

Rayane Santos Leite

**APLICAÇÃO DE MODELOS ADITIVOS  
GENERALIZADOS PARA LOCAÇÃO, ESCALA E  
FORMA EM CLONES DE *Eucalyptus* spp., NO POLO  
GESSEIRO DO ARARIPE - PE**

RECIFE-PE

FEVEREIRO DE 2019



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

**APLICAÇÃO DE MODELOS ADITIVOS GENERALIZADOS PARA LOCAÇÃO,  
ESCALA E FORMA EM CLONES DE *Eucalyptus* spp., NO POLO GESSEIRO DO  
ARARIPE - PE**

Dissertação apresentada ao curso de  
Mestrado em Biometria e Estatística  
Aplicada, como pré-requisito para ob-  
tenção do título de mestre

**Área de concentração: Biometria e Esta-  
tística Aplicada**

Orientador: Prof. Dr. José Antônio  
Aleixo da Silva  
Coorientador: Prof. Dr. Tiago Almeida  
de Oliveira

**RECIFE-PE**

**FEVEREIRO DE 2019**

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

L533a Leite, Rayane Santos  
Aplicação de modelos aditivos generalizados para locação, escala e forma em clones de *Eucalyptus* spp., no polo gesseiro do Araripe-PE / Rayane Santos Leite. – 2019.  
47 f.: il.

Orientador: José Antônio Aleixo da Silva.  
Coorientador: Tiago Almeida de Oliveira.  
Dissertação (Mestrado) – Universidade Federal Rural de Pernambuco, Programa de Pós-Graduação em Biometria e Estatística Aplicada, Recife, BR-PE, 2019.  
Inclui referências.

1. Eucalipto 2. Pesquisa florestal 3. Estatística matemática  
4. Modelos matemáticos 5. Florestas – PE I. Silva, José Antônio Aleixo da, orient. II. Oliveira, Tiago Almeida de, coorient. III. Título

CDD 519.5

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

**APLICAÇÃO DE MODELOS ADITIVOS GENERALIZADOS PARA LOCAÇÃO,  
ESCALA E FORMA EM CLONES DE *Eucalyptus* spp., NO POLO GESSEIRO DO  
ARARIPE - PE**

Rayane Santos Leite

Dissertação apresentada ao curso de Mestrado  
em Biometria e Estatística Aplicada, como  
pré-requisito para obtenção do título de mes-  
tre

Orientador:

---

**Prof. Dr. José Antônio Aleixo da Silva**  
Orientador

Banca examinadora:

---

**Prof. Dra. Ana Patricia Bastos  
Peixoto**  
Universidade Estadual de Paraiba  
(UEPB)

---

**Prof. Dr. Moacyr Cunha Filho**  
Universidade Federal Rural de  
Pernambuco (UFRPE)

*A Jesus de Nazaré,  
Aos meus avós Margarida (in memoriam) e Pedro,  
Aos meus pais Redi e Fátima,  
Ao meu irmão Lucas.  
Com amor, DEDICO.*

# 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.

Aos meus pais Redi Ferreira Leite e Maria de Fátima Santos Leite, por serem meus maiores exemplos, pelas orações em meu favor, incentivos incondicionais, orientações e preocupações para que eu pudesse sempre percorrer o caminho certo, pelo amor a mim dedicado que foi imprescindível neste percurso de formação. As palavras não expressam a minha gratidão por terem feito do meu sonho o de vocês. Ao meu irmão Lucas Santos Leite por todo seu amor e carinho que me dedica e por sempre se preocupar e acreditar nos meus sonhos.

A minha avó materna Margarida (*in memoriam*) por todo legado que deixou por ser um exemplo de inspiração para minha vida, pelos valorosos ensinamentos de grande importância na minha educação. Ao meu avô materno Pedro pelo amor e incentivo que me foi essencial.

A minha tia Sira sua contribuição foi muito valiosa neste percurso, obrigada por sempre ser presente em minha vida. A minha tia Maria do Carmo que me acolheu em sua casa e me deu todo apoio, a tia Rosália e aos primos Sérgio e Symone a me vocês foram essenciais. A minha família sou grata.

Ao meu orientador Prof. José Antônio Aleixo da Silva, pelo esforço, os ensinamentos que me passou de forma generosa, a confiança e disponibilidade durante a orientação.

Ao meu coorientador Prof. Tiago Almeida de Oliveira, pelo esforço, os ensinamentos a me transmitido de forma generosa, por toda disponibilidade, ensinamentos, confiança, incentivos, pelo imenso apoio e colaboração para que eu chegasse até aqui.

Às minhas grandes amigas de infância Ana Stefella, Roana e Anailza que mesmo na distância sempre se fizeram presentes em minha vida, colaborando com palavras de carinho.

A minha amiga Aline e família pelo acolhimento em sua casa, para as reuniões com o coorientador em Campina Grande - PB.

A minha amiga Jordanya pelos conselhos, disponibilidades, que sempre me deu um ombro amigo quando eu precisei.

Aos meus amigos acadêmicos Elielma, Nilton, Josafá, Érika, Alessandro, Rosendo, Diego, Keroly, André, Wesley, Eucymara e Fábio.

A minha amiga Elielma pela amizade foi um prazer imenso conviver com você.

A banca examinadora e em especial a minha ex-professora Ana Patricia Bastos Peixoto, pessoa a quem tenho muita admiração.

À CAPES, pela concessão de bolsa de estudos, à UFRPE, ao DEINFO, a todo o corpo docente pela oportunidade de se desenvolver academicamente, ao coordenador Prof. Moacyr Cunha Filho por sempre esta disposição para ajudar e ao secretário do programa ao Marco Antônio.

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

*Mesmo que as montanhas oscilassem e as colinas se abalassem,  
jamais meu amor te abandonará e jamais meu pacto de paz vacilará,  
diz o Senhor que se compadeceu de ti.  
(Isaías 54:10)*



## Resumo

A quantificação volumétrica é um importante parâmetro a ser avaliado nas florestas de rápido crescimento. E a modelagem é uma ferramenta que permite caracterizar a estrutura da floresta e representar o crescimento ao longo do tempo. O presente trabalho aborda características dos modelos aditivos generalizados para localização, escala e forma (GAMLSS), expondo as técnicas de estimação que inserem os algoritmos iterativos, as técnicas de inferência e previsão, as distribuições, os termos aditivos, as ferramentas de seleção de modelos e as análises de resíduos. Destaca-se abordar a flexibilidade da classe destes modelos, possibilitando ajustar todos os parâmetros em função das variáveis explicativas. Nos GAMLSS, existe uma extensa série de distribuições para a variável resposta que inclui funções de suavização e modelos flexíveis capazes de descrever a realidade. Visou-se compreender e apresentar como tal abordagem permite ser utilizada para dados de experimentação florestal, com uma aplicação prática no estudo da produtividade dos clones de *Eucalyptus* spp. no Polo Gesseiro do Araripe, com os dados provenientes de um experimento composto por 15 clones de *Eucalyptus* spp., objetivou-se assim com este trabalho selecionar os melhores clones com base na escolha do melhor modelo ajustado. Para selecionar este melhor modelo, foi utilizado o critério de Akaike generalizado (GAIC). Verificou-se a adequabilidade do modelo via análise residual, por meio do gráfico de *worm plot* e do teste de Shapiro-Wilks. Os resultados encontrados indicaram que o modelo GAMLSS ajustado com a distribuição *Skew t* do tipo 4 aplicado aos clones de *Eucalyptus* spp., foi o mais adequado e que os clones 39 e 41 de *Eucalyptus urophylla* obtiveram o melhor desempenho em relação aos demais clones.

**Palavras-chave:** *Eucalyptus urophylla*. Experimentação florestal. Modelos flexíveis.

# Abstract

Volumetric quantification is an important parameter to be evaluated in fast growing forests. And the Modeling is a tool that allows characterizing the forest structure and represents growth over time. The present work deals with the characteristics of the generalized additive models for location, scale and shape (GAMLSS), exposing the estimation techniques that insert the iterative algorithms, the inference and prediction techniques, the distributions, the additive terms, the model selection tools and analysis of residues, highlighting the flexibility of the class of these models, making it possible to adjust all the parameters according to the explanatory variables. In the GAMLSS, there is an extensive series of distributions for the response variable that includes smoothing functions and flexible models capable of describing reality. Aiming to understand and present how this approach allows to be used for data of forest experimentation, with a practical application in the productivity study of the clones of *Eucalyptus* spp. in the Gypsum Pole of Araripe, with data from an experiment composed of 15 clones of *Eucalyptus* spp., the objective of this work was to select the best clones based on the choice of the best fitted model. To select the best model, the generalized Akaike criterion (GAIC) was used. The precision of the selected equation was verified by the residue analysis and by the worm plot and the Shapiro-Wilks test. The results indicated that the GAMLSS model adjusted with the Skew  $t$  distribution of type 4 applied to the clones of *Eucalyptus* spp, was the most adequate and that the clones 39 and 41 of *Eucalyptus urophylla* performed better than the others.

**Key-words:** *Eucalyptus urophylla*. Forest experimentation. Flexible models.

# Lista de Figuras

Figura 1 – Mapa destacando as Cidades mais relevantes do Polo Gesseiro do Araripe-PE. Fonte: (GADELHA, 2010). . . . .	4
Figura 2 – Diferentes problemas que podem ser diagnosticados pelo <i>worm plot</i> (STASINOPOULOS et al., 2017). . . . .	23
Figura 3 – Desenho esquemático da parcela experimental. Fonte: Barros (2010). . . . .	24
Figura 4 – Histograma para o volume ( $m^3$ ) dos clones de <i>Eucalyptus</i> spp.. . . . .	28
Figura 5 – Boxplot para o volume ( $m^3$ ) dos clones de <i>Eucalyptus</i> spp.. . . . .	29
Figura 6 – Análise de resíduos do Modelo Normal para o <i>Eucalyptus</i> spp.. . . . .	29
Figura 7 – <i>Worm plot</i> do Modelo Normal aplicado aos clones de <i>Eucalyptus</i> spp.. . . . .	30
Figura 8 – Box-Cox da transformação da variável volume. . . . .	31
Figura 9 – Análise de resíduos e <i>worm plot</i> para o modelo Gama (M2). . . . .	33
Figura 10 – Análise de resíduos e <i>worm plot</i> para os modelos ajustados SASHo2 (M11) (a), <i>Skew t</i> com modelagem do parâmetro $\tau$ (M10) (b), Gumbel Reversa (M3) (c), Gama com modelagem do parâmetro $\sigma$ (M6) (d). . . . .	35
Figura 11 – Análise de resíduos e <i>worm plot</i> para a modelagem via <i>Skew t</i> tipo 4. . . . .	36

# Lista de tabelas

Tabela 1 – Interpretação de diferentes formatos do <i>worm plot</i> (STASINOPOULOS et al., 2017). . . . .	22
Tabela 2 – Descrição dos tratamentos de <i>Eucalyptus</i> spp. do experimento no Polo Gesseiro do Araripe. . . . .	25
Tabela 3 – Principais distribuições candidatas a modelar os dados de volume ( $m^3$ ) de <i>Eucalyptus</i> spp. sob modelo nulo. . . . .	32
Tabela 4 – Descrição dos principais modelos ajustados à <i>Eucalyptus</i> spp. do experimento no Polo Gesseiro do Araripe, com marcação dos parâmetros considerados ( $\times$ ). . . . .	33
Tabela 5 – Principais modelos ajustados via modelagem GAMLSS para os dados de <i>Eucalyptus</i> spp. . . . .	34
Tabela 6 – Estimativa dos parâmetros para o ajuste da <i>Skew t</i> do tipo 4. . . . .	36
Tabela 7 – Médias dos tratamentos e teste Tukey para modelagem GAMLSS <i>Skew t</i> tipo 4 e via distribuição normal. . . . .	37

## Lista de abreviaturas e siglas

AIC	Critério de Informação de Akaike
APL	Arranjo Produtivo Local
ASSOGESSO	Associação Nacional dos Fabricantes e Comerciantes de Gesso e seus Derivados
ATECEL	Associação Técnica Científica Ernesto Luiz de Oliveira
BIC	Critério de Informação Bayesiano
BPF	Baixo Poder de Fusão
CAP	Circunferência a Altura do Peito
CNA	Confederação da Agricultura e Pecuária do Brasil
CPEF	Companhia Paulista de Estrada de Ferro
DAP	Diâmetro a Altura do Peito
DIC	Delineamento Inteiramente Casualizado
EGB2	Exponencial Beta Generalizada Tipo II
fdp	Função Densidade de Probabilidade
fp	Função de Probabilidade
GAIC	Critério de Akaike Generalizado
GAM	Modelos Aditivos Generalizados
GAMLSS	Modelos Aditivos Generalizados para Localização, Escala e Forma
GEDV	Desvio Global
GLM	Modelos Lineares Generalizados
Ht	Altura total
IBGE	Instituto Brasileiro de Geografia e Estatística

IDH	Índice de Desenvolvimento Humano
IPA	Instituto Agrônômico de Pernambuco
JSU	Johnson SU reparação
JSUo	Johnson SU original
LO	Logística
MRL	Modelos de Regressão Linear
NO	Normal
PE	Power Exponencial
PE2	Power Exponencial 2
Pevs	Produção da Extração Vegetal e da Silvicultura
PIB	Produto Interno Bruto
RG	Gumbel Reserva
SBPCPE	Sociedade Brasileira para o Progresso da Ciência Regional Pernambuco
SBS	Sociedade Brasileira de Silvicultura
SEBRAE	Serviço Brasileiro de Apoio às Micro e Pequenas Empresas
SEP1	Skew exponencial potência tipo I
SEP2	Skew exponencial potência tipo II
SEP3	Skew exponencial potência tipo III
SEP4	Skew exponencial potência tipo IV
SHASH	Sinh-Arcsinh
SHASHo	Sinh-Arcsinh original
SHASHo2	Sinh-Arcsinh original tipo II
SINDUSGESSO	Sindicato das Indústrias do Gesso do Estado de Pernambuco
SN1	Skew Normal tipo 1
SN2	Skew Normal tipo 2

SST	Skew t de Student
ST1	Skew t tipo 1
ST2	Skew t tipo 2
ST3	Skew t tipo 3
ST4	Skew t tipo 4
ST5	Skew t tipo 5

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Objetivos</b>	<b>3</b>
2.1	Objetivo geral	3
2.2	Objetivos específicos	3
<b>3</b>	<b>Revisão de Literatura</b>	<b>4</b>
3.1	Polo Gesseiro do Araripe	4
3.1.1	<i>Eucalyptus</i>	6
3.1.2	Florestas energéticas de <i>Eucalyptus</i>	6
3.2	Modelos de Regressão	8
3.2.1	Modelos aditivos generalizados (GAM)	11
3.2.2	Modelos Aditivos Generalizados para Locação, Escala e Forma (GAMLSS)	11
3.2.2.1	Estimação	14
3.2.2.2	Inferência	17
3.2.2.3	Tipos de distribuições dentro da família GAMLSS	19
3.2.2.4	Critério de seleção dos modelos	19
3.2.2.5	Análise de resíduos	21
<b>4</b>	<b>Metodologia</b>	<b>24</b>
4.1	Localização do Experimento	24
4.2	Caracterização do Experimento	24
4.3	Distribuições probabilísticas	26
<b>5</b>	<b>Resultados e Discussão</b>	<b>28</b>
<b>6</b>	<b>Conclusão</b>	<b>39</b>
	<b>Referências Bibliográficas</b>	<b>40</b>



# 1 Introdução

O Arranjo Produtivo Local (APL) do Polo Gesseiro do Araripe é o principal produtor de gesso do Brasil. A principal fonte energética da indústria do gesso é a madeira proveniente da Caatinga, principalmente em função do baixo custo quando comparado com outras fontes energéticas. Porém, a exploração descontrolada desse recurso se reflete em um desequilíbrio ambiental desse domínio fitogeográfico. Por este fato, o reflorestamento com espécies de rápido crescimento com fins energéticos é de grande importância nessa região (ALVES, 2007). Ligado a esse fator nasce a necessidade de um planejamento energético com alta eficiência nessa região para atender a demanda da produção de gesso (GADELHA, 2010).

A implantação de florestas com elevados índices de crescimento, como é o caso das florestas do gênero *Eucalyptus* que apresentam inúmeros usos, constituem uma alternativa viável para abrandar a exploração da vegetação da Caatinga, reduzindo impactos sobre essas florestas nativas em função da oferta de madeira em maior volume em menor espaço temporal (PIERRO, 2015; PINTO et al., 2014).

No sentido de atingir êxito em plantações de eucaliptos, além da escolha adequada da espécie, é preciso o conhecimento sobre fatores geológicos e climáticos da região, dado que podem afetar direto ou indiretamente no desenvolvimento da vegetação, e consequentemente, na produtividade (ROMERO et al., 2003).

Devido ao aumento da produção, investigar com precisão as principais características da espécie se torna relevante, para tanto, o uso de uma ferramenta estatística mais robusta se faz necessário. Sabe-se também que, algumas técnicas de análise estatísticas tem fortes pressupostos que, caso sejam violadas, refletem resultados duvidosos e fora da realidade. Devido a este cenário, novas técnicas estatísticas têm surgido, sendo elas mais flexíveis e menos restritivas.

