

TRABALHO DE CONCLUSÃO DE CURSO

ENGENHARIA FLORESTAL

COMPORTAMENTO DE DIFERENTES ABORDAGENS NA MODELAGEM
DE RELAÇÃO HIPSOMÉTRICA NA PRESENÇA DE *Outliers*

NAILA THAIANE RODRIGUES LEÃO



Naila Thaiane Rodrigues Leão

**COMPORTAMENTO DE DIFERENTES ABORDAGENS NA MODELAGEM DE
RELAÇÃO HIPSOMÉTRICA NA PRESENÇA DE *Outliers***

Trabalho de Conclusão de Curso apresentado ao Instituto de Ciências Agrárias da Universidade Federal de Minas Gerais, como requisito parcial, para a obtenção do título de Bacharel em Engenharia Florestal.

Orientador: Prof. Christian Dias Cabacinha

Montes Claros
2022

Naila Thaianne Rodrigues Leão

**COMPORTAMENTO DE DIFERENTES ABORDAGENS NA MODELAGEM DE
RELAÇÃO HIPSOMÉTRICA NA PRESENÇA DE *Outliers***

Aprovada pela banca examinadora constituída por:

Prof. Adriana Leandra de Assis - ICA/UFMG

Gabriel Augusto Eça

Prof.^a Christian Dias Cabacinha - Orientador ICA/UFMG

Montes Claros, 17 de fevereiro de 2022.

Dedico aos meus pais e aos meus
irmãos, por me ensinarem que o amor
está presente em detalhes que não cabem
em uma dedicatória.

AGRADECIMENTOS

Agradeço primeiramente a Deus por ter me sondado e capacitado em todos os dias da minha graduação.

Aos meus pais, Sebastiana e Edson, por todo o amor, apoio, conselhos e eventuais reclamações, porque nem tudo são rosas.

Aos meus irmãos, Júnior e Jesus, que tornam os meus dias mais leves e felizes.

Ao meu orientador Christian, por todo apoio, paciência, por todos ensinamentos e pelas risadas em meio ao desespero.

Aos meus familiares, por serem minha base e meu motivo para sempre continuar.

À Maria Fernanda, que desde o início divide as alegrias e as lutas da graduação comigo. Você deixou essa caminhada mais fácil.

À Vitória e Paula, que compartilharam a casa e as contas comigo durante esses anos, vocês agregaram muito aos meus dias.

Aos amigos que conquistei ao longo desses anos, sobretudo aos grupos “Bonde das Maravilhas” e a “Nata do ICA”, sem vocês metade das risadas que já dei na vida não existiriam.

Aos meus colegas do ICA, os florestais e não florestais, por alegrarem minha vida pelos gramados do campus.

Ao LAMEF, por ser uma grande família e por sempre agregar ao meu currículo e à minha formação pessoal.

À UFMG e a FUMP, por todo apoio estrutural, psicológico e financeiro durante minha graduação.

À minha madrinha Dotiva, por me mostrar o amor da forma mais linda, verdadeira e dolorosa, minha admiração nunca caberá em palavras.

Ao meu primo Deleon (In memoriam), por me ensinar a ter fé e a ser forte, mesmo quando tudo estiver desabando.

E a todos que de alguma forma participaram da construção desse sonho.

Muito obrigada!

RESUMO

O conhecimento da altura das árvores é de fundamental importância para a obtenção de estimativas de produtividade de povoamentos florestais, contudo as medições de altura no campo para geração de relações hipsométricas estão sujeitas a vários erros (*outliers*) que podem gerar imprecisões nas estimativas de altura. Existem diferentes abordagens para obtenção de relações hipsométricas, tais técnicas como a regressão quantílica e as redes neurais artificiais, são apontadas pela literatura como menos sensíveis à ruídos na base de dados. Neste contexto, o presente trabalho teve como objetivo realizar simulações de diferentes cenários de ajustes de relações hipsométricas, utilizando *outliers* inseridos, em dados de altura de mogno africano (*Khaya spp.*). A base de dados foi originada da coleta de altura de 480 árvores, oriundas de 38 parcelas distribuída em 11 talhões com idades e espaçamentos diferentes. Foram utilizadas as abordagens dos Método dos Mínimos Quadrados, Regressão Quantílica, Modelagem Mista e uma Rede Neural Artificial (RNA). Foram feitos 4 experimentos, no experimento os ajustes foram realizados para a base de dados original, no experimento dois foram inseridos um *outliers* somando três metros nas altura de duas árvores de 14 parcelas aleatorizadas, no experimento três, para os *outliers* somou-se seis metros e no último experimento foram inseridos *outliers* somando-se nove metros as mesmas árvores. A avaliação das estimativas foi realizada pela correlação linear de Pearson entre os valores estimados e observados de altura ($r_{yy'}$), a raiz do erro quadrado médio (RMSE%), gráfico da dispersão dos resíduos, histograma de frequência dos erros percentuais e as estimativas de altura pelas diferentes abordagens foram comparadas pelo teste de equivalência. Os resultados mostram que, em relação a modelagem de maneira geral as técnicas de regressão geram resultados melhores que a RNA em todos os experimentos. Para os dados sem a introdução dos outliers e até possíveis erros de seis metros na medição de algumas árvores, não há diferença entre a altura média observada e estimadas pelas diferentes abordagens. A Regressão Quantílica foi a que gerou as menores tendenciosidades com correlação 0,79 e RMSE de 1,82m, mesmo na pior situação de introdução de *outliers*.

Palavras-chave: Mogno Africano; Regressão; Redes Neurais Artificiais.

