johnson SU (mu, sigma e nu)
#-----

mod.JSU <- gamlss(cp~mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
mu.fo = ~ mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
sigma.fo = ~ mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
nu.fo = ~ mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
family=JSU, data=opog, trace=FALSE)

summary(mod.JSU)

```

```
plot(mod.JSU)
wp(mod.JSU)
```

```
#-----
#Modelo Box-Cox t Original (mu, sigma e nu)
#-----
```

```
mod.BCTO <- gamlss(cp~mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
mu.fo = ~ mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
sigma.fo = ~ mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
nu.fo = ~ mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
family=BCTo, data=opog, trace=FALSE)
```

```
summary(mod.BCTO)
plot(mod.BCTO)
wp(mod.BCTO)
```

```
#-----
#Modelo Box-Cox t (mu e sigma)
#-----
```

```
mod.BCT <- gamlss(cp~mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
mu.fo = ~ mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
sigma.fo = ~ mes.a+area+ano.a+mare.lanc+pre_m_mes+mes.a:ano.a,
family=BCT, data=opog, trace=FALSE)
```

```
summary(mod.BCT)
plot(mod.BCT)
wp(mod.BCT)
```

APÊNDICE B – Distribuições citadas no trabalho

B.1 Box-Cox t , $BCT(\mu, \sigma, \nu, \tau)$

Seja Y uma variável aleatória positiva com uma distribuição Box-Cox t , denotado por $BCT(\mu, \sigma, \nu, \tau)$, definido através da variável aleatória transformada Z dada por

$$Z = \begin{cases} \frac{1}{\sigma\nu} \left[\left(\frac{Y}{\mu} \right)^\nu - 1 \right], & \text{se } \nu \neq 0 \\ \frac{1}{\sigma} \log \left(\frac{Y}{\mu} \right)^\nu, & \text{se } \nu = 0 \end{cases}$$

em que, $0 < Y < \infty$, $\mu > 0$, $\sigma > 0$ e $-\infty < \nu < \infty$ onde a variável aleatória Z é assumida para seguir um t truncado distribuição com graus de liberdade, $\tau > 0$, tratado como um parâmetro contínuo.

A Box-Cox t modela assimetria e curtose, μ é o parâmetro de escala e σ é o coeficiente de variação aproximado. A função densidade de probabilidade de $Y \sim BCT(\mu, \sigma, \nu, \tau)$, é dada por

$$f_Y(y|\mu, \sigma, \nu, \tau) = \frac{1}{B\left(\frac{1}{2}, \frac{\tau}{2}\right) \tau^{1/2}} \left[1 + \frac{z^2}{\tau} \right]^{-(\tau+1/2)} \quad (\text{B.1})$$

para, $y > 0$, $\mu > 0$, $\sigma > 0$ e $-\infty < \nu < \infty$.

Função de ligação padrão: identidade para μ , \log para σ , identidade para ν e \log para τ .

B.2 Box-Cox t original, $BCTo(\mu, \sigma, \nu, \tau)$

A Box-Cox t original modela assimetria e curtose, μ é o parâmetro de escala e σ é o coeficiente de variação aproximado. A função densidade de probabilidade de $Y \sim BCT(\mu, \sigma, \nu, \tau)$, é dada por