De modo genérico os GLM dispõem de duas mudanças relacionadas ao modelo de regressão linear (MRL) que são cruciais para a compreensão dos GAMLSS. Uma é a flexibilidade da distribuição da variável resposta que abrange não só a Normal, mas também uma classe qualquer uma distribuição que pertença da família exponencial. Outra é a função de ligação para modelar a relação entre  $E(\mathbf{Y})$  e as variáveis explicativas.

Dentre as técnicas de modelagem, a análise de regressão tem sido amplamente utilizada para modelar fenômenos biológicos, e suas diversas modalidades tem conseguido descrever o fenômeno com acurácia e precisão. Com o 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), técnica essa que, consegue reunir um grupo de modelos lineares e não-lineares, modelos lineares generalizados (*Generalized Linear Model* – GLM) (NELDER; WEDDERBURN, 1972), modelos aditivos generalizados (GAM) (HASTIE; TIBSHIRANI, 1990) e modelos mistos.

Os GAMLSS proporcionam o ajuste de modelos que admitem inúmeras distribuições para a variável resposta, sem pertencer obrigatoriamente à família exponencial. São modelos capazes de integrar as funções não paramétricas e/ou efeitos aleatórios, e todo e qualquer parâmetro pode ser modelado. Em resumo, os GAMLSS correspondem a uma modelagem apurada e flexível, que propõe uma nova classe de modelos, que abrange as demais, possuindo como pressuposto que as observações sejam independentes.

Assim, entende-se que a classe de modelos GAMLSS é uma forte ferramenta quando se trata de modelagem, visto que contribui com o pesquisador em fases que por diversas vezes são realizadas de maneira intuitiva, como a seleção da melhor distribuição para determinado banco de dados, além de achar o melhor modelo para situação imposta de maneira ágil e com certa desenvoltura em relação a linguagem de programação. Um outro ponto considerável é a questão que permite modelar não só a média, como nos modelos lineares generalizados (GLM) e modelos aditivos generalizados (GAM), mas também a dispersão, assimetria e a curtose, gerando mais possibilidades para o pesquisador e, em consequência, alcançando melhores ajustes no processo de modelagem.

## 2 Objetivos

Os objetivos propostos para execução desse trabalho são descritos abaixo.

### 2.1 Objetivo geral

Modelar a produtividade dos clones de *Eucalyptus* spp., plantados no Polo Gesseiro do Araripe – Pernambuco, Modelos Aditivos generalizados para Localização, Escala e Forma (GAMLSS) a dados assimétricos.

### 2.2 Objetivos específicos

- i) Ajustar e comparar uma sequência de Modelos Lineares Generalizados (GLM) e Modelos Aditivos generalizados para Localização, Escala e Forma (GAMLSS) com nível crescente de complexidade, em clones de *Eucalyptus* spp.;
- ii) Comparar diferentes clones no espaçamento (3,0 m × 2,0 m) usando distribuições da classe de Modelos Aditivos generalizados para Localização, Escala e Forma (GAMLSS) e selecionar a modelo que apresenta melhor desempenho (qualidade de ajuste e conformidade dos resíduos).

## 3 Revisão de Literatura

Neste capítulo serão abordados de maneira breve a motivação e contextualização das metodologias de análise de regressão clássica, GLM e GAM, proporcionando uma conexão introdutória para entender os GAMLSS e compreender como foram desenvolvidos. Apresenta-se também a estimação, inferência, critério de seleção do modelo e análise de resíduos em GAMLSS.

### 3.1 Polo Gesseiro do Araripe

É na região do Araripe que fica localizado o Arranjo Produtivo Local (APL) do gesso, denominado de Polo Gesseiro do Araripe que é composto pelas cidades de Araripina, Bodocó, Ipubi, Ouricuri e Trindade (Figura 1) (GADELHA, SILVA, FERREIRA, 2018), mundialmente conhecido por ser o principal centro de produção de gesso a partir do mineral gipsita considerada de altíssima qualidade (GRANJA et al., 2017) com a pureza variando entre 80% e 95% (JOGENSEN, 1994).



Figura 1 – Mapa destacando as Cidades mais relevantes do Polo Gesseiro do Araripe-PE.  
Fonte: (GADELHA, 2010).

A indústria do gesso em Pernambuco tem uma grande concentração das reservas de gipsita, que chegam a produzir aproximadamente mais de 95% da produção do país (MELO et al., 2017). Segundo informações da Associação Nacional dos Fabricantes e Comerciantes

de Gesso e seus Derivados (ASSOGESSO), o Polo Gesseiro do Araripe é resultante de 22 mineradoras, 154 indústrias calcificadoras de gipsita, são responsáveis pelo processo de transformação do minério gipsita em gesso (GADELHA; SILVA; FERREIRA, 2018). Na região são gerados cerca de 13,9 mil empregos diretos e 69 mil indiretos, propiciando um faturamento anual na ordem de R\$ 1,4 bilhões/ano (SINDUSGESSO, 2014), fator importante para o desenvolvimento econômico dessa região.

A transformação da gipsita em gesso é realizada pelo processo de desidratação do minério em que se consome muita energia, normalmente proveniente em especial da madeira de plantas nativas que são retiradas da Caatinga que não atende a demanda do consumo por parte da indústria do gesso (SBPCPE, 2014).

Por conta da grande demanda de lenha exigida pela indústria de desidratação da gipsita, aumenta-se a devastação do acervo florestal da região, formando uma pressão sobre a vegetação nativa, pois a matriz energética é extremamente dependente do uso da lenha da região, provocando alterações nas estruturas físicas e biológicas do ambiente (ALBUQUERQUE, 2002; SILVA, 2008/2009).

Embora, a matriz energética da indústria do gesso do Araripe possa ser considerada diversificada: 3% de energia elétrica, 5% de óleo diesel, 8% de óleo BPF (baixo poder de fusão), 10% de coque e 73% de lenha (ATECEL, 2006). Nos pequenos empreendimentos esta matriz energética é predominantemente baseada na lenha, devido os baixos custos, resultando muitas vezes em devastação de grandes áreas da Caatinga. Existem alguns planos de manejo florestal sustentável na região, mas em sua totalidade eles não têm condições de suprir a demanda da indústria do gesso, pois chegam a contribuir no máximo com 5% dessa demanda (SILVA, 2008/2009). Exige-se, portanto, estratégias para conter a devastação florestal que enquadrem a aplicação de técnicas de plantio, planos e manejo florestal sustentado, além de medidas para atenuar o processo de savanização, ocasionada pela prática inadequada de exploração florestal.

Estima-se que, anualmente, o Polo Gesseiro do Araripe consome 652.680  $m^3$  de lenha e que para atender a demanda de energéticos florestais do parque industrial do Araripe, necessita-se de 14.036 ha/ano explorados com planos de manejo sustentado da vegetação, isto implica em uma área total de 219.541,94 ha de exploração florestal sustentável (GADELHA, 2014).

A degradação ambiental dessa microrregião está ligada à industrialização de minerais, especialmente a gipsita, pois a matriz energética é composta pelos derivados de petróleo, pela eletricidade e lenha. Mas a lenha mesmo utilizada só no processo de calcinação é explorada excessivamente por métodos arcaicos que provocam a degradação ambiental danificando a fauna e a flora da região (CUNHA et al., 2008).

### 3.1.1 *Eucalyptus*

O *Eucalyptus* é um gênero pertencente à família Myrtaceae, compreende mais de 700 espécies, sendo quase todas originárias da Austrália e da Tasmânia. O nome do gênero é proveniente do grego *eũ*, “bem” + *kalyptós*, “coberto”), que significa “bem coberto”. E algumas dessas espécies é uma importante fonte de matéria-prima para a indústria madeireira e de celulose e do mesmo modo de óleos essenciais de grande valor para a indústria farmacêutica e de perfumaria (FIGUEIREDO et al., 2013).

A introdução do *Eucalyptus* no Brasil por volta de 1825, como espécie ornamental e como quebra-vento. Para finalidades econômicas seu uso se iniciou em 1903, para suprir a necessidade de dormentes para atender ao desenvolvimento das estradas de ferro da Companhia Paulista de Estradas de Ferro – CPEF (ANDRADE, 1961).

O precursor do cultivo do eucalipto no Brasil foi o Engenheiro Agrônomo Edmundo Navarro de Andrade, que realizou vários experimentos no Horto Florestal de Jundiaí – SP, nos quais detectou o rápido desenvolvimento do gênero, comparado a espécies nativas (MORA; GARCIA, 2000).

A vasta diversidade do gênero permite sua utilização na produção de papel e celulose, carvão, dormentes, estacas, lenha e postes, incluindo também atribuições de origem não-madeireira relacionadas à crédito de carbono, artesanato, medical, extração de óleo essencial e dentre outras (SILVA, 2008/2009).

### 3.1.2 Florestas energéticas de *Eucalyptus*

Em um passado recente a necessidade de madeira era suprida quase que exclusivamente por meio das florestas nativas, que ao serem exploradas sem controle, muitas vezes, provocavam danos irreversíveis a alguns ecossistemas. Um hectare de floresta plantada de eucalipto produz muitas vezes mais que um hectare de florestas tropicais nativas (SBS, 2011), o que viabiliza economicamente o plantio, além do ganho na diminuição impacto ambiental. A madeira do eucaliptos é útil para as mais variadas áreas, e isto tem incentivado a instalação de florestas de inúmeras utilidades podendo atender diversos segmentos industriais. A seleção da espécie depende especialmente das circunstâncias climáticas da área a ser plantada, relacionada a aspectos físicos e químicos do solo da região para qual se quer instalar o empreendimento (FERREIRA, 2003; SCAVINSK, 2014).

As espécies do gênero *Eucalyptus* que impressionam por seu rápido crescimento, reprodução, qualidade da madeira, fácil adaptação as mais variadas condições ambientais, solos e às modernas técnicas de manejo florestal, são as mais utilizadas no cenário das florestas energéticas no Brasil (LIMA, 1996).

O desenvolvimento tecnológico proporcionou as florestas energéticas mais utilização como fonte de matéria-prima para produção de energia elétrica, associado ao alto índice de produtividade obtido em florestas plantadas (especialmente de clones do gênero *Eucalyptus*), devido a redução dos custos da produção da eletricidade com madeira de reflorestamento, tornando o investimento mais interessante (VALE et al., 2014).

Nos últimos 10 anos foi perceptível a ampliação do gênero *Eucalyptus* no Brasil, onde o país se notabilizou no comércio internacional por ter sido um dos principais produtores de celulose, dessa forma estimulando, muitos investimentos em pesquisa e produção (GONÇALVES et al., 2017). Assim, o país se encontra em 4º lugar no ranking dos produtores de celulose contribuindo para o PIB com R\$ 69,1 bilhões o que representa 1,2% da riqueza desenvolvida no país (CNA, 2017). As florestas plantadas correspondem a uma área de 7,84 milhões de hectares de reflorestamento. Por isso, os plantios florestais no Brasil ocupam nos últimos anos cerca 9,85 milhões de hectares, e desse valor 75,2% são ocupadas por eucalipto e 20,6% com pinus, dados expressos pelo levantamento Produção da Extração Vegetal e da Silvicultura (Pevs) 2017 (IBGE, 2017).

A madeira é um dos principais recursos naturais e encontra-se praticamente em todas as regiões. Entretanto, a utilização da mesma a cerca da fonte energética pela humanidade, existe desde o princípio, principalmente quando se refere a sua forma bruta, tradicionalmente conhecida como lenha (BRITO, 2007). Contudo, apesar da sua antiga utilização, a mesma não caiu em desuso e é extensamente utilizada para os mais variados fins, sendo uma das fontes mais competitivas quando se considera sua relação benefício/custo (BUAINAIN; BATALHA, 2007).

As florestas energéticas também são instaladas com a finalidade de evitar o desmatamento de florestas naturais, pois favorecem o fornecimento de biomassa florestal, lenha e carvão de origem vegetal, e ainda desenvolvem um importante papel na utilização de terras degradadas (SEBRAE, 2015).

A mensuração florestal iniciou-se conjuntamente com a ciência florestal, tendo sido inserida como disciplina nos estados germânicos no século XVIII. Desenvolveu-se no século XIX, por adotar técnicas e informações matemáticas. Entretanto, a inserção de métodos estatísticos para análises no domínio florestal aconteceu nos Estados Unidos do século XX. A aplicação das técnicas estatísticas de modelagem para caracterizar um fenômeno florestal de maneira mais simples, aconteceu depois da elaboração das primeiras tabelas de volume, denominadas “tabelas empíricas”. Os resultados empíricos consentiram o aperfeiçoamento de técnicas de mensuração e propiciaram o uso de regressão linear na área florestal, conforme no trabalho de Trorey, que aplicou esta técnica com o intuito de mostrar a relação que existe entre a altura e o diâmetro das árvores (BATISTA, 2014).

## 3.2 Modelos de Regressão

A análise de regressão é uma técnica estatística que tem como finalidade descrever por meio de um modelo fenômenos físicos ou biológicos, investigando a relação existente entre variáveis dependente e independente(s), a partir de observações dessas variáveis, podendo ser aplicada a diversas áreas do conhecimento. A técnica foi implementada preliminarmente pelo estatístico e antropólogo Britânico, Sir Francis Galton, no século XIX (CASELLA; BERGER, 2010). O modelo de regressão linear é uma das categorias dos modelos de regressão, conforme Draper e Smith (1998) e pode ser expresso como:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad i = 1, \dots, n \quad (3.1)$$

em que,  $X_i$  representa cada observação da variável explicativa  $X$ ;  $\beta_0$  representa o parâmetro linear da reta, ou seja, representa o ponto inicial para a variável  $Y$ ;  $\beta_1$  representa o parâmetro angular da reta, isto é, o grau que a reta faz com o eixo  $X$ , e define também o quanto aumenta ou diminui o valor de  $Y$  em relação a  $X$ ; por fim,  $\varepsilon_i$ , representa o erro associado a cada observação em relação a reta de regressão linear.

O ajuste do modelo de regressão linear simples exige algumas suposições, para que se possa realizar inferências e assim ter sentido o seu uso prático. Segundo Hoffmann (2006) as suposições são:

- i. Relação linear entre  $X$  e  $Y$ ;
- ii. Os valores de  $X_i$  são fixos, ou seja,  $X$  não é uma variável aleatória;
- iii. A média do erro é nula, logo  $E(\varepsilon_i) = 0$ ;
- iv. A variância do erro, ou variância residual, é uma constante igual a  $\sigma^2$ , para todos os valores de  $X$ ;
- v. Os erros são não correlacionados,  $E(\varepsilon_i \varepsilon_j) = 0$  para todo  $i \neq j$ ;
- vi. Os erros têm distribuição normal, ou seja,  $\varepsilon \sim N(0, \sigma^2)$

De forma geral, pode-se dizer que o modelo de regressão linear múltiplo é qualquer modelo de regressão linear com duas ou mais variáveis independentes com o objetivo de prever uma variável dependente  $Y$ , isto é, propicia ao pesquisador estimar a resposta de uma variável, levando em conta os valores de duas ou mais variáveis explicativas (GANZOLA, 2002), sendo expresso por:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i \quad i = 1, \dots, n \quad (3.2)$$



Segundo Barreto (2016) no ajuste do modelo de regressão linear múltipla é mais apropriado fazer uso de operações matriciais. Assim, pode-se reescrever o modelo algébrico expresso em 3.2 na forma matricial dada por,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (3.3)$$

em que, na expressão 3.3, tem-se que

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \text{ e } \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

em que,  $\mathbf{Y}$  é o vetor com  $n$  variáveis respostas, com dimensão  $n \times 1$ ;  $\mathbf{X}$  representa a matriz de variáveis independentes, com dimensão  $n \times (k + 1)$  denominando a matriz do modelo, cuja a primeira coluna é reservada para constante 1 e as demais colunas para as  $k$  variáveis independentes;  $\boldsymbol{\beta}$  é o vetor de parâmetros desconhecidos do modelo (ou coeficientes de regressão), com dimensão  $(k + 1) \times 1$ ; e  $\boldsymbol{\varepsilon}$  o vetor de valores aleatórios (ou erros aleatórios) do modelo supostos independentes e normalmente distribuídos de média zero e variância  $\sigma^2$ , com dimensão  $n \times 1$ .

Se  $\mathbf{X}$  é a matriz do delineamento e também é uma matriz de posto completo, resulta uma solução única do sistema de equações normais, pois se refere ao modelo de regressão linear. No caso de delineamentos experimentais, a matriz  $\mathbf{X}$  consistirá em uma matriz de posto incompleto. Assim, deve-se reparar à correta caracterização da matriz de parâmetros e alguma restrição deve ser utilizada (SEARLE; CASELLA; MCCULLOCH, 2009). Em se tratando do termo linear, atribui-se à relação dos parâmetros com a esperança da variável resposta, que, presume-se, ser linear. Se  $\mathbf{X}$  é de posto incompleto então  $(X'X)$  não tem inversa simples, e permite-se utilizar a teoria de matrizes inversas generalizadas que aplicam restrições para resolver o problema (RENCHEER; SCHAALJE, 2008).