LISTA DE ABREVIATURAS

ABPMA	– Associação Brasileira dos Produtores de Mogno Africano
HT	– Altura Total
DAP	- Diâmetro à altura do peito
MMQ	- Método dos Mínimos Quadrados
RQ	- Regressão Quantílica
MM2	- Modelos Mistos com efeito aleatório nos dois parâmetros
MM0	- Modelos Mistos com efeito aleatório no parâmetro b_0
MM1	- Modelos Mistos com efeito aleatório no parâmetro b_1
RNA	- Redes Neurais Artificiais

Sumário

1- Introdução	18
2- Referencial teórico	20
2.1- Relações hipsométricas	20
2.2- Regressão na Mensuração Florestal	21
2.3- Redes Neurais Artificiais	22
3- Metodologia	24
3.1- Área de estudo.....	24
3.2- Caracterização dos dados	25
3.3- Ajuste por Regressão.....	27
3.4- Teste de equivalência	30
3.5- Experimentos Avaliados	31
4- Resultados e Discussão	32
4.1- Experimento 1	32
4.2- Experimento 2	39
4.3- Experimento 3	45
4.4- Experimento 4	51
5- Conclusão	58
6- Referências	59

1- Introdução

As florestas plantadas são uma excelente alternativa para suprir a demanda madeireira, além de restaurar as áreas desmatadas e serem usadas no setor comercial, apresentando, de acordo com Maggi (2018), elevada relevância para economia do país.

Atualmente, muitos agricultores despertaram o interesse pelo cultivo de madeiras nobres, vistas como uma boa alternativa para ganhos futuros. Hoje se destaca no Brasil o cultivo do mogno africano (*Khaya spp.*). Dentre estas a mais cultivada no país é a *Khaya ivorensis* A. Chev., tendo como características o rápido crescimento, fuste retilíneo e resistência à praga causadora da broca dos ponteiros (*Hypsiphyla grandella*) (OLIVEIRA, 2017).

A estimação de estoques de crescimento e de colheita é um importante elemento no manejo florestal, uma vez que fornece informações quantitativas sobre a floresta, auxiliando na definição de planos de manejo e em análises econômicas de prescrições de manejo (BINOTI, M.L.M.S. et al., 2013). Neste sentido a relação hipsométrica é um aspecto importante a ser considerado no sistema de coleta de informações, pois a altura é uma variável difícil de ser mensurada, implicando em maior tempo para a sua mensuração, além de ter maior margem de erro na coleta, por ser geralmente uma medida indireta.

Existem muitos modelos para estimação da altura total de árvores, muitos deles podendo ser utilizados em várias espécies (VENDRUSCOLO et al., 2017). Estudos sobre modelos, procedimentos e equipamentos para a mensuração da altura de árvores são extremamente importantes, visto que a medição da altura das árvores é considerada componente significativo no custo do inventário florestal (LEITE; ANDRADE, 2004).

Atualmente, além das técnicas de modelagem por regressão, o aprendizado de máquinas está se destacando no meio florestal, devido a sua capacidade de gerar estimativas mais precisas que as equações obtidas com os modelos. Na mensuração florestal, as Redes Neurais Artificiais (RNA) estão sendo estudadas e ganhando espaço nesse meio, tais como: Binoti et al. (2013) que utilizaram RNA para estimar a altura de povoamentos equiâneos de *Eucalyptus spp.*; Vendruscolo et al. (2017) estimaram altura de árvores de *Tectona grandis* L.f.; entre outros.

Valores estranhos nas bases de dados, normalmente chamados de outliers geralmente impactam muito os parâmetros obtidos por regressão, sobretudo em métodos de regressão baseados na média. Algumas abordagens para obtenção de estimativas mostram-se menos sensíveis à presença de tais valores, destacando-se a regressão quantílica e a RNA.

Objetivou-se com este trabalho avaliar e comparar o comportamento de diferentes abordagens de ajustes hipsométricos em dados sem e com a presença de *outliers*, utilizando diferentes métodos de regressão e uma rede neural artificial para estimar a altura de árvores de um povoamento de *Khaya spp.*

2- Referencial teórico

2.1- Relações hipsométricas

A altura de uma árvore é a distância linear ao longo do eixo principal, partindo do solo até o topo, ou até outro ponto de referência a depender do tipo de altura que se quer medir, já que na mensuração florestal ela divide-se em altura total, comercial, do fuste e da copa (MACHADO; FIGUEIREDO FILHO, 2014).

Ela constitui uma importante característica da árvore e pode ser medida ou estimada. Sua medição ou estimativa é muito importante para o cálculo do volume, de incrementos em altura e, em determinadas situações, pode servir como indicadora da qualidade produtiva de um local (SILVA et al., 2012).

A medição da altura de todas as árvores na parcela, em um inventário florestal, é uma atividade que requer um elevado tempo de dedicação, pois é dificultada pela ação de alguns fatores como: inclinação acentuada do terreno, e da operação inadequada dos equipamentos para medição de alturas (SANQUETTA et al., 2014).

A relação entre diâmetro e altura das árvores é chamada de relação hipsométrica. Essa relação é de fundamental importância nos procedimentos de inventário florestal (SOARES; PAULA NETO; SOUZA, 2011), pois permite que apenas algumas árvores tenham sua altura medida no campo, aumentando a velocidade dos levantamentos e reduzindo seus custos. Qualquer tipo de modelo de diâmetro-altura pode nem sempre ser adequado para todos os tipos de condições em que uma espécie de árvore em particular pode ser encontrada, porque as condições do local podem afetar a relação diâmetro-altura (BAYAT et al. 2019).

Diversas variáveis podem influenciar a relação hipsométrica, dentre elas: idade, região de plantio, variação genética, tratamentos silviculturais, sistema de amostragem, densidade, tamanho da copa, posição sociológica e o regime de manejo adotado. Dessa forma, existem equações que se adaptam melhor a uma determinada condição do que outras, havendo, em alguns casos, dificuldade em se definir que modelo deve ser utilizado (Sousa et al., 2013).

A grande dificuldade da escolha do melhor modelo para representar essas relações hipsométricas se deve à não linearidade da relação entre as variáveis envolvidas e restrições

impostas aos parâmetros dos modelos, por razões biológicas (BARTOSZECK et al., 2004; SOARES et al., 2004).