$$f_Y(y|\mu, \sigma, \nu, \tau) = \frac{1}{B\left(\frac{1}{2}, \frac{\tau}{2}\right) \tau^{1/2}} \left[1 + \frac{z^2}{\tau} \right]^{-(\tau+1/2)} \quad (\text{B.2})$$

para, $y > 0$, $\mu > 0$, $\sigma > 0$ e $-\infty < \nu < \infty$.

Função de ligação padrão: \log para μ , \log para σ , identidade para ν e \log para τ .

B.3 Beta generalizada tipo II, GB2(μ, σ, ν, τ)

A Beta generalizada tipo II modela assimetria e curtose. A função densidade de probabilidade de $Y \sim GB2(\mu, \sigma, \nu, \tau)$, é dada por

$$f_Y(y|\mu, \sigma, \nu, \tau) = |\sigma|y^{\sigma\nu-1} \left\{ \mu^{\sigma\nu} B(\nu, \tau) [1 + (y/\mu)^\sigma]^{\nu+\tau} \right\}^{-1} = \frac{\Gamma(\nu + \tau)}{\Gamma(\nu)\Gamma(\tau)} \frac{|\sigma|(y/\nu)^{\sigma\nu}}{y [1 + (y/\mu)^\sigma]^{\nu+\tau}} \quad (\text{B.3})$$

para $y > 0$, onde $\mu > 0$, $\sigma > 0$, $\nu > 0$ e $\tau > 0$.

Função de ligação padrão: log para μ , log para σ , log para ν e log para τ .

B.4 Johnson SU, JSU(μ, σ, ν, τ)

Esta é uma reparametrização da distribuição original da *Johnson Su*, de modo que os parâmetros μ e σ são a média e o desvio padrão da distribuição. A distribuição *Johnson Su* é leptocúrtica e modela assimetria e curtose. A função densidade de probabilidade de $Y \sim JSU(\mu, \sigma, \nu, \tau)$, é dada por

$$f_Y(y|\mu, \sigma, \nu, \tau) = \frac{\tau}{c\sigma(s^2 + 1)^{1/2}\sqrt{2\pi}} \exp\left[-\frac{1}{2}z^2\right] \quad (\text{B.4})$$

para $-\infty < y < \infty$, onde $-\infty < \mu < \infty$, $\sigma > 0$, $-\infty < \nu < \infty$ e $\tau > 0$, e também

$$z = -\nu + \tau \sinh^{-1}(s) = -\nu + \tau \log \left[s + (s^2 + 1)^{1/2} \right],$$

$$s = \frac{y - \mu + c\sigma w^{1/2} \sinh(\nu/\tau)}{c\sigma},$$

$$c = \left\{ \frac{1}{2}(w - 1)[w \cosh(2\nu/\tau) + 1] \right\}^{-1/2}$$

where $w = \exp(1/\tau^2)$.

Função de ligação padrão: identidade para μ , log para σ , identidade para ν e log para τ .

B.5 Skew t tipo 1, ST1(μ, σ, ν, τ)

A distribuição Skew t tipo 1 modela assimetria e curtose. A função densidade de probabilidade de $Y \sim ST1(\mu, \sigma, \nu, \tau)$, é dada por

$$f_Y(y|\mu, \sigma, \nu, \tau) = \begin{cases} \frac{c}{\sigma_0} \left[1 + \frac{\nu^2 z^2}{\tau} \right]^{-(\tau+1)/2} & \text{se } y < \mu_0 \\ \frac{c}{\sigma_0} \left[1 + \frac{z^2}{\nu^2 \tau} \right]^{-(\tau+1)/2} & \text{se } y \geq \mu_0 \end{cases},$$

para $-\infty < y < \infty$, onde $-\infty < \mu < \infty$, $\sigma > 0$, $-\infty < \nu < \infty$, $\tau > 0$ e $z = (y - \mu)/\sigma$

Função de ligação padrão: identidade para μ , log para σ , identidade para ν e log para τ .

B.6 Skew t tipo 2, $ST2(\mu, \sigma, \nu, \tau)$

A distribuição Skew t tipo 2 também modela assimetria e curtose. A função densidade de probabilidade de $Y \sim ST2(\mu, \sigma, \nu, \tau)$, é dada por