A declaração de que a variável resposta tem distribuição normal é um pressuposto rigoroso e necessário para o seu uso correto, mas elimina uma série de situações práticas, como variáveis respostas binárias, de proporção, de contagem ou ainda categóricas. Tem muita aplicabilidade pela sociedade acadêmica, especialmente por volta do século passado (STASINOPOULOS et al., 2017), sendo amplamente aplicado até hoje, mas os modelos de regressão linear (MRL) (RODGERS; NICEWANDER, 1988) são inadequados para várias situações do cotidiano da pesquisa.

Nesta circunstância, Nelder e Wedderburn (1972) sugeriram os Modelos Lineares Generalizados (GLM) sua ideia básica é expandir as opções de distribuições de probabilidades 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).

Essencialmente, os GLM relacionam a variável resposta com as variáveis explicativas mediante uma função de ligação, que têm por finalidade assegurar que os parâmetros consigam atribuir quaisquer valores reais mantendo os valores ajustados no domínio da distribuição. Essa teoria unificadora de diversos modelos estatísticos considera de preferência as distribuições da família exponencial. Deste modo, expressa-se como  $Y \sim F\varepsilon(\mu, \phi)$  em que  $\mu$  e  $\phi$  são vetores dos parâmetros de locação e escala da distribuição, respectivamente, e  $\phi$  é constante. Distribuições comuns como a Normal, Beta normal inversa, Binomial, Binomial negativa, Gama e Poisson, são exemplos de distribuições que pertencem à família exponencial e podem ser usadas no ajuste de GLM que envolve três componentes.

No GLM, a variável aleatória  $\mathbf{Y}$  é associada a uma função de probabilidade (fp) ou uma função densidade de probabilidade (fdp) denominada de componente aleatório. Então, seja  $\mu$  a média da distribuição, dispõe de,

$$E(\mathbf{Y}) = \mu$$

O componente sistemático é formado pelas variáveis explicativas ( $\mathbf{X}$ ) e os parâmetros do modelo ( $\beta$ ). Esse componente produz o denominado preditor linear ( $\eta$ ), e  $\eta$  é um vetor com dimensão  $n$ , então, dispõe de,

$$\eta = \mathbf{X}\beta.$$

A denominada função de ligação é o terceiro componente de um GLM, cujo objetivo é relacionar a média ao preditor linear, estabelecendo uma relação entre o componente aleatório e o componente sistemático do modelo. Assim,

$$(g(\mu) = \eta),$$

sendo as funções de ligação ( $g(\cdot)$ ) e devem ser diferenciável e monótona, determinadas segundo Demétrio (2001) como,

$$g(E[\mathbf{Y}]) = g(\mu) = \eta = \mathbf{X}\beta.$$

### 3.2.1 Modelos aditivos generalizados (GAM)

Embora flexibilizem o pressuposto de normalidade dos MRL, os GLM ainda revelam limitações dado que, julgam que a relação da média é linear às variáveis explicativas. Outra alternativa corresponde ao 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 Aditive 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 ( $\boldsymbol{\eta}$ ), que geralmente, decorre de maneira não-linear. Os GAM permitem casualmente, solucionar questões dos resíduos assimétricos e/ou heterocedásticos. O GAM é expresso como:

$$\mathbf{Y} \stackrel{ind}{\sim} \mathcal{FE}(\boldsymbol{\mu}, \boldsymbol{\phi})$$

$$\boldsymbol{\eta} = g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + s_1(\mathbf{x}_1) + \dots + s_J(\mathbf{x}_J),$$

em que,  $s_j$  é uma função de suavização não-paramétrica aplicada a covariável  $\mathbf{x}_j$  para  $j = 1, \dots, J$  (STASINOPOULOS et al., 2017). Nota-se que não são todas as covariáveis que precisam de funções de suavização, pois as mesmas têm que ser contínuas caso não sejam, não se aplica.

Então, permite-se mencionar algumas funções de suavização mais comuns são: *P-splines*, *splines* cúbicas, *loess* e redes neurais. 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; MARX; DURBÁN, 2015).

Ao aplicar o GAM é necessário ter atenção ao realizar a análise do modelo, pois os erros padrões são inválidos na medida que só é possível serem calculados com base na própria covariável, pressupondo uma relação linear, não se importando com o efeito das funções de suavização como um todo. O ajuste de modelos semi-paramétricos tem como um dos seus aspectos não ser facilmente descritos em uma forma matemática, porém, eles se permitem ser expostos de forma gráfica, que ao mesmo tempo pode nos remeter várias informações sobre a adequação do ajuste e resíduos.

### 3.2.2 Modelos Aditivos Generalizados para Localização, Escala e Forma (GAMLSS)

Técnicas de suavização começaram a ficar populares nos anos 1980. Hastie e Tibshirani (1990) introduziram dentro da abordagem das classes de modelos GLM os

parâmetros de suavização talhando o termo GAM. Nos anos 2000 o modelo GAM se popularizou com a implementação no software R por meio do pacote mcgv (WOOD, 2006). Porém, nem todas as limitações foram resolvidas com advento dos modelos GLM e GAM. Neste sentido, Rigby e Stasinopoulos (2005) propuseram a classe modelos de regressão denominada GAMLSS - Modelos Aditivos Generalizados para Localização, Escala e Forma com a vantagem de modelar os parâmetros de assimetria e curtose, possibilitando uma maior flexibilidade e uma expansão ao número de distribuições consideradas, pois agora as distribuições não precisam ser parte da família exponencial.

Admitindo o ajuste de modelos que a distribuição da variável resposta não necessariamente segue uma distribuição pertencente à família exponencial, propôs-se a substituição por uma família de distribuição geral. No GAMLSS, a parte sistemática do modelo é expandida para permitir modelar não apenas a média, mas todos os parâmetros de localização, escala e forma da distribuição da variável resposta. De modo, que todos esses parâmetros permitem ser modelados em função das variáveis explicativas e, ainda, os preditores também podem inserir funções não-paramétricas de suavização, efeitos aleatórios, ou outros termos aditivos, ou seja, os GAMLSS são modelos de regressão semi-paramétricos. Paramétricos, por requererem a suposição de distribuição paramétrica para a variável resposta, e 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. Esta nova categoria de modelos, os GAMLSS dispõem dos modelos MRL, GLM e GAM como casos especiais. Entretanto estes modelos ainda adotam as observações da variável resposta  $Y$  sejam independentes entre si.

Um modelo GAMLSS é descrito por,

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

$$\begin{aligned} \boldsymbol{\eta}_1 &= g_1(\boldsymbol{\mu}) = \mathbf{X}_1\boldsymbol{\beta}_1 + s_{11}(\mathbf{x}_{11}) + \dots + s_{1J_1}(\mathbf{x}_{1J_1}) \\ \boldsymbol{\eta}_2 &= g_2(\boldsymbol{\sigma}) = \mathbf{X}_2\boldsymbol{\beta}_2 + s_{21}(\mathbf{x}_{21}) + \dots + s_{2J_2}(\mathbf{x}_{2J_2}) \\ \boldsymbol{\eta}_3 &= g_3(\boldsymbol{\nu}) = \mathbf{X}_3\boldsymbol{\beta}_3 + s_{31}(\mathbf{x}_{31}) + \dots + s_{3J_3}(\mathbf{x}_{3J_3}) \\ \boldsymbol{\eta}_4 &= g_4(\boldsymbol{\tau}) = \mathbf{X}_4\boldsymbol{\beta}_4 + s_{41}(\mathbf{x}_{41}) + \dots + s_{4J_4}(\mathbf{x}_{4J_4}) \end{aligned} \tag{3.4}$$

em que,  $D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau})$  é uma distribuição de quatro parâmetros,  $\boldsymbol{\mu}$  normalmente é um parâmetro de localização,  $\boldsymbol{\sigma}$  é constantemente, um parâmetro de escala,  $\boldsymbol{\nu}$  e  $\boldsymbol{\tau}$  são os parâmetros de forma da distribuição, comumente associados à assimetria e curtose, respectivamente.  $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$  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).

Na função 3.4 que descreve o modelo GAMLSS com a máxima quantidade de parâmetros implementados no *software* R (R Core Team, 2018), que de modo teórico, ainda é possível admitir uma distribuição de probabilidade para a variável resposta com mais de quatro parâmetros, inserindo mais preditores no modelo. Além disso, nos permite definir modelos intermediários para distribuições com dois ou três parâmetros e ajustar  $\sigma$  e/ou  $\nu$  em função de covariáveis, considerando diminuir a quantidade de preditores.

Esta configuração da distribuição  $D(\mu, \sigma, \nu, \tau)$  é geral e somente implica que a distribuição deve estar na configuração paramétrica. O pacote *gamlls* (STASINOPOULOS; RIGBY, 2007) implementado exclusivamente no *software* R, apresenta cerca de 200 distribuições discretas, contínuas e mistas implementadas na família GAMLSS, integrando algumas distribuições exclusivamente assimétricas, platicúrticas ou leptocúrticas. Também permite implementar uma nova distribuição à critério do pesquisador.

As distribuições implementadas neste pacote permitem ser truncadas à direita, à esquerda ou ambos os lados truncados; censuradas em qualquer intervalo de resposta; misturadas criando distribuições de mistura; inflacionadas em zero e/ou um. Em relação às distribuições contínuas, é possível serem discretizadas, possibilitando serem criadas para modelar variáveis de resposta discreta. Também, as distribuições contínuas no intervalo  $(-\infty, \infty)$  podem ser transformadas em distribuições com intervalos de  $(0, \infty)$  e  $(0, 1)$  (RIGBY et al., 2017).

Vale enfatizar, que para cada parâmetro pode ser atribuído uma função de ligação diferente. A função de ligação aplicada ao preditor, como propôs Rigby e Stasinopoulos (2005) está associada à amplitude dos valores dos parâmetros, o que difere de como é definido no GLM, pois a função de ligação apresenta-se mais relacionada à distribuição selecionada para a variável resposta. Em se tratando das mesmas funções de ligação, em GAMLSS, a ligação identidade é escolhida quando o parâmetro apresenta valores entre  $(-\infty, \infty)$ , logarítmica para  $(0, \infty)$  e logística para  $(0, 1)$ .

As variáveis explicativas conseguem atingir os parâmetros da distribuição de diversas maneiras. Os GAMLSS permitem ajustes por meio de funções paramétricas lineares ou não lineares, e funções não-paramétricas. O pacote *gamlls* permite suavizar termos aditivos: *P-splines* (*B-splines* penalizadas); *P-splines* monótonas; *P-spline* cíclica; *P-spline* de coeficientes variantes; *splines* cúbicas; *loess*; polinômios fracionários; efeitos aleatórios; regressão *ridge* e ajustes não-paramétricos. E unido a outros pacotes, possibilita, de modo simultâneo, o ajuste de árvores de decisão, efeitos aleatórios, redes neurais e suavizações multidimensionais (STASINOPOULOS et al., 2017).

### 3.2.2.1 Estimação

Os modelos GAMLSS preliminarmente determinados na Equação 3.4 em que  $D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau})$  é a distribuição da variável de resposta  $\mathbf{Y}$ ,  $\mathbf{X}_k$  são as matrizes que incluem os termos aditivos lineares no modelo,  $\beta_k$  são os parâmetros lineares e  $s_{kj}(x_{kj})$  representam diferentes funções de suavização para diferentes variáveis explicativas  $x_{kj}$  para  $k = 1, 2, 3, 4$  e  $j = 1, \dots, J_k$ . Considere-se que a maioria das suavizações podem ser escritas na forma (STASINOPOULOS et al., 2017),

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

em que, a matriz  $\mathbf{Z}$  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  $\mathbf{G} = \mathbf{D}^T\mathbf{D}$  em que  $\mathbf{D}$  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.

Desta maneira o modelo 3.4 pode ser generalizado e reescrito como,

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

$$\begin{aligned} \boldsymbol{\eta}_1 &= g_1(\boldsymbol{\mu}) = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{Z}_{11}\boldsymbol{\gamma}_{11} + \dots + \mathbf{Z}_{1J_1}\boldsymbol{\gamma}_{1J_1} \\ \boldsymbol{\eta}_2 &= g_2(\boldsymbol{\sigma}) = \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{Z}_{21}\boldsymbol{\gamma}_{21} + \dots + \mathbf{Z}_{2J_2}\boldsymbol{\gamma}_{2J_2} \\ \boldsymbol{\eta}_3 &= g_3(\boldsymbol{\nu}) = \mathbf{X}_3\boldsymbol{\beta}_3 + \mathbf{Z}_{31}\boldsymbol{\gamma}_{31} + \dots + \mathbf{Z}_{3J_3}\boldsymbol{\gamma}_{3J_3} \\ \boldsymbol{\eta}_4 &= g_4(\boldsymbol{\tau}) = \mathbf{X}_4\boldsymbol{\beta}_4 + \mathbf{Z}_{41}\boldsymbol{\gamma}_{41} + \dots + \mathbf{Z}_{4J_4}\boldsymbol{\gamma}_{4J_4} \end{aligned} \quad (3.5)$$

Refere-se como modelos GAMLSS semi-paramétrico o 3.5, os parâmetros de efeitos fixos são expressos por

$$\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \boldsymbol{\beta}_3^T, \boldsymbol{\beta}_4^T)^T,$$

e Stasinopoulos et al. (2017) definem,

$$\boldsymbol{\gamma} = (\boldsymbol{\gamma}_{11}^T, \dots, \boldsymbol{\gamma}_{1J_1}^T, \boldsymbol{\gamma}_{21}^T, \dots, \boldsymbol{\gamma}_{4J_4}^T)^T,$$

conforme os parâmetros de um efeito aleatório assumindo que todos os  $\boldsymbol{\gamma}_{kj}$  são independentes entre si e que,

$$\boldsymbol{\gamma}_{kj} \sim N\left(0, [\mathbf{G}_{kj}(\boldsymbol{\lambda}_{kj})]^{-1}\right),$$

consiste em  $[\mathbf{G}_{kj}(\boldsymbol{\lambda}_{kj})]^{-1}$  a inversa (generalizada) de ordem  $q_{kj} \times q_{kj}$  da matriz simétrica  $\mathbf{G}_{kj}(\boldsymbol{\lambda}_{kj})$ .

Se no modelo 3.5 não existem os efeitos aleatórios, logo o modelo resume-se à

$$\begin{aligned}
\mathbf{Y} &\overset{ind}{\sim} D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau}) \\
\boldsymbol{\eta}_1 &= g_1(\boldsymbol{\mu}) = \mathbf{X}_1\boldsymbol{\beta}_1 \\
\boldsymbol{\eta}_2 &= g_2(\boldsymbol{\sigma}) = \mathbf{X}_2\boldsymbol{\beta}_2 \\
\boldsymbol{\eta}_3 &= g_3(\boldsymbol{\nu}) = \mathbf{X}_3\boldsymbol{\beta}_3 \\
\boldsymbol{\eta}_4 &= g_4(\boldsymbol{\tau}) = \mathbf{X}_4\boldsymbol{\beta}_4.
\end{aligned} \tag{3.6}$$

Stasinopoulos et al. (2017) apontam 3.6 sendo modelo GAMLSS paramétrico e 3.5 sendo modelo GAMLSS com efeito aleatório. Quando um GAMLSS é paramétrico, isto é, que não existem funções de suavizações (efeitos aleatórios determinados anteriormente), exige só a estimação de  $\boldsymbol{\beta}$ . Em se tratando de um modelo GAMLSS com efeitos aleatórios exige não apenas o  $\boldsymbol{\beta}$ , mas  $\boldsymbol{\gamma}$  e  $\boldsymbol{\lambda}$  determinado por,

$$\boldsymbol{\lambda} = \left( \boldsymbol{\lambda}_{11}^T, \dots, \boldsymbol{\lambda}_{1J_1}^T, \boldsymbol{\lambda}_{21}^T, \dots, \boldsymbol{\lambda}_{4J_4}^T \right)^T.$$

Os GAMLSS paramétricos 3.6 pode ser entendido também no contexto dos modelos de delineamento experimentais (FISHER, 1937; MONTGOMERY, 2017), no qual modela-se o parâmetro de locação, porém com o ganho de poder modelar o parâmetro de escala, assimetria e curtose. Os modelos GAMLSS paramétricos são ajustados pelas estimativas de máxima verossimilhança em relação à estimação de  $\boldsymbol{\beta}$ . Expressa que  $\mathbf{Y} \overset{ind}{\sim} D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau})$  implica que o logaritmo da função de verossimilhança, determinada pela verossimilhança observada da amostra é,

$$l = \sum_{i=1}^n \ln [f(y_i | \mu_i, \sigma_i, \nu_i, \tau_i)] \tag{3.7}$$

de acordo com o pressuposto de independência das observações.

Enquanto o modelo GAMLSS não 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. Desta maneira se dá a função de verossimilhança penalizada que é,

$$l_p = l - \frac{1}{2} \sum_{k=1}^4 \sum_{j=1}^{J_k} \boldsymbol{\gamma}_{kj}^T \mathbf{G}_{kj}(\boldsymbol{\lambda}_{kj}) \boldsymbol{\gamma}_{kj}. \tag{3.8}$$

Para ajustar um modelo GAMLSS, recomenda-se dois algoritmos iterativos de estimação, quando se trata de valores fixos de hiperparâmetros, a fim de maximizar a função de verossimilhança penalizada (RIGBY; STASINOPOULOS, 2005).