2.2- Regressão na Mensuração Florestal

Na mensuração florestal, a técnica mais empregada para modelar a relação entre variáveis é a regressão (linear e não linear), por meio da qual são ajustadas equações para prever atributos (variáveis) de árvores, como por exemplo, peso, altura e volume e a produção presente e futura por unidade de área das florestas, sejam elas equiâneas ou inequiâneas (FARIAS, 2018).

Os modelos hipsométricos se dividem em lineares e não lineares. Os modelos lineares descrevem a variável Y como a soma de uma quantidade determinística (X) e uma quantidade aleatória (ϵ) de inúmeros fatores que podem, conjuntamente, ter influência sobre Y e podem assumir as seguintes formas: lineares simples e lineares múltiplos (CHARNET et al., 2008). Os modelos não lineares são aqueles que possuem seus parâmetros agregados na forma não aditiva e podem ser classificados em linearizáveis e não linearizáveis (DRAPER; SMITH, 1981).

Nos modelos de regressão, as estimativas dos parâmetros dos modelos lineares, normalmente são obtidas pela solução do sistema de equações normais, considerando a minimização da soma do quadrado dos erros (Mínimos Quadrados Ordinários ou Método dos Mínimos Quadrados) (CAMPOS; LEITE, 2017).

De acordo com Koenker e Basset (1978), o método dos Mínimos Quadrados Ordinários (regressão pela média) é extremamente sensível a valores discrepantes, tornando os seus estimadores pouco precisos, sendo recomendado, nesta situação, o uso da Regressão Quantílica (RQ) (regressão pela mediana). O método da Regressão Quantílica utiliza estimativas para os quantis condicionais de interesse da variável dependente, ao invés de uma única regressão para a média (SILVA, 2003).

Outra metodologia utilizada é a aplicação de modelos lineares e não lineares mistos ou generalizados, nos quais a estimação dos parâmetros é realizada considerando a função de probabilidade mais adequada à distribuição da variável resposta, bem como a matriz completa de variâncias e covariâncias (McCULLAGH; NELDER, 1989; CALEGÁRIO et al., 2005; CORDEIRO; DEMÉTRIO, 2011).

2.3- Redes Neurais Artificiais

As Redes Neurais Artificiais (RNA), em termos conceituais, consistem em um sistema paralelo composto por unidades especializadas (chamadas de nodos) com a função de calcular funções matemáticas (lineares e não lineares), sendo estas unidades interligadas por uma gama de conexões, associadas a pesos, que cumprem a função de armazenar o conhecimento adquirido através da entrada inicial de dados, assemelhando-se ao cérebro humano (SILVA et al., 2010).

Uma RNA consiste em múltiplas unidades de processamento simples, denominados neurônios artificiais, que estão conectados entre si e organizados em camadas, formando um sistema computacional paralelo para executar uma determinada tarefa (BULLINARIA, 2014). Estes fornecem estimativas precisas, tornando o alcance dos resultados relacionados à dendrometria florestal mais célere, menos onerosa e dispendiosa economicamente, além de ser menos laboriosa (FERREIRA et al., 2014).

O procedimento inicial é uma fase chamada de aprendizagem, que consiste em extrair características para representar a informação e armazenar o conhecimento. Depois do aprendizado, é feita a generalização, que nada mais é do que aplicar a RNA em dados não conhecidos (BRAGA; CARVALHO; LUDEMIR, 2007).

Em um processo de treinamento, essas redes têm a capacidade de aprender com exemplos e, a partir disso, extrair informações importantes. Esta etapa atribui à rede a possibilidade de fazer interpolações e extrapolações daquilo que aprenderam inicialmente, gerando um processo iterativo de ajuste de parâmetros, dando pesos às conexões entre as unidades de processamento, que guardam, ao final do processo, o conhecimento que a rede adquiriu do ambiente em que está operando (BRAGA et al., 2000; 2014; BINOTI, 2010).

O uso das RNA para o manejo de florestas teve seu início no final da década de 1980 (COULSON et al., 1987); No Brasil, as RNA têm sido utilizadas por empresas florestais, com as vantagens de simplificação nas rotinas de coleta e processamento de dados, diminuição no número de medições de alturas e consequente redução do custo do inventário (CAMPOS; LEITE, 2017).

Em alguns casos, as RNA têm apresentado desempenho superior aos modelos de regressão devido a diversos fatores, como: estrutura maciça e paralelamente distribuída

(camadas); habilidade de aprender e generalizar, que as tornam capazes de resolver problemas complexos; são tolerantes a falhas e ruídos; podem modelar diversas variáveis e suas relações não lineares; possibilidade de modelagem com variáveis categóricas (qualitativas), além das numéricas (quantitativas); e analogia neurobiológica (HAYKIN, 2001).