$$f_Y(y|\mu, \sigma, \nu, \tau) = \frac{2}{\sigma} f_{Z_1}(z) F_{Z_2}(w) \quad (\text{B.5})$$

para $-\infty < y < \infty$, onde $-\infty < \mu < \infty$, $\sigma > 0$, $-\infty < \nu < \infty$, $\tau > 0$ e $z = (y - \mu)/\sigma$, $w = \nu \lambda^{1/2} z$, $\lambda = (\tau + 1)/(\tau + z^2)$.

Função de ligação padrão: identidade para μ , log para σ , identidade para ν e log para τ .

B.7 Skew t tipo 3, $ST3(\mu, \sigma, \nu, \tau)$

A Skew t tipo 3 é uma distribuição *spliced-scale* útil para quando existe interesse na moda, porque o parâmetro μ é a própria moda da distribuição. A função densidade de probabilidade de $Y \sim ST3(\mu, \sigma, \nu, \tau)$, é dada por

$$f_Y(y|\mu, \sigma, \nu, \tau) = \begin{cases} \frac{c}{\sigma} \left[1 + \frac{\nu^2 z^2}{\tau}\right]^{-(\tau+1)/2} & \text{se } y < \mu \\ \frac{c}{\sigma} \left[1 + \frac{z^2}{\nu^2 \tau}\right]^{-(\tau+1)/2}, & \text{se } y \geq \mu \end{cases}$$

para $-\infty < y < \infty$, onde $-\infty < \mu < \infty$, $\sigma > 0$, $\nu > 0$, $\tau > 0$ e $z = (y - \mu)/\sigma$, $c = 2\nu[(1 + \nu^2)B(1/2, \tau/2)\tau^{1/2}]^{-1}$.

Função de ligação padrão: identidade para μ , log para σ , log para ν e log para τ .

B.8 Skew t tipo 4, $ST4(\mu, \sigma, \nu, \tau)$

A Skew t tipo 4 é uma distribuição *spliced-shape*, seus parâmetros de forma estão associados às caudas esquerda e direita da distribuição. A função densidade de probabilidade de $Y \sim ST4(\mu, \sigma, \nu, \tau)$, é dada por

$$f_Y(y|\mu, \sigma, \nu, \tau) = \begin{cases} \frac{c}{\sigma} \left[1 + \frac{z^2}{\nu}\right]^{-(\nu+1)/2} & \text{se } y < \mu \\ \frac{c}{\sigma} \left[1 + \frac{z^2}{\tau}\right]^{-(\tau+1)/2}, & \text{se } y \geq \mu \end{cases}$$

para $-\infty < y < \infty$, onde $-\infty < \mu < \infty$, $\sigma > 0$, $\nu > 0$, $\tau > 0$ e $z = (y - \mu)/\sigma$, $c = 2[\nu^{1/2}B(1/2, \nu/2)\tau^{1/2}B(1/2, \tau/2)]^{-1}$.

Função de ligação padrão: identidade para μ , log para σ , log para ν e log para τ .

B.9 Skew t tipo 5, $ST5(\mu, \sigma, \nu, \tau)$

A distribuição Skew t tipo 5 modela assimetria e curtose. A função densidade de probabilidade de $Y \sim ST5(\mu, \sigma, \nu, \tau)$, é dada por

$$f_Y(y|\mu, \sigma, \nu, \tau) = \frac{c}{\sigma} \left[1 + \frac{z}{(a+b+z^2)^{1/2}} \right]^{a+1/2} \left[1 + \frac{z}{(a+b+z^2)^{1/2}} \right]^{b+1/2} \quad (\text{B.6})$$

para $-\infty < y < \infty$, onde $-\infty < \mu < \infty$, $\sigma > 0$, $-\infty < \nu < \infty$, $\tau > 0$ e $z = (y - \mu)/\sigma$, $c = [2^{a+b-1}(a+b)^{1/2}B(a, b)]^{-1}$, $\nu = (a - b/[ab(a+b)]^{1/2})$, $\tau/2/(a+b)$. Onde $a = \tau^{-1}[1 + \nu(2\tau + \nu^2)^{-1/2}]$ e $b = \tau^{-1}[1 - \nu(2\tau + \nu^2)^{-1/2}]$.