O algoritmo CG é uma generalização do algoritmo de Cole e 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  $\boldsymbol{\theta} = (\mu, \sigma, \nu, \tau)$  com quatro parâmetros. Porém, para muitas funções de densidade de probabilidade  $f(\gamma|\boldsymbol{\theta})$ , os parâmetros  $\boldsymbol{\theta}$  dispõem de informação ortogonal, isto quer dizer que, os valores das derivadas cruzadas da função de verossimilhança são iguais à zero. Nesta situação, aplica-se o algoritmo mais simples e que não utiliza o valor esperado das derivadas cruzadas, chamado de RS, que é uma generalização do algoritmo proposto por Rigby e Stasinopoulos (2005). Existe ainda uma terceira metodologia, que agrega passos dos dois algoritmos, começando o processo por RS e concluindo por CG (ROQUIM, 2018).

O objetivo do algoritmo RS é maximizar a verossimilhança em cada parâmetro  $(\mu, \sigma, \nu, \tau)$  até alcançar a convergência. E se tratando do algoritmo CG, a partir das derivadas cruzadas desde que não sejam nulas, irá atualizar os parâmetros em conjunto com cada iteração. Esta é a divergência fundamental entre eles. Ressalta-se também, que em geral, o algoritmo RS é mais estável e mais rápido, quando comparado ao CG (STASINOPOULOS et al., 2017).

Algumas vantagens relevantes dos algoritmos: (i) possibilita diferentes modelos de diagnóstico para cada parâmetro da distribuição, indica permitir o procedimento de ajuste modular; (ii) favorece a adição de termos aditivos extra e de distribuições extras, e (iii) comumente, são rápidos e estáveis (RIGBY; STASINOPOULOS, 2005; FLORENCIO, 2010).

De acordo com o trabalho de Rigby e Stasinopoulos (2005), os algoritmos induzem ao máximo da função de verossimilhança penalizada, proporcionando estimadores dos parâmetros de efeitos fixos e aleatórios,  $\hat{\beta}$  e  $\hat{\gamma}$  para  $\boldsymbol{\lambda}$  constante. Até agora os algoritmos RS e CG, para estimar  $\alpha$ ,  $\beta$  e  $\gamma$  dados os hiperparâmetros  $\boldsymbol{\lambda}$  constante, e assim, alcançar o máximo da função de verossimilhança penalizada. Mas usualmente, pretende-se estimar o grau de suavização.

Existem diversos métodos para estimar  $\boldsymbol{\lambda}$  e um dos primeiros métodos são os nomeados de globais, pois é aplicado fora das iterações dos algoritmos RS e CG, constituem em aplicar determinado procedimento baseado em otimização numérica no sentido de minimizar o critério de Akaike e desta forma, escolher a melhor estimativa para  $\boldsymbol{\lambda}$  ou aplicar a minimização do desvio global de um banco de dados de validação para conseguir os parâmetros de suavização. Ainda assim, esses métodos determinam que diversos modelos fossem reajustados, o que possivelmente seria um processo lento computacional principalmente quando o modelo dispor de diversas funções de suavização (RIGBY; STASINOPOULOS, 2014).

Rigby e Stasinopoulos (2014) sugeriram forma de selecionar este parâmetro de



forma automática, que denominam de estimação interna ou local, pelo fato do método ser aplicado em cada passo do algoritmo RS e CG, fazendo o uso das teorias de máxima verossimilhança. Esta opção tem como fundamental vantagem à rapidez em comparação às demais e seu desempenho é bastante semelhante às técnicas anteriores, quando for utilizado o modelo 3.6, estas considerações não se fazem necessárias.

### 3.2.2.2 Inferência

Em se tratando de modelos paramétricos, a estimação dos parâmetros 3.6 é realizada via máxima verossimilhança, já quando se trata de modelos GAMLSS semi-paramétricos 3.5, junto a termos de suavização, é aplicada a máxima verossimilhança penalizada. Referindo-se a modelos paramétricos, é importante que o modelo tenha um bom ajuste, para que via o método da máxima verossimilhança, se disponha assintoticamente, que todos os estimadores dos parâmetros sejam consistentes, acompanhados de erros padrões assintóticos corretos, conduzindo assim à cobertura correta dos intervalos de confiança. Porém, caso o modelo GAMLSS não seja adequado, isto é, desfavorável aos dados, logo os parâmetros podem não ser consistentes, levando a inferências errôneas (STASINOPOULOS et al., 2017).

Inferências acerca dos modelos de regressão paramétricos, como um GLM, normalmente selecionam ou testam diferentes modelos, estimam os parâmetros, testam valores de um ou mais parâmetros, estimam intervalos de confiança para tais parâmetros e se faz predições de valores futuros da variável resposta. Nos modelos com funções de suavização como um GAM é fundamental a existência de ferramentas extras para realizar inferências que admitem obter conclusões sobre os parâmetros e o comportamento das funções de suavização. Para os GAMLSS faz-se necessário utilizar ferramentas que sejam capazes de lidar com a diversidade de distribuições e termos no modelo, e ainda de diferentes conjuntos de variáveis explicativas e função de ligação que cada parâmetro do modelo pode admitir.

A respeito dos GAMLSS, não é obrigatório que todos os parâmetros da distribuição dependam de covariáveis, isto é, observa-se algum parâmetro constante. Em tais casos é possível obter erros padrões aproximados e, em consequência, intervalos de confiança para tal parâmetro. Para esta situação sobre os parâmetros serem ajustados em função de variáveis explicativas, permite-se obter os erros padrões e intervalos de confiança para cada coeficiente do preditor linear, desde que não sejam aplicadas funções de suavização naquela covariável. Refere-se a fatores, usualmente, não há interesse em analisar os coeficientes, mas sim a contribuição daquele fator no modelo (STASINOPOULOS et al., 2017).

No modelo que contém componentes não-paramétricos, é importante conhecer a

forma das curvas ajustadas para o termo de suavização e seus respectivos erros padrões. A estimação do parâmetro de suavização ( $\lambda$ ) é de total relevância porque indicará o formato da curva de suavização. Não é comum o interesse em erros padrões deste parâmetro, mas, é possível que eles sejam aproximados por técnicas de *bootstrapping* (ROQUIM, 2018).

Quando existe o interesse em prever um valor futuro para qualquer parâmetro da distribuição da variável resposta, seja ele  $\mu$ ,  $\sigma$ ,  $\nu$  ou  $\tau$  ou ainda prever a própria distribuição ajustada da variável resposta como um todo, o que é possível de ser obtida ao se substituir na distribuição todos os parâmetros preditos. É importante destacar que, para qualquer distribuição resultante não se considera as incertezas dos parâmetros preditos (STASINOPOULOS et al., 2017).

Stasinopoulos et al. (2017) atribuem os métodos inferenciais utilizados para resolver as questões citadas anteriormente sobre classe da inferência baseada em verossimilhança e a classe em *bootstrapping*, os métodos com base em verossimilhança. A princípio um GAMLSS paramétrico e  $\theta$  o vetor genérico de parâmetros. Diante disso, surge a ideia de um grupo de todos os coeficientes lineares para  $\mu$ ,  $\sigma$ ,  $\nu$  e  $\tau$ , por exemplo,  $(\beta_1, \beta_2, \beta_3, \beta_4)$ . A teoria de máxima verossimilhança clássica considerada, temos que, assintoticamente,

$$\hat{\theta} \sim N(\theta_T, i(\theta_T)^{-1}),$$

em que  $\hat{\theta}$  é o estimador de máxima verossimilhança e

$$i(\theta_T) = -E \left[ \frac{\partial^2 l(\theta)}{\partial \theta \partial \theta^T} \right]_{\theta_T},$$

sendo,  $(i(\theta_T))$  a matriz de informação esperada de Fisher tomando valores verdadeiros de  $\theta_T$ . As vezes não é possível derivar a matriz de informação esperada analiticamente. Logo, utiliza a de informação observada de Fisher, que é a seguinte

$$I(\theta_T) = - \left[ \frac{\partial^2 l(\theta)}{\partial \theta \partial \theta^T} \right]_{\theta_T}.$$

Nota-se que  $I(\theta_T)$  é equivalente ao negativo da matriz hessiana do logaritmo da verossimilhança em  $\theta_T$ . A matriz de variâncias e covariâncias da distribuição assintótica de  $\hat{\theta}$  é aproximado por  $I(\theta_T)^{-1}$  ao invés de  $i(\theta_T)^{-1}$ . O  $\theta_T$  substitui por  $\hat{\theta}$  quando  $\theta_T$  for desconhecido, recorrendo às matrizes de informação esperada e observada, dadas por  $i(\hat{\theta})$  e  $I(\hat{\theta})$ , respectivamente (STASINOPOULOS et al., 2017).

Portanto, para GAMLSS paramétricos, distribuição assintótica para  $\hat{\theta}$  é a seguinte

$$\hat{\theta} \sim N\left(\theta_T, \mathbf{I}(\hat{\theta})^{-1}\right).$$

Os erros padrões dos parâmetros estimados são obtidos em geral pela raiz quadrada da diagonal da matriz de variâncias e covariâncias associada à  $\theta$ . Para obter os erros padrões ( $EP$ ) de  $\hat{\beta}$ , no caso de a matriz de informação para todos os parâmetros ser de difícil obtenção, recomenda-se a solução sugerida por Aitkin et al. (2009):

$$EP(\hat{\beta}) \approx \frac{|\hat{\beta}|}{\sqrt{\Delta GDEV}},$$

em que  $\Delta GDEV$  é a diferença no desvio global obtida pela omissão das variáveis explicativas associadas ao coeficiente do modelo o  $\beta$ . Este procedimento é justificado por aproximações da estatística de teste de razão de verossimilhanças à estatística do teste de Wald, sob a hipótese nula de que  $\beta = 0$ , em que (STASINOPOULOS et al., 2017)

$$\left[ \frac{\hat{\beta}}{EP(\hat{\beta})} \right]^2 \approx \Delta GDEV.$$

### 3.2.2.3 Tipos de distribuições dentro da família GAMLSS

Dentre as mais de 200 distribuições existentes na classe GAMLSS e implementadas no pacote gamlss do R. O tipo de distribuição para ser utilizada depende da variável resposta que está sendo modelada. Dentro da família GAMLSS há três tipos distintos de família de distribuições:

1. Distribuições Contínuas;
2. Distribuições Discretas;
3. Mistura de Distribuições.

Maiores informações, bem como uma tabela com os nomes das distribuições, espaço paramétrico e domínio das funções, respectivas funções de ligação e nome da distribuição na implementação do R, podem ser vistos em Stasinopoulos et al. (2017) páginas 148 - 151.

### 3.2.2.4 Critério de seleção dos modelos

Os critérios de seleção de modelos de regressão tem o intuito de solucionar o problema da seleção de preditores apropriados dentre um grande número de possíveis preditores. A seleção de modelos em GAMLSS inclui a seleção da melhor distribuição

para a variável resposta, dos preditores adequados para os parâmetros da distribuição selecionada, das funções de ligação e dos hiperparâmetros.

O critério de Akaike generalizado (GAIC) e este 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; FREIRE; CECATTI, 2008). Ele é utilizado no mesmo contexto do critério de informação de Akaike (AIC) (AKAIKE, 1998). O GAIC é definido por Voudouris et al. (2012), como:

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

em que  $l$  é o logaritmo da função verossimilhança e  $gl$  são os graus efetivos de liberdade do modelo ajustado,  $k$  é constante e  $e$  é a penalidade para cada grau de liberdade utilizado. Refere-se a  $-2l(\hat{\theta})$ , como desvio global, pois o  $GAIC(k)$  é a estatística obtida pela adição do desvio global.

O desvio global (GDEV), importante medida para a seleção de modelos em GAMLSS, tal define-se como,

$$GDEV = -2l(\hat{\theta}),$$

em que,  $l(\hat{\theta})$  é o logaritmo da função de verossimilhança ajustada, mostrado nas Equações 3.7 e 3.8.

O modelo com o menor valor do critério  $GAIC(k)$ , para algum  $k$  escolhido é então selecionado e considerado melhor ajuste, pois o  $GAIC(k)$  penaliza modelos com muitos parâmetros. O critério (AIC) (AKAIKE, 1998) e o critério de informação bayesiano (BIC) (SCHWARZ, 1978) são casos especiais do  $GAIC(k)$  um caso se  $k = 2$  recai no (AIC) e um outro caso  $k = \ln(n)$  recai no (BIC) (VOUDOURIS et al., 2012).

Segundo Stasinopoulos et al. (2017) afirmam que na busca de um modelo GAMLSS adequado, para qualquer banco de dados, os componentes relacionados à distribuição, à função de ligação, às possíveis variáveis explicativas a serem inseridas nos preditores e aos hiperparâmetros é necessário serem especificados de modo mais objetiva possível.

A seleção da distribuição apropriada é feita 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*.

Na prática a seleção de variáveis explicativas é de extrema importância no ajuste do modelo estatístico. Se  $\chi_k$  é uma coleção de variáveis explicativas acessíveis para

consideração na modelagem do parâmetro  $\theta_k$  de um modelo GAMLSS, em que  $\theta = (\theta_1, \theta_2, \theta_3, \theta_4) = (\mu, \sigma, \nu, \tau)$ . Geralmente,  $\mathbf{x}_k$  incluirá fatores e variáveis quantitativas que podem vir a fazer parte do modelo de forma linear ou por intermédio de funções de suavização (STASINOPOULOS et al., 2017).

### 3.2.2.5 Análise de resíduos

Os resíduos viabilizam extrair informações que permitem diagnosticar ou demonstrar a qualidade do ajuste de um modelo (estágio de diagnóstico), além de averiguar se as suposições do modelo foram satisfeitas. Muitas ferramentas gráficas são utilizadas para identificar incongruências entre o modelo ajustado e as observações coletadas, um equilíbrio entre essas duas fontes de informação é necessário para que o modelo possa ser analisado com algum sentido prático.

A vantagem principal dos resíduos quantílicos (aleatorizados) normalizados é que, para qualquer distribuição da variável resposta, os resíduos verídicos sempre terão uma distribuição normal padrão quando o modelo assumido está correto. A averiguação de pressupostos por meio da normalidade dos resíduos já é bastante estabelecido na literatura estatística. Porém, os resíduos quantílicos (aleatorizados) normalizados fornecem-nos uma maneira comum de verificar a adequação de um modelo (STASINOPOULOS et al., 2017).

A análise de resíduos utiliza os resíduos dos quantis aleatórios normalizados que podem ser utilizados para analisar exclusivamente a distribuição utilizada para ajustar tal modelo, introduzidos (DUNN; SMYTH, 1996), é definido por,

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

em que  $\Phi^{-1}$  denota a inversa da função de distribuição acumulada de uma normal padrão,  $F(\cdot)$  é a função de distribuição acumulada adequada aos dados e  $\theta$  é o vetor de parâmetros. Observe que, um modelo adequado tem seus resíduos  $r_i$  seguindo a distribuição Normal Padrão, para o modelo ser considerado correto.

Uma técnica primordial aplicada para analisar os resíduos de um GAMLSS são os gráficos de minhoca (*worm plot*), o estágio de diagnóstico envolve o uso deste gráfico bem específico, que permite a detecção de inadequações no modelo globalmente ou dentro de um intervalo de uma variável explicativa (BUUREN; FREDRIKS, 2001; VAN BUUREN, 2007). O *worm plot* funciona como o gráfico normal *quantil-quantil* (Q-Q) sem tendência a fim de evidenciar alterações locais no domínio de uma dada covariável.

Uma importante propriedade do *worm plot* está no estabelecimento de intervalos com  $(1 - \alpha)$  de confiança para os *quantis* normais teóricos. Para tal, considerando um *quantil*  $z$  associado a uma probabilidade  $p$  e um tamanho amostral  $n$ , possibilitando

determinar os limites do intervalo baseado na expressão  $\pm z_{\alpha/2} f(z)^{-1} \sqrt{p(1-p)/n}$ , em que  $f(z)$  é a função de densidade de probabilidade da distribuição normal. Um diferencial do GAMLSS é justamente os gráficos *worm plots* por serem uma ferramenta de diagnóstico para análise de resíduos que por ser expresso como um único gráfico, mas que abrange todo o intervalo da variável explicativa ou em diferentes regiões (intervalos).

Os gráficos *worm plots* permitem identificar regiões em que o modelo não é bem ajustado aos dados, seu eixo vertical descreve para cada observação, a diferença entre a sua localização nas distribuições teórica e empírica. Os pontos observados em conjunto, representam uma curva que se assemelha a uma minhoca. Esta forma gráfica aponta como os dados se distanciam da distribuição assumida e, quando tomadas em conjunto, sugere modificações úteis no modelo, como pode ser observado na Tabela 1 as interpretações de diferentes *worm plot*. Logo, se os pontos estão situados no interior da banda de confiança de 95% (entre as duas curvas elípticas), o ajuste do modelo é considerado satisfatório.

Tabela 1 – 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	platicurtose	calda muito pesada

Como pode ser visto na Tabela 1, 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. Graficamente este comportamento pode ser visto na Figura 2.