3- Metodologia

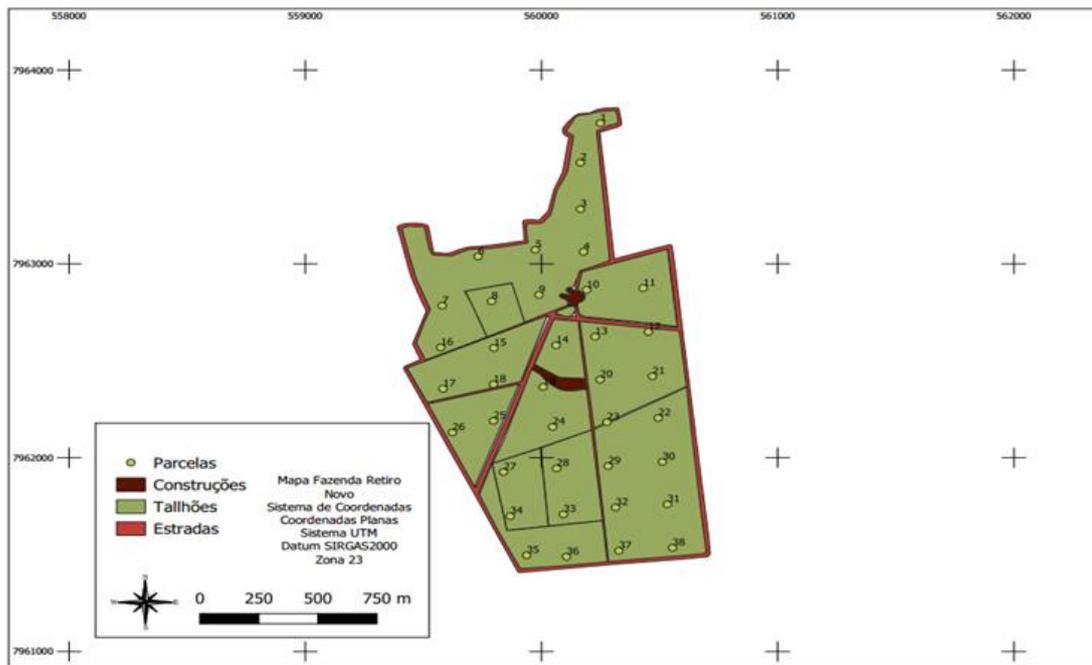
3.1- Área de estudo

Os dados utilizados neste estudo foram provenientes de medições de parcelas permanentes de um inventário florestal contínuo (IFC) realizado na fazenda Retiro Novo (Figura 1). A Fazenda possui 245,96 hectares sendo 172,97 ha de plantio de mogno africano (*Khaya spp.*), localizando-se a aproximadamente 10 quilômetros do município de Corinto-MG.

A propriedade possui uma área de 50 hectares de Reserva Legal averbada em cartório. O clima da região é caracterizado como tropical conforme a classificação de Köppen e Geiger. A temperatura média anual varia entre 22°C e 28°C onde a temperatura do mês mais quente varia entre 27,5°C e 30,5°C, e do mês mais frio varia entre 14,5°C e 19°C. A precipitação média anual está entre 800 a 1.200 mm. Com estiagem que pode variar de 4 a 7 meses, no período mais frio no ano. A evapotranspiração potencial anual varia entre 1.100 a 1.200 mm, com déficit hídrico acentuado que pode chegar a 210 mm. A altitude média da região situa-se entre 500 e 700 metros.

O relevo da propriedade pode ser caracterizado como plano a suave ondulado. A área do plantio não apresenta restrições à mecanização em nenhuma das etapas da produção de florestas plantadas. As declividades da área do plantio estão entre 0,5 e 4%. A propriedade está inserida dentro da Bacia Hidrográfica do Rio das Velhas.

Figura 1 - Mapa do plantio de mogno estudado com a distribuição das parcelas inventariadas nos diferentes talhões.



Fonte: Da autora.

3.2- Caracterização dos dados

A área da fazenda utilizada como fonte de dados divide-se em onze talhões, com diferentes idades e diferentes espaçamentos. O espaçamento dos talhões varia entre 4m X 5m, 4,5m X 5,5m e 6m X 4m. Os dados foram coletados em 38 parcelas permanentes circulares distribuídas sistematicamente na área. O raio das parcelas é de 18 metros, totalizando uma área de 1.018 m². As medições foram realizadas no ano de 2021, o centro de cada parcela foi georreferenciado e demarcado com tubos de PVC numerados. A primeira árvore da parcela foi denominada com referência ao lado direito do norte magnético marcada por um “X” à tinta, seguindo em forma de “zigue-zague” até o último indivíduo que foi marcado à tinta com um “O”.

Em todas as parcelas foram mensurados: o diâmetro a 1,30 m do solo de todas as árvores (DAP), e nas árvores das duas fileiras centrais da parcela também foram coletadas a altura total (Ht) e altura comercial (Hc). Após a coleta, esses dados foram tabulados para o processamento. Na Tabela 1, está apresentada uma síntese de informações das parcelas.

Tabela 1 - Síntese das informações das parcelas inventariadas.

Parcela	Árvores por parcela	DAP				HT			
		Média	Max	Mín	Desvio Padrão	Média	Max	Mín	Desvio Padrão
1	11	16,09	18	11,6	2,05	15,52	17,4	14	1,02
2	16	18,24	22,9	13,2	2,81	16,15	21,8	13	2,17
3	12	19,08	22,3	15,4	2,03	17,72	20,4	14,6	1,55
4	13	19,72	25,4	12	3,70	18,87	22,9	15	1,82
5	14	20,82	24,4	14,5	2,85	19,06	21,6	15,9	1,77
6	14	18,70	20,9	15,5	1,70	16,67	18,9	14	1,56
7	14	20,14	23	17	1,89	18,59	20,5	15,5	1,51
8	13	20,68	25,4	14,8	3,24	18,74	21	15,6	1,52
9	12	20,36	25,4	13,1	3,14	19,13	21,6	16,8	1,64
10	11	27,76	39,1	20	5,84	22,25	27,7	18,5	2,92
11	11	20,57	29,6	9,5	5,85	18,75	22,9	11,8	3,07
12	10	19,65	23,8	14,6	3,57	19,13	21,5	14,6	2,06
13	11	22,45	31,5	16,8	3,94	22,40	26,9	17,2	3,25
14	12	22,10	26,7	17,6	2,80	20,58	23,7	18,4	1,83
15	15	19,97	26,7	12,8	4,51	19,13	23,7	13	2,83
16	15	20,29	25,1	11,4	4,98	19,33	23,6	14,9	2,71
17	14	19,58	26,9	11,2	5,06	19,39	22,7	14	2,46
18	13	23,27	29,8	17,2	3,86	20,98	24,1	17,8	2,10
19	12	20,88	23,7	13,4	3,05	20,11	22,3	18,1	1,64
20	12	21,06	25,5	10,9	3,84	20,95	24,8	13,9	3,02
21	13	19,52	27,1	14,1	3,47	19,59	26,9	15,5	3,33
22	13	20,91	25,9	16,4	2,62	19,82	22,3	16,8	1,77
23	10	26,16	33,1	17,9	5,19	23,21	26,3	17,6	2,56
24	14	19,26	23,5	12,8	3,44	19,56	31,3	14,9	4,08
25	12	22,46	28,1	14,1	4,43	21,48	25,3	18,5	2,29
26	13	21,15	24,2	14,9	3,14	21,22	23,7	14,9	2,22
27	13	22,94	27,7	12,9	4,14	21,08	26,6	11	3,79
28	13	21,56	27,4	15,1	3,68	19,14	21,9	14,9	2,08
29	14	22,19	26,7	16	2,71	22,72	25,7	20	1,62
30	14	20,18	25,7	14,2	3,48	19,18	26,4	15,2	3,27
31	11	19,14	24	14,6	3,12	16,09	19,9	12,2	2,44
32	13	21,29	33,1	13,7	4,54	19,49	24,2	16,7	2,17
33	12	20,26	23,1	14,1	3,14	18,51	22,1	15,6	1,79
34	12	22,60	27	15	3,77	21,44	25,1	16,5	2,84
35	11	20,55	25,5	13,9	4,16	18,67	22,1	14,4	2,11
36	13	20,25	25,7	15,6	3,27	19,16	22,4	15,5	2,08
37	13	21,22	25,3	15,1	3,21	19,42	23,7	14	2,98
38	11	17,92	22,8	9,2	3,93	15,32	19,7	10,2	2,74
	480	20,76	39,1	9,2	4,05	19,41	31,3	10,2	2,96