Função de ligação padrão: identidade para μ , log para σ , identidade para ν e log para τ .

B.10 Skew t de Student, $SST(\mu, \sigma, \nu, \tau)$

A distribuição Skew t de Student: é uma reparametrização de ST3. Modela assimetria e curtose. Seu parâmetro σ é o próprio desvio padrão da distribuição. A função densidade de probabilidade de $Y \sim SST(\mu, \sigma, \nu, \tau)$, é dada por

$$f_Y(y|\mu, \sigma, \nu, \tau) = \begin{cases} \frac{c}{\sigma_0} \left[1 + \frac{\nu^2 z^2}{\tau} \right]^{-(\tau+1)/2} & \text{se } y < \mu_0 \\ \frac{c}{\sigma_0} \left[1 + \frac{z^2}{\nu^2 \tau} \right]^{-(\tau+1)/2} & \text{se } y \geq \mu_0 \end{cases}$$

para $-\infty < y < \infty$, onde $-\infty < \mu < \infty$, $\sigma > 0$, $\nu > 0$, $\tau > 2$ e $\mu_0 = \mu - \sigma m/s$, $\sigma_0 = \sigma/s$, $z = (y - \mu_0)/\sigma_0$, $c = 2\nu[(1 + \nu^2)B(1/2, \tau/2)\tau^{1/2}]^{-1}$, $m = \frac{2\tau^{1/2}(\nu - \nu^{-1})}{(\tau-1)B(1/2, \tau/2)}$, $s = \frac{\tau}{(\tau-2)}(\nu^2 + \nu^{-2} - 1) - m^2$.

Função de ligação padrão: identidade para μ , log para σ , log para ν e $\log(\tau - 2)$ para τ .

B.11 Box-Cox exponencial potência, $BCPE(\mu, \sigma, \nu, \tau)$

A distribuição Box-Cox exponencial potência serve para modelagem da assimetria e curtose. Também μ é o parâmetro de escala e σ é o coeficiente de variação aproximado.

função densidade de probabilidade de $Y \sim BCPE(\mu, \sigma, \nu, \tau)$, é dada por

$$f_Y(y|\mu, \sigma, \nu, \tau) = \frac{1}{B\left(\frac{1}{2}, \frac{\tau}{2}\right) \tau^{1/2}} \left[1 + \frac{z^2}{\tau}\right]^{-(\tau+1/2)} \quad (\text{B.7})$$

para, $y > 0$, $\mu > 0$, $\sigma > 0$ e $-\infty < \nu < \infty$.

Função de ligação padrão: identidade para μ , log para σ , identidade para ν e log para τ .

B.12 Box-Cox exponential potência original, BCPEo(μ, σ, ν, τ)

A distribuição Box-Cox exponencial potência original tem como função densidade de probabilidade de $Y \sim BCPEo(\mu, \sigma, \nu, \tau)$ dada por

$$f_Y(y|\mu, \sigma, \nu, \tau) = \frac{1}{B\left(\frac{1}{2}, \frac{\tau}{2}\right) \tau^{1/2}} \left[1 + \frac{z^2}{\tau}\right]^{-(\tau+1/2)} \quad (\text{B.8})$$

para, $y > 0$, $\mu > 0$, $\sigma > 0$ e $-\infty < \nu < \infty$.

Função de ligação padrão: log para μ , log para σ , identidade para ν e log para τ .

B.13 Exponencial Gaussiana, exGAUS(μ, σ, ν)

A função densidade de probabilidade da distribuição Exponencial Gaussiana é definida por

$$f_Y(y|\mu, \sigma, \nu) = \frac{1}{\nu} \exp\left[\frac{\mu - y}{\nu} + \frac{\sigma^2}{2\nu^2}\right] \Phi\left(\frac{y - \mu}{\sigma} - \frac{\sigma}{\nu}\right) \quad (\text{B.9})$$

para, $-\infty < y < \infty$, $-\infty < \mu < \infty$, $\sigma > 0$ e $\nu > 0$.

Função de ligação padrão: identidade para μ , log para σ , identidade para ν e log para τ .