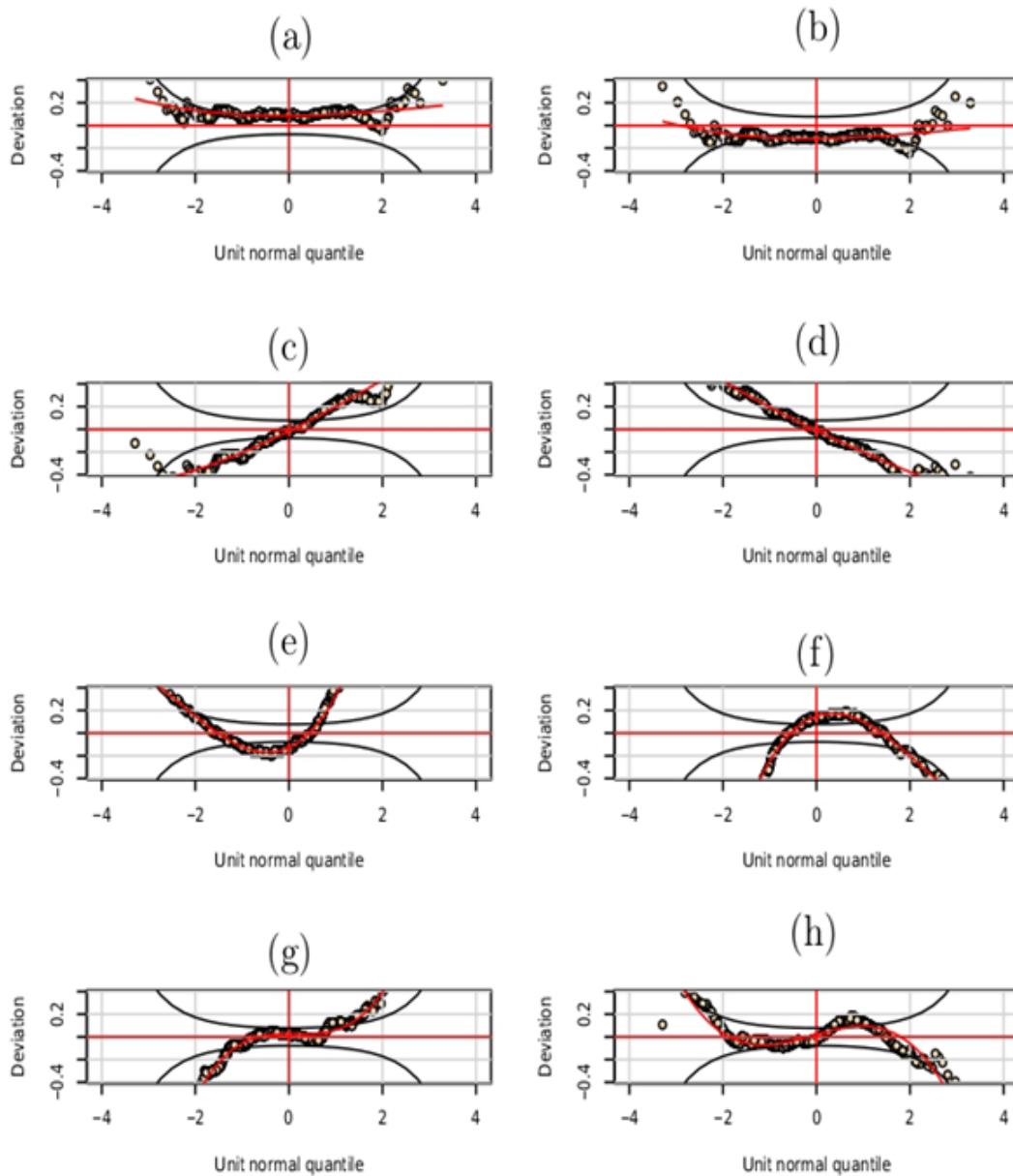


Figura 2 – Diferentes problemas que podem ser diagnosticados pelo *worm plot* (STASINOPOULOS et al., 2017).

Os diagnósticos por meio de gráficos na Figura 2 apontam como os dados se distanciam da distribuição assumida e, quando tomadas em conjunto, sugerem modificações úteis no modelo, a interpretação pode ser combinada com as informações da Tabela 1. O *worm plot* é uma ferramenta bastante útil para verificar adequabilidade e qualidade de ajuste. Desde que a modelagem via GAMLSS foi implementada este gráfico vem sendo utilizado por diversos autores (MACERA, 2011; ROQUIM, 2018).

## 4 Metodologia

Nesta seção está descrito o local da implantação do experimento, a caracterização do experimento via desenho esquemático da parcela experimental e o grupo de dados explorados na pesquisa. E ainda, são apresentados os principais modelos utilizados no ajuste dos dados.

### 4.1 Localização do Experimento

Os dados utilizados no presente trabalho são oriundos de um experimento relativo a clones de *Eucalyptus* spp. que foi instalado na Estação Experimental do Instituto Agrônomo de Pernambuco (IPA), localizado na Chapada do Araripe, território do município de Araripina, com coordenadas geográficas entre 07°27'37"S e 40°24'36"W com altitude de 831 metros e solos definidos como latossolo amarelo + latossolo vermelho-amarelo (GOUVEIA et al., 2015).

O clima é caracterizado como BshW', semiárido, com chuvas de verão, com temperatura média de 25°C, segundo classificação de Köppen. A precipitação média anual é aproximada à 700 mm, com concentração de 70% nos meses de dezembro e março (ROCHA, 2012).

### 4.2 Caracterização do Experimento

O experimento teve sua implantação no início do mês de março de 2002, com tratamentos (clones) dispostos em 60 parcelas com dimensão (14,0 m × 21,0 m) e área de 294 m<sup>2</sup>, espaçamentos de (3,0 m × 2,0 m) contendo 49 plantas em cada parcela e 25 na área útil (Figura 3).

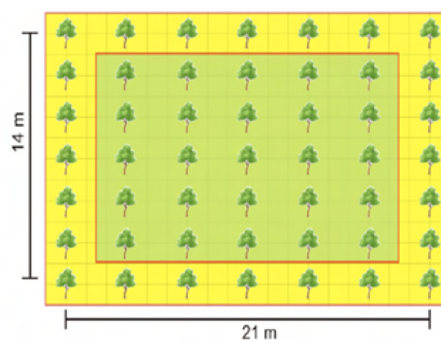


Figura 3 – Desenho esquemático da parcela experimental. Fonte: Barros (2010).



Os dados experimentais tiveram duas rotações, sendo que nas rotações se coletou dados a cada seis meses, a primeira rotação foi finalizada em 2009, com um povoamento de 84 meses. Os dados que foram utilizados neste trabalho, são provenientes da segunda rotação, que se finalizou no primeiro semestre de 2015, com um povoamento de 66 meses.

Tabela 2 – Descrição dos tratamentos de *Eucalyptus* spp. do experimento no Polo Gesseiro do Araripe.

Tratamento	Clone	Descrição
T1	C49	Híbrido de <i>Eucalyptus tereticorn</i> (cruzamento natural)
T2	C80	Híbrido de <i>E. urophylla</i> vs <i>E. tereticorn</i> de Laura (polinização controlada)
T3	C315	Híbrido de <i>E. urophylla</i> vs <i>E. tereticorn</i> de Laura (polinização controlada)
T4	C101	Híbrido de <i>E. urophylla</i> vs <i>E. tereticorn</i> de Laura (polinização controlada)
T5	C78	Híbrido de <i>E. urophylla</i> vs <i>E. tereticorn</i> de Laura (polinização controlada)
T6	C156	Híbrido de <i>E. urophylla</i> vs <i>E. tereticorn</i> vs <i>E. pellita</i> de Laura (polinização controlada)
T7	C39	Híbrido de <i>E. urophylla</i> (cruzamento natural)
T8	C27	Híbrido de <i>E. brassiana</i> (cruzamento natural)
T9	C51	Híbrido de <i>E. urophylla</i> vs <i>E. tereticorn</i> de Laura (polinização controlada)
T10	C158	Híbrido de <i>E. urophylla</i> vs <i>E. tereticorn</i> de Laura (polinização controlada)
T11	C41	Híbrido de <i>E. urophylla</i> (cruzamento natural)
T12	C31	Híbrido de <i>E. brassiana</i> (cruzamento natural)
T13	C25	Híbrido de <i>E. brassiana</i> (cruzamento natural)
T14	C33	Híbrido de <i>E. urophylla</i> (cruzamento natural)
T15	C11	Híbrido de <i>E. brassiana</i> (cruzamento natural)

As mudas dos clones C80, C315, C101, C78, C51, C158 foram produzidas com material genético de procedência da região de Laura, Queensland, Austrália.

O software utilizado para as análises estatísticas foi o R (versão 3.5.1) e para tabulação foi feita via planilha eletrônica Excel (MICROSOFT OFFICE, 2010). As variáveis coletadas foram: altura total ( $Ht$ ), diâmetro a altura do peito ( $DAP = CAP/\pi$ ) em que a circunferência a altura do peito ( $CAP$ ) igual ou superior a 6 cm e Volume ( $Vol\ m^3$ ) de todas as plantas úteis das parcelas. O volume é medido com o auxílio da ferramenta suta com a qual foi medido a DAP por 1,30 m em relação ao solo. As árvores foram derrubadas, mediu-se a altura de corte 30 cm e a altura comercial. A estimação do volume se baseia na relação entre o volume do cilindro e o volume real da árvore, oriundo da cubagem rigorosa das árvores, pelo o método de Smalian calculou-se o volume para as seções de fuste (SILVA; NETO, 1979) pela seguinte expressão:

$$v_i = \frac{g_1 + g_2}{2} \cdot l,$$

o volume total se dá por:

$$V_i = \sum_{i=1}^n v_i,$$

em que:

$V_i$  = Volume da árvore  $i$  ( $m^3$ );

$v_i$  = Volume da tora  $i$  ( $m^3$ );

$g_1$  = Área basimétrica no início da tora ( $m^2$ );

$g_2$  = Área basimétrica no fim da tora ( $m^2$ );

$l$  = Comprimento da tora ( $m$ ).

Desta forma, o volume do fuste se dá pelo somatório do volume de todas as toras e a área basimétrica calculou-se via a seguinte expressão,

$$g_1 = \frac{\pi \cdot d^2}{4},$$

em que,  $d$  = diâmetro da tora ( $m$ ).

Neste experimento, foram utilizados 15 tratamentos representados por clones de híbridos de eucaliptos com quatro repetições (Tabela 2), instalados em um delineamento inteiramente casualizado (DIC) e se expressa da seguinte forma:

$$Y_{ij} = m + \tau_i + \sigma \varepsilon_{ij},$$

em que,  $Y_{ij}$  é o valor observado do tratamento  $i$ ,  $m$  representa o efeito da média geral,  $\tau_i$  é o efeito do tratamento  $i$  e  $\varepsilon_{ij} \sim ST4(\mu, \sigma, \nu, \tau)$  que configuram os efeitos do fator não controlado do ensaio experimental, com  $i = 1, \dots, I$  e  $j = 1, \dots, J$  e  $I$  representa o número de tratamentos e  $J$  o número de repetições.

### 4.3 Distribuições probabilísticas

A forma geral das distribuições consideradas foi na classe GAMLSS paramétrico 3.6 na qual é apresentada pela seguinte expressão,

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

A função de densidade de probabilidade de uma variável aleatória *Skew t* do tipo 4 (ST4) é caracterizada como uma distribuição de quatro parâmetros, dos quais dois deles modelam assimetria a direita e a esquerda. Tal distribuição é definida como,

$$f_y(y|\mu, \sigma, \nu, \tau) = \frac{c}{\sigma} \left\{ \left[ 1 + \frac{z^2}{\nu} \right]^{-1(\nu+1)/2} I(y < \mu) + \left[ 1 + \frac{z^2}{\tau} \right]^{-(\tau+1)/2} \right\},$$

para  $-\infty < y < \infty$ , em que,  $-\infty < \mu < \infty$ ,  $\sigma > 0$ ,  $\nu > 0$  e  $\tau > 0$ . Com  $z = (y - \mu) / \sigma$  e  $c = 2 \left[ \nu^{1/2} B\left(\frac{1}{2}, \frac{\nu}{2}\right) + \tau^{1/2} B\left(\frac{1}{2}, \frac{\nu}{2}\right) \right]^{-1}$ . Aqui  $E(Y) = \mu + \sigma E(Z)$ , e  $Var(Y) = \sigma^2 V(Z)$ , na qual,

$$E(Z) = c \left[ \frac{1}{\tau - 1} - \frac{1}{\nu - 1} \right]$$

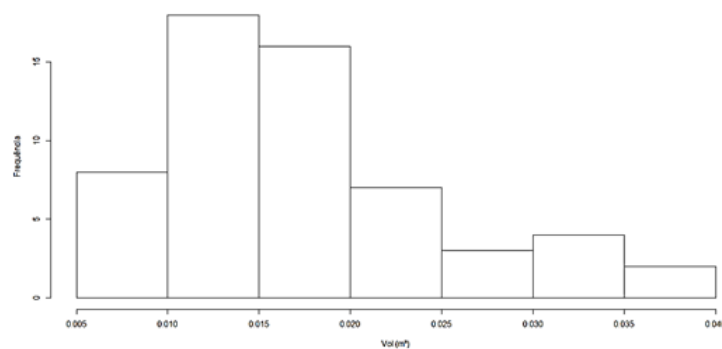
com,

$\nu > 1$  e  $\tau > 1$  e  $E(Z^2) = \frac{c}{2} \left[ \tau^{3/2} B\left(\frac{1}{2}, \frac{\tau}{2}\right) / (\tau - 2) \right] + \left[ \nu^{3/2} B\left(\frac{1}{2}, \frac{\nu}{2}\right) / (\nu - 2) \right]$ , de forma que,  $\nu > 2$  e  $\tau > 2$  (JONES; FADDY, 2003).

Outras distribuições ajustadas foram a Exponencial gaussiana (exGAUS), *Skew* Normal tipo 1 (SN1), *Skew* Normal tipo 2 (SN2), *Skew* exponencial potência tipo IV (SEP4), Sinh-Arcsinh (SHASH) Johnson SU reparação (JSU), Johnson SU original (JSUo), Exponencial beta generalizada tipo II (EGB2), dentre outras pertencentes ou não a família exponencial mas pertencentes a classe GAMLSS. O teste Tukey foi aplicado via pacotes multcomp (HOTHORN; BRETZ; WESTFALL 2008) e BSagri versão 0.1-10. (SCHAARSCHMIDT, 2018), que faz adaptação do teste para os modelos da classe GAMLSS. O critério de GAIC foi aplicado para auxiliar na escolha do melhor modelo, com  $k = 1$ , além de gráficos de diagnósticos de resíduos e *worm plot*. Também foi considerado o teste de Shapiro-Wilks para os resíduos dos modelos GAMLSS bem como estatísticas dos quatro primeiros momentos da distribuição em estudo.

## 5 Resultados e Discussão

Os principais resultados encontrados são apresentados em uma sequência mostrando os procedimentos adotados. Inicialmente foi ajustado um histograma aos dados de volume ( $m^3$ ) de *Eucalyptus* spp., e desta forma foi possível observar a assimetria encontrada nos dados. Este fato, vem a indicar que dificilmente uma distribuição simétrica (distribuição normal, por exemplo) se adequa aos dados (Figura 4).

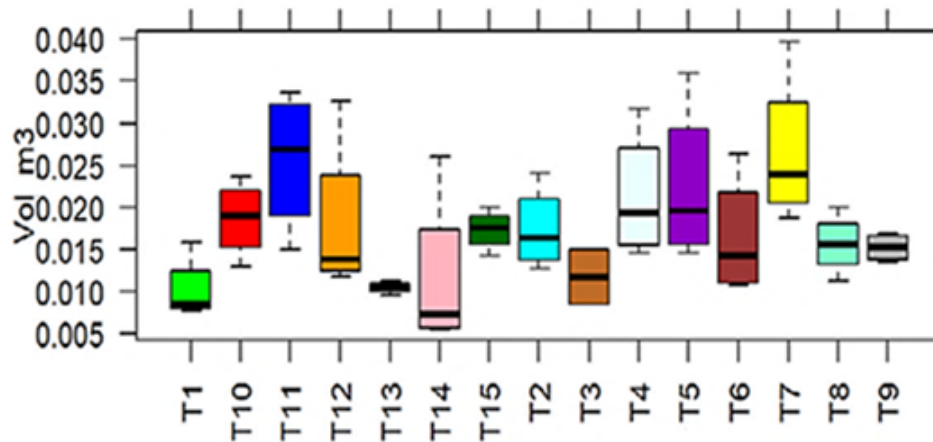


Fonte: Elaborada pelo Autor.

Figura 4 – Histograma para o volume ( $m^3$ ) dos clones de *Eucalyptus* spp..