Fonte: Da autora.

O trabalho contou com a coleta de altura de 480 árvores. Foi obtido então quatro bases de dados. A primeira base contou com as alturas originais. Logo em seguida foi selecionada de forma aleatória no Microsoft Excel 14 parcelas e em cada parcela foi sorteada duas árvores aleatórias para se inserir os *outliers*. A segunda base de dados teve 3 metros inseridos nas árvores selecionadas. A terceira e quarta base tiveram 6 metros e 9 metros inseridos respectivamente.

3.3- Ajuste por Regressão

O modelo utilizado nos ajustes por regressão, foi o modelo de Curtis, que possui ampla aplicação na área florestal, cuja representação é dada por:

$$\ln H_i = \beta_0 + \beta_1 DAP^{-1} + \varepsilon_i$$

Em que: H_i representa a altura total da árvore i ; DAP é o diâmetro do tronco da árvore i a uma altura de 1,30 m do solo; e β_0 e β_1 são os parâmetros do modelo que, por razões biológicas, estão sujeitos às restrições do tipo $\beta_0 > 0$ e $\beta_1 > 0$, sendo ε_i o erro associado do modelo.

Este modelo foi ajustado por diferentes abordagens de regressão. Os métodos testados foram os mais usados na literatura, sendo eles: Regressão pelo método dos Mínimos Quadrados (MMQ); Regressão Quantílica (RQ); Regressão com Modelagem Mista (MM), onde o efeito aleatório foi a parcela e testou-se a modelagem com este efeito aleatório no parâmetro β_0 (MM0), no parâmetro β_1 (MM1) e nos dois parâmetros conjuntamente (MM2); e por Rede Neural Artificial (RNA).

Os ajustes foram realizados no ambiente estatístico R. O Script utilizado foi adaptado para esse processamento, de forma que para a regressão pelo Método dos Mínimos Quadrados, utilizou-se o código “*lm*” e os ajustes foram realizados por parcela. Para as demais abordagens foi necessária a ativação de algumas bibliotecas do R.

Para a Regressão Quantílica, usou-se o pacote “*quantreg*”, o código “*rq*” e o argumento “*tau*” para especificar o quantil, que foi 0,5 (quantil da mediana) e os ajustes também foram por parcela.

Para a Modelagem Mista, usou-se o pacote “*nlme*” e o código “*lme*”. Para introdução do efeito aleatório foi necessário especificar o parâmetro do modelo e o efeito aleatório aplicado foi a parcela.

As redes neurais artificiais (RNA) treinadas foram do tipo Perceptron de múltiplas camadas, comumente conhecidas como MLP (Multilayer Perceptron). As arquiteturas do tipo MLP constituem os modelos neurais artificiais mais utilizados e conhecidos. Uma rede MLP é subdividida em camadas: camada de entrada, camada(s) intermediária(s) ou escondida(s) e camada de saída (NIED, 2007). De acordo com Machado e Fonseca Junior (2013), na

arquitetura da RNA de múltiplas camadas as entradas são prolongadas da camada de entrada para a camada de saída, passando por uma ou mais camadas ocultas.

O treinamento foi realizado com uso do software R, utilizando o pacote “*neuralnet*” e o código “*neuralnet*”. A rede foi configurada com apenas uma camada oculta, contando com quatro neurônios, com a variável de entrada DAP (quantitativa), e a variável de saída, a altura (quantitativa).

De acordo Vendruscolo et al. (2015) o uso de três a quatro neurônios na camada oculta da RNA é o suficiente para aplicações em inventário e biometria florestal, pois o excesso de neurônios acarreta em erros maiores, podendo ocorrer superajuste da RNA.

O algoritmo utilizado foi o Resilient Propagation (RPROP+), com função de ativação sigmoidal na camada oculta e de saída. A Função sigmoidal trata-se da função mais comum. É definida como uma função crescente com balanceamento adequado entre o comportamento linear e não linear e assume um intervalo de variação entre 0 e 1. Um exemplo de função sigmóide é a função logística (usada neste estudo), definida por:

$$f(\mu) = \frac{1}{1 + \exp(-a\mu)}$$

Sendo a o parâmetro de inclinação da função sigmóide (quanto maior o valor de a , mais inclinada se torna a curva).

Este algoritmo tem como vantagem a facilidade de calcular e adquirir aprendizagem sobre determinado problema, pelo fato de seu ajuste de pesos depender mais do sinal dos gradientes de erro, além de ser mais eficiente e recomendado para RNA do tipo Multilayer Perceptron (RIEDMILLER & BRAUN, 1993).

Os critérios de parada do treinamento das RNA foram, o número total de ciclos igual a 10.000 ou o erro quadrático médio inferior a 1%, portanto o treinamento foi encerrado quando um dos critérios foi atingido. Foram treinadas 30 redes e selecionada a com menor erro quadrático médio.