Por meio do histograma dos dados de volume ( $m^3$ ) de *Eucalyptus* spp., foi possível perceber uma assimetria à direita, esta já era esperada para este tipo de cultura, conforme abordado por (BARTOSZECK et al., 2004; MACERA, 2011) que relatam assimetria nos dados de clones de *Eucalyptus* spp.. E diversas abordagens já foram utilizadas para poder corrigir ou amenizar assimetrias. Fontenele (2016) procedeu com transformação de dados, para variáveis altura da árvore e diâmetro do tronco, bem como outros autores como Hess et al. (2015) utilizaram a distribuição Gama sob a ótica dos GLM, obtendo bons resultados para a altura da planta em *Eucalyptus*.

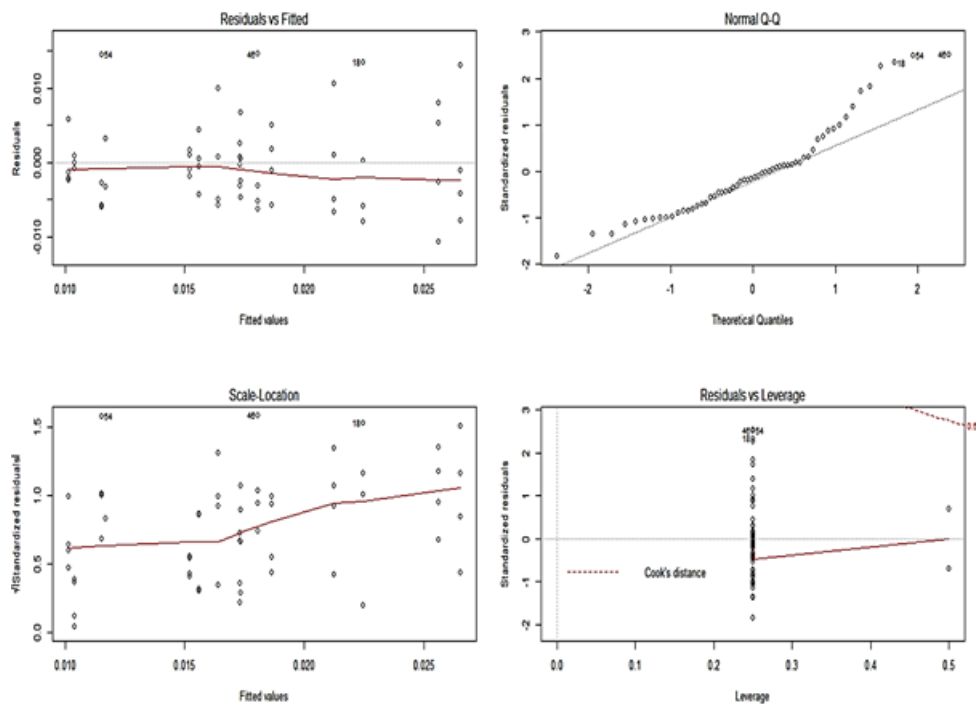
O gráfico de Boxplot é apresentado na Figura 5, este tipo de visualização é útil para o melhor entendimento do comportamento dos clones de *Eucalyptus* spp., pois é possível notar a variabilidade e assimetria em cada clone do experimento, bem como quais clones aparentemente tiveram melhores desempenho no estudo, como por exemplo os clones 39 e 41 de *Eucalyptus. urophylla* (denominado pelo tratamento 7 e tratamento 11), o clone 49 *Eucalytus tereticorn* (denominado pelo tratamento 1) de menor produção. Destacando que neste experimento foi considerado um único espaçamento (3,0 m x 2,0 m) e segundo diversos autores o espaçamento influência de maneira direta na produtividade florestal (COELHO et al., 1970; BERNARDO et al., 1998; BARTON, MONTAGU, 2006; ALCORN, 2007).



Fonte: Elaborada pelo Autor.

Figura 5 – Boxplot para o volume ( $m^3$ ) dos clones de *Eucalyptus* spp..

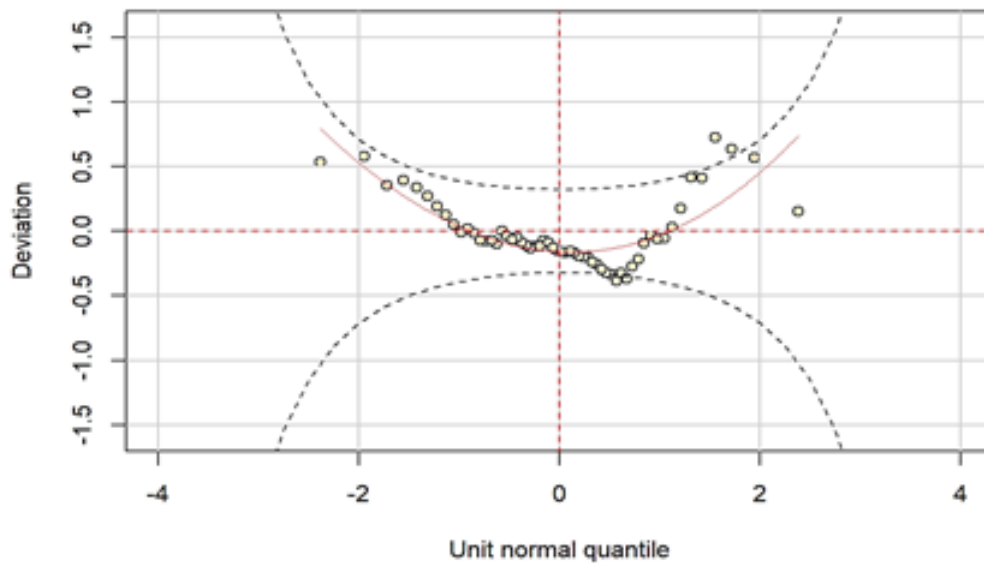
Na Figura 6, apresenta a análise de resíduos para o modelo Normal aplicado aos dados de *Eucalyptus* spp., percebe-se pelo gráfico de quantil quantil que os resíduos não seguem uma distribuição Normal, indicando que a modelagem padrão via análise de variância não se adequa aos dados, fato este corroborado pela Figura 4 e pela análise inferencial para normalidade.



Fonte: Elaborada pelo Autor.

Figura 6 – Análise de resíduos do Modelo Normal para o *Eucalyptus* spp..

A Figura 7 corrobora os resultados encontrados na Figura 6, pois por meio do *worm plot* percebe-se que os dados não estão bem modelados pela distribuição Normal havendo alguns pontos que ultrapassam as bandas de confiança, fato este que sugere falta de ajuste do modelo aos dados. Neste sentido, é possível afirmar por meio da Figura 7 que a curva tem inclinação negativa o que sugere uma variância muito baixa com o parâmetro de escala superestimado (TABELA 1; FIGURA 2d).

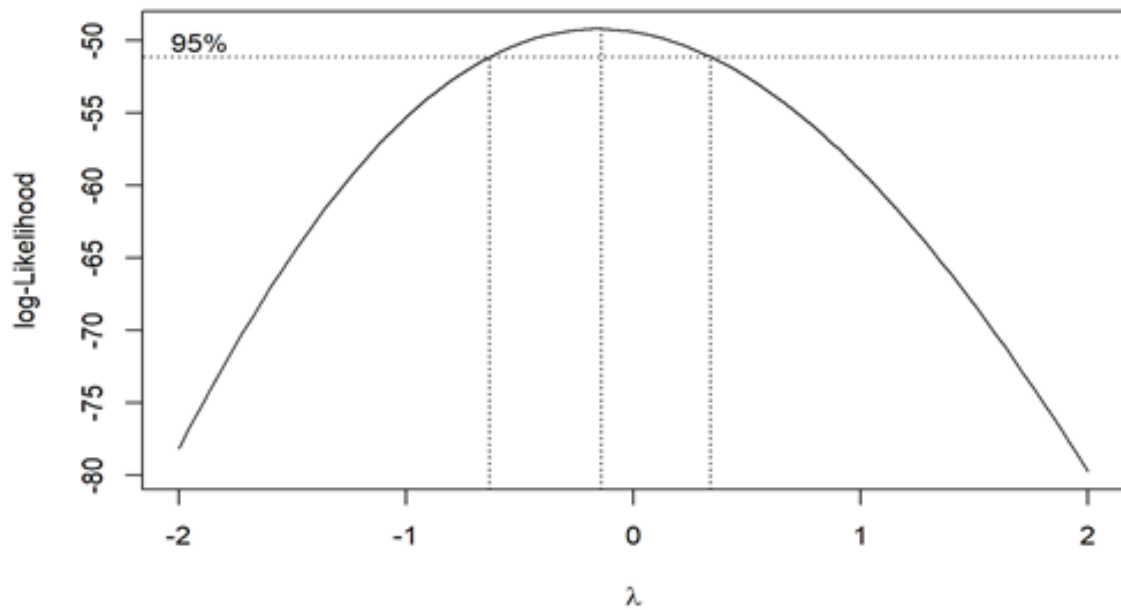


Fonte: Elaborada pelo Autor.

Figura 7 – *Worm plot* do Modelo Normal aplicado aos clones de *Eucalyptus* spp..

Após os resultados gráficos, foi aplicada análise inferencial por meio de testes de hipóteses. De acordo com o teste de Shapiro-Wilks utilizado para análise dos resíduos do modelo sob suposição de normalidade, o resultado obtido foi de um Valor P de 0,001861, que indica a ausência de normalidade dos resíduos, ao nível de 5% de significância. Foi aplicado ainda o teste de homogeneidade de variâncias de Bartlett, o Valor P foi de 0,03068. Estes resultados indicam que os pressupostos da análise de variância não foram atendidas, indicando que outras abordagens devem ser preferidas para que as inferências e a escolha do melhor clone de *Eucalyptus* spp. seja efetiva. Uma delas é a transformação de variável, mas traz como desvantagem o fato de não utilizar a distribuição inerente dos dados, devido a isto a literatura apresenta a metodologia dos modelos lineares generalizados (NELDER; WEDDERBURN, 1972). Estas duas abordagens citadas foram exaustivamente realizadas, mas não se obteve êxito.

Na Figura 8 tem-se a transformação Box-Cox, no qual o intervalo de confiança não abrangeu o valor 1 para  $\lambda$ , por isso pode-se recomendar algumas transformações para os dados, porém nenhuma das transformações tentadas  $(\sqrt{y}, \log(y), 1/\sqrt{y})$  atendeu as pressuposições da modelagem via análise de variância, obtendo Valor P para o teste de Shapiro-Wilks de 0,0107; 0,0254; 0,0255, respectivamente, todos menores que o nível de significância adotado.



Fonte: Elaborada pelo Autor.

Figura 8 – Box-Cox da transformação da variável volume.

Optou-se então por utilizar a abordagem de modelos lineares generalizados aditivos para localização, escala e forma – GAMLSS (RIGBY; STASINOPOULOS, 2005), por esta apresentar grande flexibilidade na modelagem dos dados, sendo uma alternativa recente para ajustar dados de grande complexidade. Na Tabela 3, foi ajustado via o comando *fitDist* do pacote *gamlss*, sem se considerar os efeitos fixos do modelo, sob o modelo nulo.

Dentre os resultados apresentados na Tabela 3, chama-se atenção ao fato de que a distribuição Normal não foi a melhor das candidatas a modelar os dados de Volume ( $m^3$ ), e que a *Skew t* tipo 4 (ST4) foi uma distribuição intermediária quando se leva em consideração os valores do GAIC ordenados do menor valor ao maior. Outro ponto a ser destacado é que a Gumbel reversa (RG) foi a que teve menor valor de GAIC, sendo a melhor das distribuições candidatas.

Tabela 3 – Principais distribuições candidatas a modelar os dados de volume ( $m^3$ ) de *Eucalyptus* spp. sob modelo nulo.

Distribuição	GAIC
Gumbel reversa (RG)	-408,5609
Exponencial gaussiana (exGAUS)	-406,5360
Skew Normal tipo 1 (SN1)	-406,5288
Skew Normal tipo 2 (SN2)	-405,9879
Skew exponencial potência tipo IV (SEP4)	-405,2879
Sinh-Arcsinh (SHASH)	-405,2426
Johnson SU reparação (JSU)	-404,8554
Johnson SU original (JSUo)	-404,8550
Exponencial beta generalizada tipo II (EGB2)	-404,8157
Sinh-Arcsinh original (SHASHo)	-404,7896
Sinh-Arcsinh original tipo II (SHASHo2)	-404,7896
Skew t tipo 5 (ST5)	-404,6035
Skew t tipo 1 (ST1)	-404,5384
Skew t tipo 2 (ST2)	-404,5363
Skew exponencial potência tipo III (SEP3)	-404,2211
Skew t tipo 3 (ST3)	-404,0059
Skew t de Student (SST)	-404,0059
Skew exponencial potência tipo I (SEP1)	-402,4980
Skew t tipo 4 (ST4)	-400,5613
Skew exponencial potência tipo II (SEP2)	-399,2017
Logística (LO)	-398,9670
Power Exponencial 2 (PE2)	-397,8507
Power Exponencial (PE)	-397,8507
Normal (NO)	-397,5509

Fonte: Elaborada pelo Autor.

Os principais modelos probabilísticos com respectivas configurações paramétricas, efeitos fixos de tratamento, adotadas nas distribuições consideradas, em função dos resultados encontrados na Tabela 3, são apresentados na Tabela 4, em que são listados os modelos que obtiveram melhor desempenho, considerando (análise diagnóstica e convergência). O número de modelos (distribuições com combinações de parâmetros  $(\mu, \sigma, \tau, \nu)$  ajustados aos dados de *Eucalyptus* spp., superou duas dezenas.



Tabela 4 – Descrição dos principais modelos ajustados à *Eucalyptus* spp. do experimento no Polo Gesseiro do Araripe, com marcação dos parâmetros considerados ( $\times$ ).

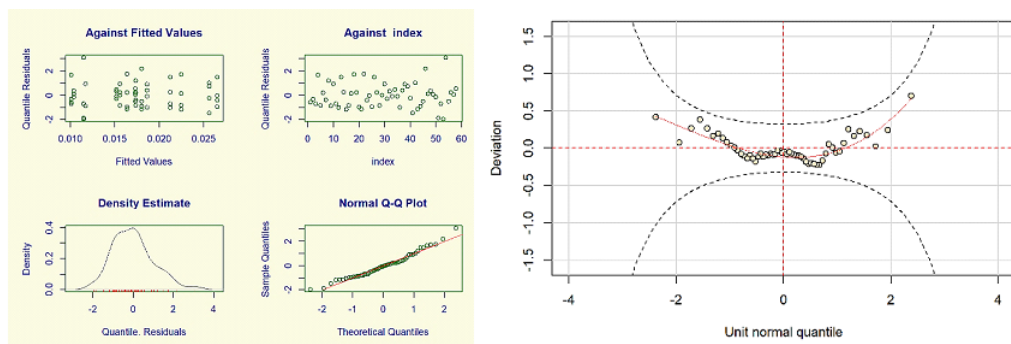
Modelos Ajustados	Parâmetros			
	$\mu$	$\sigma$	$\nu$	$\tau$
M1: Normal	$\times$	-	-	-
M2: Gama	$\times$	-	-	-
M3: Gumbel Reversa	$\times$	-	-	-
M4: Inversa Gaussiana	$\times$	-	-	-
M5: Logístico	$\times$	-	-	-
M6: Gama	$\times$	$\times$	-	-
M7: Inversa Gaussiana	$\times$	$\times$	-	-
M8: Logístico	$\times$	$\times$	-	-
M9: Skew t tipo 4	$\times$	$\times$	$\times^*$	$\times^*$
M10: Skew t tipo 4	$\times$	$\times$	$\times$	$\times^*$
M11: Sinh-Arcsinh original tipo II	$\times$	$\times^*$	$\times^*$	$\times^*$

$\times$  Símbolo que indica a consideração do efeito de Tratamento no parâmetro;

$\times^*$  Apenas Intercepto considerado.

Fonte: Elaborada pelo Autor.

Apesar destas distribuições listadas serem promissoras, optou-se neste trabalho ainda por procurar outras distribuições da classe da família exponencial que não se destacaram na Tabela 3, dentre as distribuições estudadas a que melhor ajuste teve foi a distribuição Gama (M2 e M6) e a distribuição Gumbel Reversa (M3), a análise de resíduos é apresentada na Figura 9 do modelo (M2) e seus resultados de GAIC na Tabela 5. Aplicação do GLM via distribuição Gama, apresentou um resultado aparentemente satisfatório (Figura 9), principalmente no que se refere à análise residual, pois não foram notados grandes desvios, porém como o gráfico de densidade estimada foi não simétrico, e como o valor obtido no critério de GAIC foi inferior a alguns modelos ajustados via distribuições GAMLSS, Tabela 5, além de que desvios nos coeficientes de assimetria e curtose, estimados para esta distribuição serem acentuados, estes motivos levaram a descartar esta abordagem.



Fonte: Elaborada pelo Autor.

Figura 9 – Análise de resíduos e *worm plot* para o modelo Gama (M2).