Nessa etapa foi utilizada toda a base de dados, sem o controle da parcela. Isso ocorreu porque as RNA têm uma precisão maior, uma maior capacidade de generalização, quando trabalha com uma maior massa de dados. Como pode ser verificado na Tabela 1, as parcelas

possuem em média 12 árvores medidas apenas e este número de observações não seria suficiente para um treinamento por parcela.

Para treinamento e validação da rede, foi utilizado 70% dos dados para treinamento e 30% para validação (ARAÚJO, 2015). A etapa de validação visa determinar a capacidade da rede em estimar satisfatoriamente para dados desconhecidos, ou seja, se a rede é capaz de extrapolar saídas coerentes na inserção de novos dados de entrada. (GOMES,2017)

Os ajustes e treinamento da RNA, foram feitos para a base de dados original e para a base de dados com os *outliers* introduzidos, com a finalidade de se observar o comportamento nas diferentes situações. Como saída do processamento, uma nova base de dados com as alturas estimadas por cada abordagem e em cada situação (com e sem *outliers*), foi gerada.

A avaliação das estimativas obtidas por regressão e RNA, foi realizada por critérios estatísticos, que foram a correlação linear de Pearson entre os valores estimados e observados de altura ($r_{yy'}$), a raiz do erro quadrado médio (RMSE%), gráfico da dispersão dos resíduos, histograma de frequência dos erros percentuais. A análise gráfica consistiu na inspeção estatística da dispersão dos erros (resíduos) percentuais em relação aos valores observados para o histograma de frequência dos erros percentuais, conforme Binoti (2010), sendo:

$$\text{erro}\% = \frac{(Y - \hat{y})}{Y} * 100$$

Em que, Y é a altura observada e \hat{Y} é a altura estimada.

A correlação entre os valores observados e estimados ($r_{y\hat{y}}$) indica a força e a direção da relação entre as duas variáveis. Essa correlação visa atribuir um grau associativo entre os valores estimados e os observados e, com o auxílio da análise de resíduos, permite inferir sobre a qualidade da estimativa. Quanto mais próxima de 1, maior a correlação entre as variáveis (BINOTI, 2010). Então:

$$r_{y\hat{y}} = \frac{\text{cov}(Y, \hat{Y})}{\sqrt{s^2(Y)s^2(\hat{Y})}}$$

Sendo que, *cov* é a covariância e s^2 é a variância.

A raiz do erro quadrado médio (RMSE) avalia a diferença quadrática média entre os valores observados e os valores estimados (BINOTI, 2010). Quanto menor o RMSE, melhor a precisão da estimativa (MEHTÄTALO et al., 2006), sendo:

$$\text{RMSE}\% = \frac{100}{\bar{Y}} * \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}}$$

Em que n é o número total de dados e \bar{Y} a média do volume observado.

3.4- Teste de equivalência

O teste de equivalência foi usado para determinar se as médias para a altura observada e a altura estimada pelas diferentes abordagens são suficientemente próximas para serem consideradas equivalentes.

Em um teste padrão de comparação de médias, como o teste t de Student por exemplo, a hipótese nula (H_0) supõe que a média (M) da amostra 1 é a mesma da amostra 2 ($M_1=M_2$), assim o ônus da prova recai sobre provar que a média de uma amostra difere da média da outra ($M_1 \neq M_2$). Já no teste de equivalência, a H_0 é que a média da amostra 1 difere da amostra 2 ($M_1 \neq M_2$), colocando assim o ônus da prova, em provar que a média de uma amostra é a mesma ou equivalente à da outra amostra ($M_1=M_2$).

O teste de equivalência é chamado de procedimento TOST (*Two One Sided T-test*) e consiste em dois testes unilaterais onde são definidos os limites de equivalência superior e inferior (com $p < 0,05$). A avaliação foi realizada no ambiente estatístico R, com a utilização do pacote “*TOSTER*” e o código “*TOSTtwo*”.

Na determinação dos parâmetros do teste é necessário definir o valor “d” do Cohen’s, que é a diferença entre duas médias expressa em desvios-padrão, conforme a fórmula:

$$D = \frac{M_1 - M_2}{S_p}$$

Onde: M_1 e M_2 são as médias da altura total observada e altura total estimada, e S_p é o desvio-padrão estimado para os dois grupos de altura e é dado por:

$$S_p = \sqrt{\frac{(N_1 - 1) \cdot S_1^2 + (N_2 - 1) \cdot S_2^2}{N_1 + N_2 - 2}}$$

No entanto, sua interpretação não é direta e os pesquisadores costumam usar diretrizes gerais, como pequeno (0,2), médio (0,5) e grande (0,8) ao interpretar um efeito.

Optou-se pela utilização de $d=0,2$ por ser um efeito pequeno, garantindo assim uma maior rigorosidade para o teste. Valores maiores que 0,2 acarretam num efeito maior e conseqüentemente tornam a análise menos rígida, portanto maiores diferenças entre as médias podem ser não significativas.

Durante a aplicação do teste também é possível calcular o poder do teste em função do tamanho da amostra (480 árvores) e o valor de Cohen's ($d=0,2$). O teste de equivalência apresentou um poder de ~85%, ou seja, para o teste tem-se 85% de chance de detectar o efeito (diferenças entre as médias do valor observado e valor estimado), caso ele aconteça.

3.5- Experimentos Avaliados

Neste estudo foram considerados quatro experimentos para avaliação da qualidade e precisão das estimativas como segue:

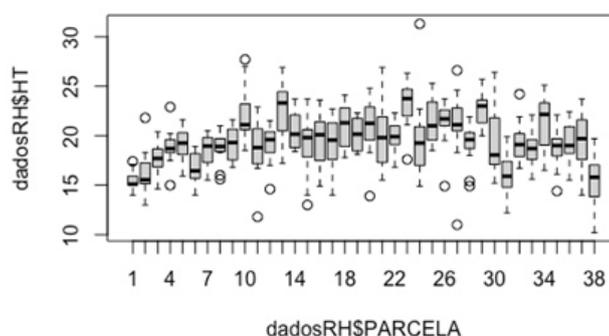
1. **Experimento 1:** Neste experimento, considerou-se as estimativas obtidas por cada abordagem sem a introdução de outliers;
2. **Experimento 2:** Foi introduzida na base de dados em duas árvores de 14 parcelas aleatorizadas, *outliers*, somando-se a altura observada 3 metros, valor equivalente a aproximadamente 1 desvio padrão das alturas observadas (Tabela 1).
3. **Experimento 3:** Foi introduzida na base de dados em duas árvores de 14 parcelas aleatorizadas, *outliers*, somando-se a altura observada 6 metros, valor equivalente a aproximadamente 2 desvios padrão das alturas observadas.
4. **Experimento 4:** Foi introduzida na base de dados em duas árvores de 14 parcelas aleatorizadas, *outliers*, somando-se a altura observada 9 metros, valor equivalente a aproximadamente 3 desvios padrão das alturas observadas.

4- Resultados e Discussão

4.1- Experimento 1

Gráficos que auxiliam na percepção do comportamento dos outliers são os boxplots. O Boxplot é um diagrama de caixa construído utilizando as referências de valores mínimos e máximos, primeiro e terceiro quartil, mediana e outliers da base dados, como mostra a figura 2.

Figura 2: Boxplot do conjunto de dados sem a presença de outlier.



Realizando uma análise de consistência na base de dados foi possível comprovar que apesar do gráfico indicar a presença de dados estranhos (pequenos círculos observados em algumas parcelas na Figura 2), eles se tratam de valores que fogem do comportamento padrão das parcelas dos dados de altura observados, mas não erros de medição.

A Tabela 2, mostra os resultados das estatísticas obtidas para cada abordagem de modelagem avaliada para os dados sem a introdução de *outliers*.

Tabela 2: Estatísticas das diferentes abordagens avaliadas.

Modelos	RMSE (m)	RMSE %	$r_{y\hat{y}}$
Método do Mínimos Quadrados (MMQ)	1,583	8,157	0,845
Regressão Quantílica (RQ)	1,641	8,455	0,833
Modelagem Mista (efeito aleatório nos 2 parâmetros – MM2)	1,643	8,463	0,833
Modelagem Mista (efeito aleatório no parâmetro b0 – MM0)	1,686	8,687	0,823
Modelagem Mista (efeito aleatório no b1 – MM1)	1,643	8,463	0,833
RNA	2,011	10,359	0,734

Onde: $r_{y\hat{y}}$ = correlação entre altura observada e altura estimada.

Ao analisar os resultados, verifica-se que todas as abordagens de regressão retornaram RMSE% inferiores a 10%, correlações acima 0,80. O erro quadrático médio variou para essas abordagens entre 1,583 (mínimos quadrados) e 1,686 m (modelagem mista com efeito aleatório no b0). A RNA gerou erro maior (2,011 m) e correlação inferior (~0,7).

Os resultados relacionados às redes durante o treinamento e validação (generalização), estão apresentados na Tabela 3.

Tabela 3: Estatísticas do treino e validação das RNA.

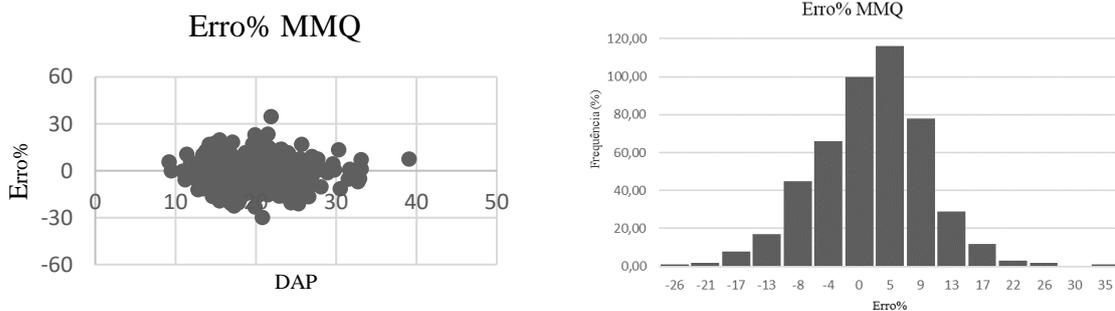
Etapas	RMSE	RMSE%	$r_{y\hat{y}}$
Treino	1,945	9,948	0,765
Validação	2,552	13,367	0,638

Onde: $r_{y\hat{y}}$ = correlação entre altura observada e altura estimada.

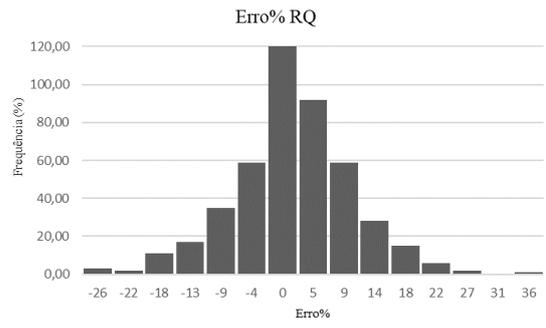
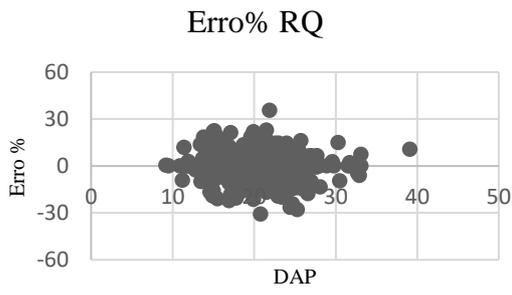
Observa-se para o treinamento RMSE% inferior a 10%, contudo ao generalizar a RNA passa a apresentar RMSE% superior a 10%. Esse erro para o treinamento e validação são de 1,945 e 2,552 m, respectivamente.