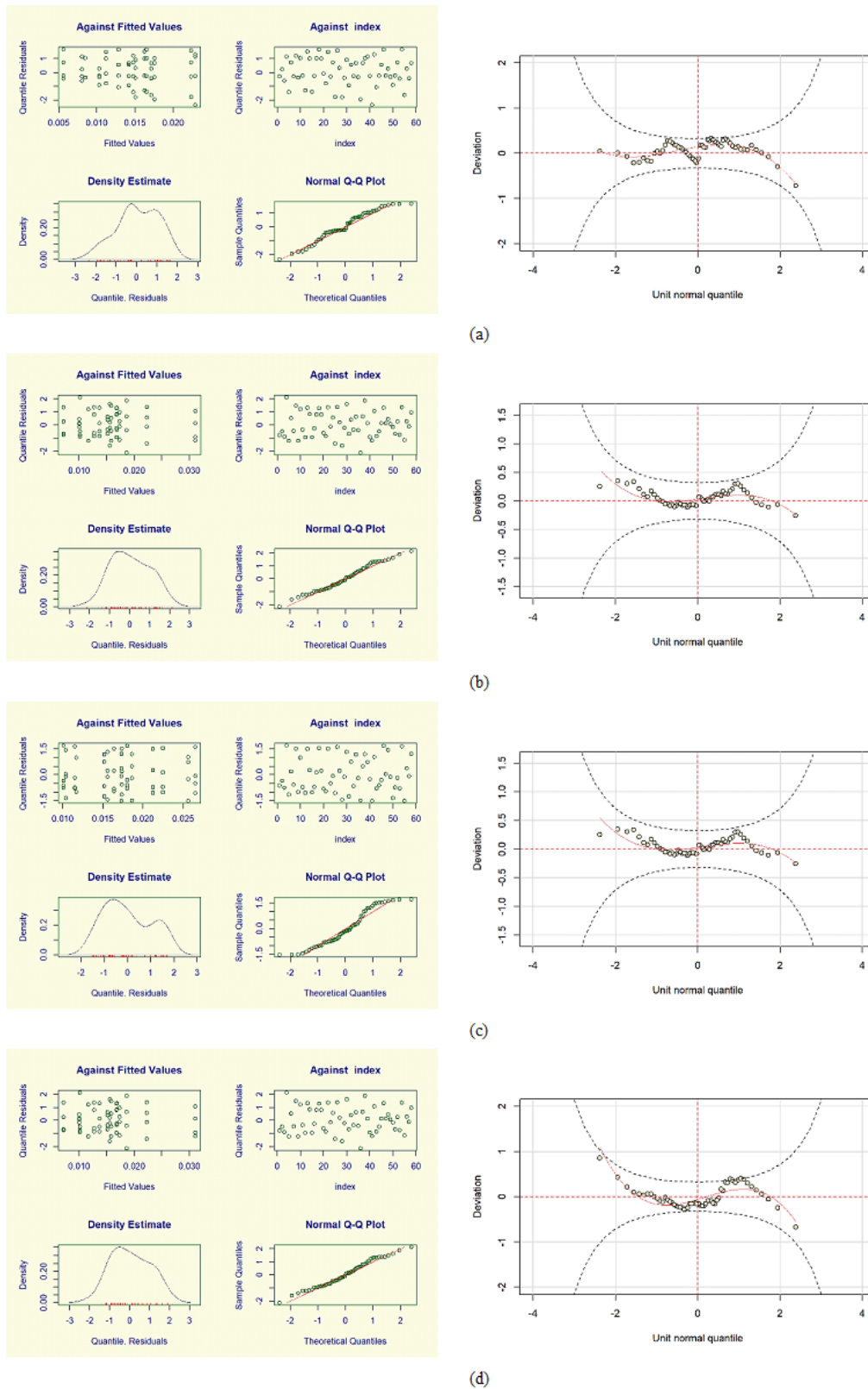
Na Tabela 5, os resultados apresentados são referentes ao GAIC dos modelos ajustados, expostos alguns de melhor resultado na Figura 10. Apesar de ter menores valores de GAIC que a distribuição Normal e que a *Skew t* tipo 4, as distribuições SHASHo2 (M11), Inversa Gaussiana (M4 e M7), Gama com modelagem do parâmetro  $\alpha$  (M6), foram descartadas devido a problemas encontrados nos worm plots, análise de resíduos e valores dos coeficientes de assimetria e curtose. Sem obter bons resultados com a transformação de dados e embora a natureza dos dados utilizados neste trabalho tenha sugerido intuitivamente a distribuição Gama para modelagem da variável resposta - com a restrição de modelamos apenas os parâmetros de locação e escala (M2 e M6) - os modelos GAMLSS foram preferidos por 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 para a modelagem GLM, se tornando portanto uma alternativa recente para ajustar dados de grande complexidade.

Tabela 5 – Principais modelos ajustados via modelagem GAMLSS para os dados de *Eucalyptus* spp.

Modelos Ajustados	GL	GAIC
M11: SHASHo2	18	-454,4374
M7: Inversa Gaussiana	30	-453,1166
M6: Gama	30	-450,0557
M8 : Logística	30	-438,7704
M3: Gumbel Reversa	16	-436,7074
M2: Gama	16	-432,2141
M4: Inversa Gaussiana	16	-429,6533
M9: Skew t tipo 4	32	-425,0601
M5: Logística	16	-421,1904
M10: Skew t tipo 4	46	-418,9769
M1: Normal	16	-417,6236

Fonte: Elaborada pelo Autor.

Com os gráficos dos resíduos e *worm plots* dos principais modelos considerados na Tabela 5 apresentados na Figura 10, (a) SASHo2 (M11), (b) *Skew t* com modelagem do parâmetro  $\tau$  (M10), (c) Gumbel Reversa (M3), (d) Gama com modelagem do parâmetro *alpha* (M6). Então, à exceção do modelo Gama (M6) (Figura 10 (d)) que não teve bom ajuste quanto a assimetria, os modelos obtiveram resultados satisfatórios quanto à análise de resíduos e *worm plot* consistentes, porém ao se olhar os parâmetros de locação, escala e forma via cálculos da média, coeficientes de assimetria e curtose de todos os modelos da Figura 10, percebeu-se que todos os ajustes tiveram uma diferença superior no parâmetro de assimetria e de curtose na ordem de 0,70 unidades, resultado este pior que o apresentado pela distribuição *Skew t* tipo 4 (ST4) (M9).



Fonte: Elaborada pelo Autor.

Figura 10 – Análise de resíduos e *worm plot* para os modelos ajustados SASHo2 (M11) (a), *Skew t* com modelagem do parâmetro  $\tau$  (M10) (b), Gumbel Reversa (M3) (c), Gama com modelagem do parâmetro  $\sigma$  (M6) (d).

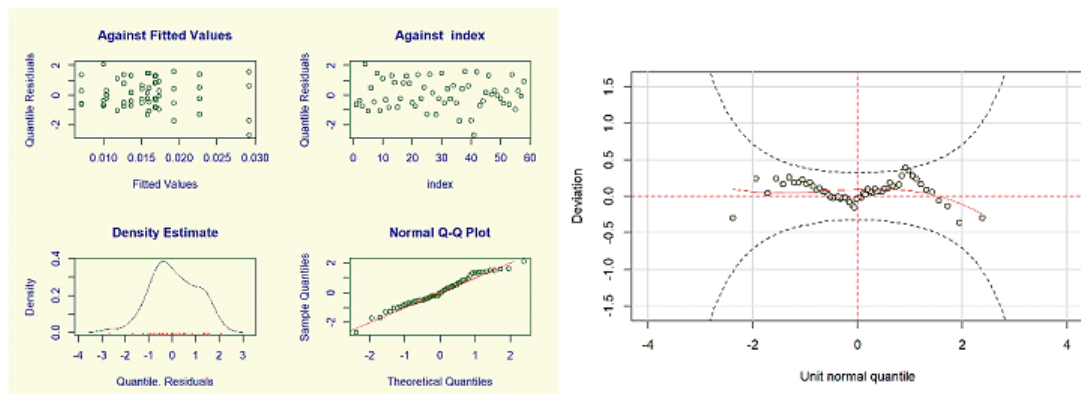
Na Tabela 6 se tem as estatísticas obtidas pelo ajuste do modelo (M9) com média igual a 0,0674 que é relativamente próximo de zero, a variância de 0,9755 bem próxima de 1, logo se aproxima de uma normal padrão, assimetria de -0,1328 relativamente próxima de zero, curtose de 2,6270 próxima de 3 e por fim o coeficiente de correlação de Filliben igual a 0,9888 próximo de 1, também foi aplicado o teste de Shapiro-Wilks para os resíduos do modelo *Skew t* do tipo 4, e o mesmo obteve um Valor P superior a 0,05 (Valor P de 0,37). Estes resultados dão maior segurança do ponto de vista da aplicação desta distribuição aos dados.

Tabela 6 – Estimativa dos parâmetros para o ajuste da *Skew t* do tipo 4.

Parâmetros	Valores Estimados
Média	0,0674
Variância	0,9755
Coefficiente de Assimetria	-0,1328
Coefficiente de Curtose	2,6270
Coefficiente de correlação de Filliben	0,9888

Fonte: Elaborada pelo Autor.

Os resultados encontrados dos gráficos de resíduos e *worm plot* do modelo (M9), evidenciam melhores resultados que os encontrados nos modelos anteriores, Figura 10 e Figura 9, um ponto que chama atenção porém é que no *worm plot* do modelo (M9), há um ponto que viola as bandas de confiança, porém este não corresponde a 5% dos pontos que poderiam estar fora do intervalo com 95% de confiança.



Fonte: Elaborada pelo Autor.

Figura 11 – Análise de resíduos e *worm plot* para a modelagem via *Skew t* tipo 4.

Uma vez ajustado o modelo com a distribuição *Skew t* tipo 4, com o efeito de tratamento para o parâmetro  $\alpha$  e o intercepto para o parâmetro  $\tau$  (M9) (Tabela 5). Realizou-se o teste de Tukey para o efeito de tratamentos (Tabela 7). Observa-se, que os tratamentos T7 e T11, indicam maior produtividade de volume dos híbridos de *Eucalyptus urophylla* em relação aos demais o que podemos dizer que é o mais indicado em relação a adequabilidade do cultivo para a região. E em relação aos demais resultados obtidos o tratamento menos produtivo foi o T1 híbrido de *Eucalyptus tereticorn* e este fato pode ser provado por diversos motivos dentre eles características genotípicas não favoráveis às condições climáticas e edáficas da região, logo apresenta desvantagens para o plantio nesta área.

Tabela 7 – Médias dos tratamentos e teste Tukey para modelagem GAMLSS *Skew t* tipo 4 e via distribuição normal.

Tratamentos	Médias	Grupos Tukey GAMLSS ST	Grupos Tukey Normal
T7	0,0265	a	a
T11	0,0256	a	a
T5	0,0224	a b	a
T4	0,0212	a b	a
T10	0,0186	a b	a
T12	0,0180	a b	a
T2	0,0173	a b	a
T15	0,0173	a b	a
T6	0,0164	a b	a
T8	0,0155	a b	a
T9	0,0152	a b	a
T3	0,0146	b c	a
T14	0,0115	b c	a
T13	0,0103	b c	a
T1	0,0101	b c	a

As médias seguidas das mesmas letras não diferem estatisticamente entre si nas colunas, ao nível de 5% de probabilidade e  $CV\% = 38,43$ .

Fonte: Elaborada pelo Autor.

É possível notar também por meio da Tabela 7, que se tivesse optado por modelar os dados via distribuição normal, não se teria condições de escolher um único clone que apresenta-se maior produção, devido ao teste de Tukey não encontrar nenhuma diferença mínima significativa ao nível de 5%, nestas condições.

A maior produtividade em relação ao volume dos clones em pauta foram os T11 e T7 pode estar atribuída ao fato da espécie ter bom crescimento, enraizamento, capacidade

de brotação de cepas e ser tolerante à falta de água, alta resistência a doenças e plasticidade da espécie (VALLE, 2009; FREITAG, 2013; FONSECA et al., 2010). Segundo Moura et al. (1995) a espécie *E. brassiana* identificada originalmente na Papua Nova Guiné, crescendo em solos mal drenados às margens de pântanos, várzeas inundáveis e em depressões, efetivamente diferente do tipo de solo de Araripina o que pode ter sido o fator limitante para suas brotações. E ainda, observa-se o crescimento desta espécie em encostas com solos com escassez de estrutura rochosa e bem drenadas. De acordo com Lima (1996), algumas espécies incluindo a *Eucalytus tereticorn* possuem a capacidade de crescer em áreas encharcadas e drenar o excesso, mas na localidade instalada o referido experimento não possui essas características.

## 6 Conclusão

Durante a elaboração deste trabalho foram apresentadas as características e propriedades da classe de modelos de regressão, mais precisamente denominada de modelos aditivos generalizados para locação, escala e forma (GAMLSS), além das vertentes relacionadas à inferência e diagnóstico, assim como, a flexibilidade que a análise do modelo permite por possuir uma ampla família de distribuições para a variável resposta. Vale salientar que o uso da classe desses modelos se mostrou-se eficiente por sua acurácia e precisão na obtenção dos resultados e surge como uma ferramenta poderosa para lidar com situações de dados com comportamentos assimétricos, que exigem o emprego de modelos mais complexos.

Para o mesmo conjunto de dados, os resultados foram comparados com os obtidos por meio da aplicação do modelo normal e dos modelos de lineares generalizados. As análises realizadas revelaram que os modelos estimados via GAMLSS viabilizaram um ajuste superior aos obtidos via MLG, segundo o critério GAIC e as análises dos resíduos (gráficos *worm plot*), mostrando que a classe do modelos GAMLSS aparenta ser mais apropriada do que as tradicionais modelagens via GLM. Vale destacar a presença de efeito do teste de Tukey aplicado ao GAMLSS em relação ao teste aplicado em condições de normalidade, haja vista que, sem a técnica não seria possível a seleção de um clone se não fosse utilizada uma distribuição adequada aos dados.

Por meio do emprego do modelo GAMLSS, ajustado com a distribuição *skew t* do tipo 4, foi possível verificar que os clones 39 e clone 41 híbridos de *Eucalyptus urophylla* (cruzamento natural) apresentaram o melhor rendimento volumétrico na região do Polo Gesseiro do Araripe - PE.

## Referências Bibliográficas

- AITKIN, A. M.; FRANCIS, B.; HINDE, J.; DARNELL, R. **Statistical modelling in R**. Oxford University Press Oxford, 2009.
- AKAIKE, H. Information theory and an extension of the maximum likelihood principle. In: **Selected Papers of Hirotugu Akaike**. [S.l.]: Springer, p. 199-213, 1998.
- ALBUQUERQUE, J. L. **Diagnóstico ambiental e questões estratégicas: uma análise considerando o Pólo Gesseiro do sertão do Araripe – Estado de Pernambuco**. 2002, 185f. Tese (Doutorado em Engenharia Florestal) – Universidade Federal do Paraná, Curitiba, 2002.
- ALCORN, P. J.; BAUHUS, J.; THOMAS, D. S.; JAMES, R. N.; SMITH, R. G. B.; NICOTRA, A. B. Effect on initial planting density on branch development in 4-year-old plantation grown *Eucalyptus pilularis* and *Eucalyptus cloeziana* trees. **Forest Ecology and Management**, n. 252, p. 41-51, 2007.
- ALVES, A. M. C. **Quantificação da produção de biomassa e do teor de carbono fixado por clones de Eucalipto, no pólo gesseiro do Araripe-PE**. 75 p. Dissertação (Ciências Florestais) — Universidade Federal Rural de Pernambuco, Departamento de Ciências Florestais, 2007.
- ANDRADE, E. N. **O Eucalipto**. 2 Ed. Impresso nas Oficinas Tipográficas da Campanhia Paulista de Estradas de Ferro. São Paulo, 667 p., 1961.
- ATECEL. Associação Técnica Científica Ernesto Luiz de Oliveira. **Diagnóstico energético do setor industrial do polo gesseiro da meso região de Araripina-PE**. Campina Grande, 126 p., 2006.
- BARRETO, V. C. S.; BATISTELA, G. C.; GAIOTTO, M. R.; SIMÕES, D. Regressão linear múltipla aplicada ao preço do leite. **Revista Eletronica Paulista de Matemática**, v. 7, 111 p., 2016.
- BARROS, K. N. N. O. **Abordagem clássica e Bayesiana em modelos simétricos transformados aplicados à estimativa de crescimento em altura de Eucalyptus urophylla no Pólo gesseiro do Araripe**. 51 p. Dissertação (Biometria e Estatística Aplicada) — Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, 2010.



- BARTON, C.V.M.; MONTAGU, K.D. Effect of spacing and water availability on root: shoot ratio in *Eucalyptus camaldulensis*. **Forest Ecology and Management**, 221, p. 52-62, 2006.
- BARTOSZECK, A. C. P. S.; MACHADO, S. do, A.; FILHO, A. F.; OLIVEIRA, E. B. A distribuição diamétrica para bracatingais em diferentes idades, sítios e densidades da região metropolitana de Curitiba. **Revista Floresta**, Curitiba, v. 34, n. 3, p. 305-324, set./dez. 2004.
- BATISTA, J. L. F. **Biometria Florestal segundo o Axioma da Verossimilhança Com Aplicações em Mensuração Florestal**. Tese de Livre-Docente. Departamento de Ciências Florestais – ESALQ-USP, Piracicaba, 2014.
- BERNARDO, A. L.; REIS, M. G. R.; REIS, G. G.; HARRISON, R. B.; FIRME, D. J. Effect of spacing on growth and biomass distribution in *Eucalyptus camaldulensis*, *E. pellita* and *E. urophylla* plantations in southeastern Brazil. **Forest Ecology and Management**, v. 104, p. 1-13, 1998.
- BRITO, J. O. O uso energético da madeira. **Estudos Avançados**, São Paulo, v. 21, n. 59, p. 185-193, 2007. Disponível em: <<https://www.revistas.usp.br/eav/article/view/10215/11820>> doi: <http://dx.doi.org/10.1590/S010340142007000100015>
- BUAINAIN, A. M.; BATALHA, M. O. **Cadeia produtiva de madeira**. Brasília, DF: Ministério da Agricultura, Pecuária e Abastecimento: IICA, 2007. 82 p. (MAPA. Agronegócios, v. 6). Disponível em: <<http://www.florestal.gov.br/snif/producao-florestal/cadeia-produtiva>> Acessado em 08 abr. 2018
- BUUREN, S. V.; FREDRIKS, M. Worm plot: a simple diagnostic device for modelling growth reference curves. **Statistics in medicine**, Wiley Online Library, v. 20, n. 8, p. 1259-1277, 2001.
- CASELLA, G.; BERGER, R. L. **Inferência Estatística**. 2 Ed. São Paulo: Cengage Learning, 588 p., 2010.
- COELHO, A. S. R.; MELLO, H. A.; SIMÕES, J. W. Comportamento de espécies de eucaliptos face ao espaçamento, Piracicaba, **IPEF**, Piracicaba, n. 1, p. 29-55, 1970.
- COLE, T. J.; GREEN, P. J. Smoothing reference centile curves: the lms method and penalized likelihood. **Statistics in medicine**, Wiley Online Library, v. 11, n. 10, p. 1305-1319, 1992.
- CONFEDERAÇÃO DA AGRICULTURA E PECUÁRIA DO BRASIL (CNA). **Brasil é um dos maiores produtores de florestas plantadas do mundo**. 2017. Disponível em: <<http://www.cnabrasil.org.br/noticias/o-brasil-e-um-dos-maiores-produtores-de>

florestas-plantadas-do-mundo>Acesso em: 18 dez. 2018.

CUNHA, A.; BARBOSA, M.; FELISMINO, D.; DANTAS, I. Consumo da lenha na calcinação da gipsita e impactos ambientais no polo gesseiro da mesorregião do Araripe - PE. **Revista de Biologia e Farmácia**, v. 2, n. 1, 2008.

DEMÉTRIO, C. G. B. Modelos lineares generalizados em experimentação agronômica. [S.l.]: **USP/ESALQ**, 2001.

DRAPER, N.; SMITH, H. **Applied Regression Analysis**. 3 Ed. New York, USA: John Wiley e Sons, 736 p., 1998.

DUNN, P. K.; SMYTH, G. K. Randomized Quantile Residuals. **Journal of Computational and Graphical Statistics**, JSTOR, v. 5, n. 3, p. 236, set 1996. Disponível em: <<http://dx.doi.org/10.2307/1390802>>. Acesso em: jun. 2018.

EILERS, P. H.; MARX, B. D.; DURBÁN, M. Twenty years of p-splines. **SORT: statistics and operations research transactions**, v. 39, n. 2, p. 149-186, 2015.

FERREIRA, C. A. **Cultivo do eucalipto**: Manejo de plantações para desdobro. In: Embrapa Florestas Sistemas de Produção, 4 ISSN 1678-8281 Versão Eletrônica Ago./2003. Disponível em: <[http://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Eucalipto/-CultivodoEucalipto/08\\_conducao\\_para\\_serraria.htm](http://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Eucalipto/-CultivodoEucalipto/08_conducao_para_serraria.htm)>. Acesso em: 05 de agosto de 2018.

FIGUEIREDO, A. C.; PEDRO, L. G.; BARROSO, J. G.; TRINDADE, H.; SANCHES, J.; OLIVEIRA, C.; CORREIA, M. **Óleos essenciais de espécies de eucalyptus**, 2013. Disponível em: <[http://cbv.fc.ul.pt/Agrotec\\_8.pdf](http://cbv.fc.ul.pt/Agrotec_8.pdf)>. Acesso em: 1 Fev. 2018.

FISHER, R. A. **The design of experiments**. Oliver And Boyd; Edinburgh; London, 1937.

FLORENCIO, L. A. **Engenharia de avaliações com base em modelos GAMLSS**. Tese (Doutorado) — Universidade Federal de Pernambuco, 2010.

FONSECA, S. M.; RESENDE, M. D. V.; ALFENAS, A. C.; GUIMARÃES, L. M. S.; ASSIS, T. F. **Manual prático de melhoramento genético do eucalipto**. Viçosa: UFV, 200 p., 2010.

FONTENELE, N. M. **Comparação entre a produtividade e análise econômica de clones de Eucalyptus spp. em sistemas de alto fuste e talhadia no Polo Gesseiro do Araripe – PE**. 2016. Dissertação (Programa de Pós-Graduação em Ciências Florestais) - Universidade Federal Rural de Pernambuco, Recife, 62 p., 2016.

FREITAG, A. S. **Crescimento de brotações de um clone de *Eucalyptus urophylla* x *Eucalyptus grandis* em função da disponibilidade de nutrientes no solo e da**

- aplicação de fitorreguladores na cepa.** Tese (Doutorado em Recursos Florestais) – Universidade de São Paulo, Escola Superior de Agricultura “Luiz de Queiroz”, 82 p., 2013.
- GADELHA, F. H. L. **Rendimento volumétrico e energético de clones de híbridos de *Eucalyptus urophylla* (Cruzamento Natural) e *Eucalyptus brassiana* (Cruzamento Natural) na chapada do Araripe-PE.** Dissertação (Ciências Florestais) — Universidade Federal Rural de Pernambuco, Departamento de Ciências Florestais, 65p., 2010.
- GADELHA, F. H. L. **Desempenho silvicultural e avaliação econômica de clones híbridos de eucaliptos plantados em diferentes regimes de manejo para fins energéticos.** Tese de doutorado (Ciências Florestais) - Universidade Federal Rural de Pernambuco. Recife, 2014.
- GADELHA, F. H. L.; SILVA, J. A. A.; FERREIRA, R. L. C. Viabilidade Econômica do Cultivo de Clones de *Eucalyptus* spp. Em função do Espaçamento e do sistema Silvicultural, em Araripina – PE. Recife, **Anais da Academia Pernambucana de Ciência Agrônoma**, v.15, n.2, p.105-123, 2018.
- GANZOLA, S. **Construção de um modelo de regressão para avaliação de imóveis.** 2002. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Florianópolis, 104 p., 2002.
- GONÇALVES, J. C.; OLIVEIRA, A. D.; CARVALHO, S. P. C.; GOMIDE, L. R. Análise Econômica da Rotação Florestal de Povoamentos de Eucalipto Utilizando a Simulação de Monte Carlo. **Ciência Florestal**, v. 27, n. 4, p. 1339-1347, 2017.
- GOUVEIA J. F; SILVA, J. A. A.; FERREIRA, R. L. C.; GADELHA, F. H. L.; FILHO, L. M. A. L. Modelos Volumétricos mistos em clones de *Eucalyptus* no polo gesseiro do Araripe, Pernambuco. **Floresta**, v. 45, n. 3, p. 587-598, 2015.
- GRANJA, C. V. A.; CAVALCANTE, E. P.; FILHO, H. P. C.; SIQUEIRA, M. S.; NASCIMENTO, W. Degradação Ambiental Exploração de Gipsita no Polo Gesseiro do Araripe. **Id on Line Revista Multidisciplinar e de Psicologia**, v.11, n. 36, p. 239-264, 2017.
- HASTIE, T.; TIBSHIRANI, R. **Generalized additive models.** [S.l.]: Wiley Online Library, 1990.
- HESS, A. F. CIANORSCHI; L.D., SILVESTRE; R., SCARIOT, R.; RICKEN, P. Aplicação dos modelos lineares generalizados para estimativa do crescimento em altura. **Pesquisa Florestal Brasileira**. v. 35 n. 84. p. 427-433. 2015.
- HOFFMANN, R. **Análise de Regressão: Uma introdução à econometria.** 4. ed. São Paulo: Hucitec, 378 p., 2006.

- HOTHORN, T.; BRETZ, F.; WESTFALL, P. Simultaneous Inference in General Parametric Models. **Biometrical Journal**, v. 50, n. 3, p. 346-363, 2008.
- Instituto Brasileiro de Geografia e Estatística – IBGE. **Produção de Extração Vegetal e da Silvicultura**, 2017. Disponível em: <[https://biblioteca.ibge.gov.br/visualizacao/periodicos/74/pevs\\_2017\\_v32\\_informativo.pdf](https://biblioteca.ibge.gov.br/visualizacao/periodicos/74/pevs_2017_v32_informativo.pdf)> Acesso em: 28 Nov. 2018.
- JONE, M. C.; FADDY, M. J. A Skew extension of the t distribution, with applications. **J. Roy. Statist. Soc B**, v. 65, p. 159-174, 2003.
- JORGENSEN, D. B. **Industrial Minerals and Rocks**. Ed. D. D. Carr, Soc. Mining, Metallurgy Explorationn Inc., Littleton, Colorado, EUA, 1994.
- LIMA, W. P. **Impacto ambiental do eucalipto**. 2. Ed. EDUSP, 203 p., 1996.
- MACERA, M. H. C. **Uso dos métodos clássico e bayesiano para os modelos não-lineares heterocedásticos simétricos**. Dissertação (Ciências de Computação e Matemática Computacional) – Universidade de São Paulo. São Carlos, 2011.
- MELO, K. K. S.; LIMA, A. P. C.; SANTANA, M. C.; ANDRADE, V. C. P.; BRAGA, A. L. C.; CORREIA, K. V. Caracterização química e Mineralógica dos Resíduos da Mineração de Gipsita no Semiárido Pernambucano. **HOLOS**. v. 6, n. 33, 195 p. 2017.
- MONTGOMERY, D. C. **Design and analysis of experiments**. John wiley & sons, 2017.
- MORA, A. L.; GARCIA, C. H. **A cultura do eucalipto no Brasil** (Eucalypt cultivation in Brazil). São Paulo: SBS – Sociedade Brasileira de Silvicultura, 112 p., 2000.
- MOURA, V. P. G.; OLIVEIRA, J. B.; VIEIRA, V. M.; Avaliação de procedências de *Eucalyptus brassiana* s. t. blake em Planaltina, Distrito Federal, área de Cerrado. **IPEF** n. 48/49, p. 87-97, jan./dez.1995.
- NELDER, J. A.; WEDDERBURM, R. W. M. Generalized linear models. **Journal of the Royal Statistical Society**, v. 135, n. 3, p. 370-384, 1972.
- 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, 302 p., 2008.
- PAULA, G. A. **Modelos de Regressão com apoio computacional**. São Paulo, 2013. Disponível em: <[http://www.ime.usp.br/giapaula/texto\\_2013.pdf](http://www.ime.usp.br/giapaula/texto_2013.pdf)>.
- PIERRO, B. Modos de restaurar as florestas. **Pesquisa FAPESP**, v. 238, p. 32-35, 2015.
- PINTO, D. S.; RESENDE, R. T.; MESQUITA, A. G. G.; ROSADO, A. M.; CRUZ, C. D. Seleção precoce para características de crescimento em testes clonais de *Eucalyptus*.

**Scientia. Forestalis**, Piracicaba, v. 42, n. 102, p. 251-257, 2014.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2018. Disponível em: <<http://www.R-project.org/>>.

RENCHER, A. C.; SCHAALJE, G. B. **Linear models in statistics**. [S.l.]: John Wiley & Sons, 2008.

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.

RIGBY, R. A.; STASINOPOULOS, D. M. Automatic smoothing parameter selection in gamlss with an application to centile estimation. **Statistical methods in medical research**, Sage Publications Sage UK: London, England, v. 23, n. 4, p. 318-332, 2014.

RIGBY, R.; STASINOPOULOS, M.; HELLER, G.; BASTIANI, F. **Distributions for Modelling Location, Scale, and Shape: Using GAMLSS in R**. [S.l.: s.n.], 2017.

ROCHA, K. D. **Produtividade volumétrica de clones de Eucalyptus spp. Na região do Pólo gesseiro do Araripe**. Tese (Ciências Florestais) — Universidade Federal Rural de Pernambuco, Departamento de Ciências Florestais, 111 p., 2012.

RODGERS, J. L.; NICEWANDER, W. A. Thirteen ways to look at the correlation coefficient. **The American Statistician**, Taylor & Francis, v. 42, n. 1, p. 59-66, 1988.

ROMERO, E. C., SEIXAS F., MOREIRA, R. M., SIXEL, R.M.M. Avaliação das Taxas de Crescimento Inicial e de Sobrevivência das Espécies de Eucalyptus do Projeto TUME (Teste de Uso Múltiplo de Eucalyptus). **Revista Arvore**, v. 48, n. 3, p. 316-321, abr/jun, 2003.

ROQUIM, F. V. **GAMLSS na experimentação agropecuária: um estudo em doenças parasitárias de bovinos de leite**. Dissertação (Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras. Lavras, 2018.

SBPCPE. Sociedade Brasileira Para o Progresso da Ciência Regional Pernambuco. Disponível em: <[sbpcpe.org/antigosite/anuncios/pologesseiro/simposiopologesseiro/introdução.html](http://sbpcpe.org/antigosite/anuncios/pologesseiro/simposiopologesseiro/introdução.html)> Acesso em: 19 out. 2014.

SBS. Sociedade Brasileira de Silvicultura. **Eucalipto com ciência**, plantio consciente, 2011. Disponível em: <[http://www.sbs.org.br/destaques\\_eucalipto\\_ciencia.htm](http://www.sbs.org.br/destaques_eucalipto_ciencia.htm)>. Acesso em: 19 out. 2018.

SCAVINSKI, V. **Projeção da Produção Utilizando Curvas de Sítio Anamórficas e Polimórficas para Plantios de *Eucalyptus grandis* W. Hill. Ex. Maiden**.

- Dissertação (Ciências Florestais) – Universidade Estadual do Centro-Oeste. Paraná, 2014.
- SCHAARSCHMIDT, F. BSagri: Safety Assessment in Agricultural Field Trials. R package version 0.1-10. Disponível em: <<https://CRAN.R-project.org/package=BSagri>>, 2018.
- 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.
- SEARLE, S. R.; CASELLA, G.; MCCULLOCH, C. E. **Variance components**. [S.l.]: John Wiley & Sons, v. 391, 2009.
- SEBRAE. Serviço Brasileiro de Apoio às Micro e Pequenas Empresas. **O que são florestas energéticas**. 2015. Disponível em: <<http://www.sebrae.com.br/sites/PortalSebrae/artigos/o-que-sao-florestas-energeticas,50a3438af1c92410VgnVCM100000b272010aRCRD>> Acesso em: 16 set. 2018.
- SILVA, J. A. A.; NETO, F. P. Princípios básicos de dendrometria. Recife, PE: Imprensa Universitária da UFRPE, 1979. 191 p.
- SILVA, J. A. A. Potencialidades de florestas energética de eucalyptus no pólo gesseiro do araripe, pernambuco. **Anais da Academia Pernambucana de Ciências Agrônomicas**, v. 5-6, p. 301-319, 2008/2009.
- SINDUSGESSO. Sindicato das Indústrias do Gesso do Estado de Pernambuco . **O Gesso como Recurso Mineral Estratégico para Pernambuco**. 2014. Disponível em: <[www.abema.org.br/site/.../PALESTRANASEINFRAJUNHO2014.pdf](http://www.abema.org.br/site/.../PALESTRANASEINFRAJUNHO2014.pdf)>. Acesso em: 28 Mai. 2018.
- STASINOPOULOS, M. D.; RIGBY, R. A. Generalized additive models for location scale and shape (gamlss) in r. **Journal of Statistical Software**, v. 23, n. 7, p. 1-46, 2007.
- STASINOPOULOS, M. D.; RIGBY, R. A.; HELLER, G. Z.; VOUDOURIS, V.; BASTIANI, F. **Flexible Regression and Smoothing: Using GAMLSS in R**. [S.l.]: CRC Press, 2017.
- VALE, A. B. do; MACHADO, C. C.; PIRES, J. M. M.; VILAR, M. B.; COSTA, C. B.; NACIF, A. de P. (Ed.). **Eucaliptocultura no Brasil – Silvicultura, Manejo e Ambiente**. 2014. Viçosa, MG, SIF, 551p.
- VALLE, M. L. A. **Propriedades da madeira de eucalipto de primeira e segunda rotação, visando a sua utilização como madeira preservada**. Dissertação (Mestrado em Ciência Florestal) – Universidade Federal de Viçosa, Viçosa, 96 p., 2009.
- VAN BUUREN, S. Worm Plot to Diagnose Fit in Quantile Regression. **Statistical Modelling**. **SAGE Publications**, v. 7, n. 4, p. 363-376, 2007.

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.

WOOD, S. N. Generalized Additive Models: An Introduction with R. **Boca Raton: Chapman & Hall**, 2006.