As figuras que seguem apresentam a dispersão gráfica dos resíduos e histograma de frequência dos erros para cada abordagem avaliada (Figura 3) e das etapas de treino e validação da RNA (Figura 4), para os dados sem outliers.

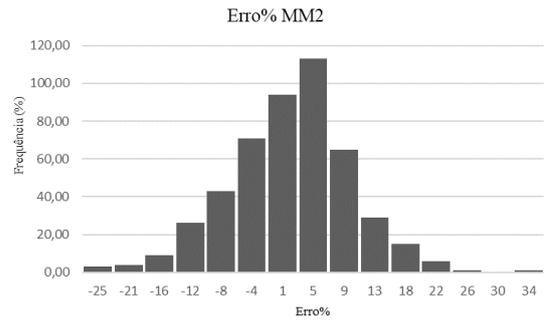
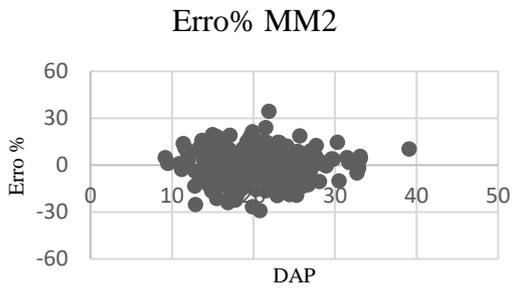
Figura 3: Dispersão gráfica dos resíduos e histograma de frequência dos erros.



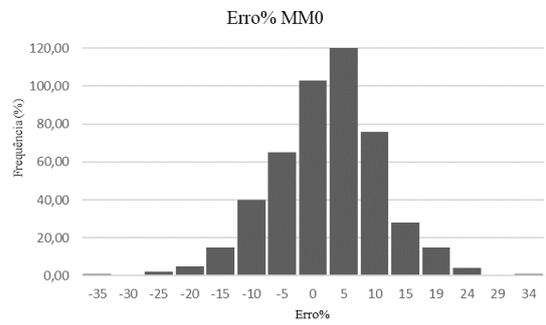
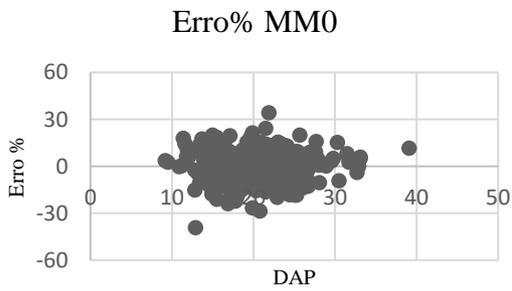
A



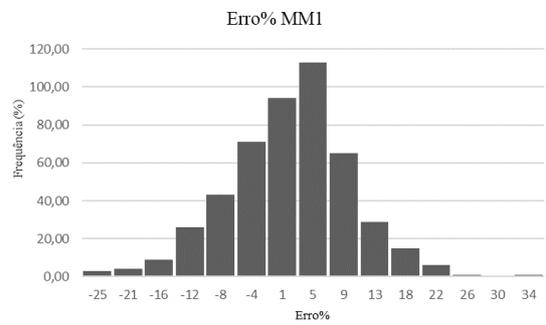
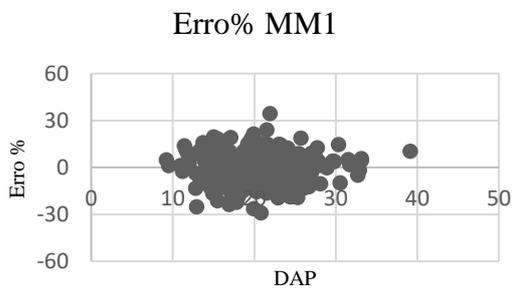
B



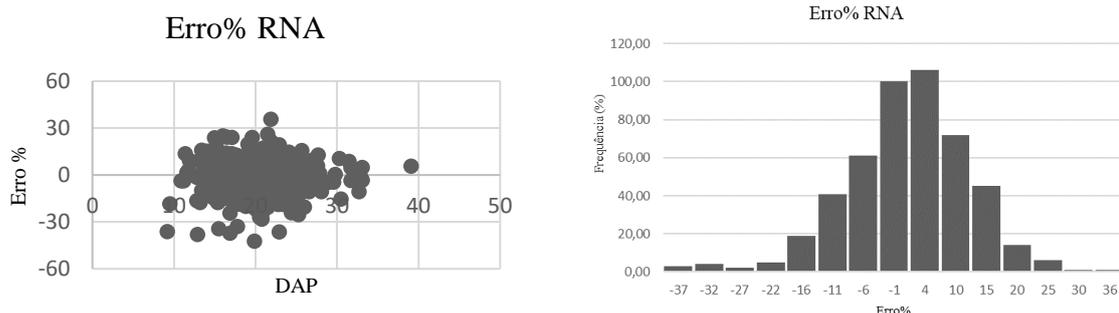
C



D



E



F

Onde: MMQ: Método dos Mínimos Quadrados(A); RQ: Regressão Quantílica(B); MM2:Modelo Misto com efeito aleatório nos dois parâmetros(C); MM0: Modelo Misto com efeito aleatório no parâmetro b_0 (D); MM1: Modelo Misto com efeito aleatório no parâmetro b_1 (E); RNA: Redes Neurais Artificiais(F).

O método que gerou os melhores resultados foi o dos Mínimos Quadrados, apresentando RMSE% mais próximo de zero e uma correlação maior que os demais. Constatou-se que os valores de correlação foram muito próximos entre as abordagens. Por outro lado, a RNA apresenta o pior resultado, o que não era esperado pois alguns estudos apontam essa abordagem como uma opção precisa para estimativa de altura e semelhante a outras abordagens tradicionais (BINOTI, 2010; BINOTI, 2012; LIMA, 2014; BINOTI et al., 2014).

Pode-se perceber que a dispersão residual entre todas as abordagens foi bastante semelhante.

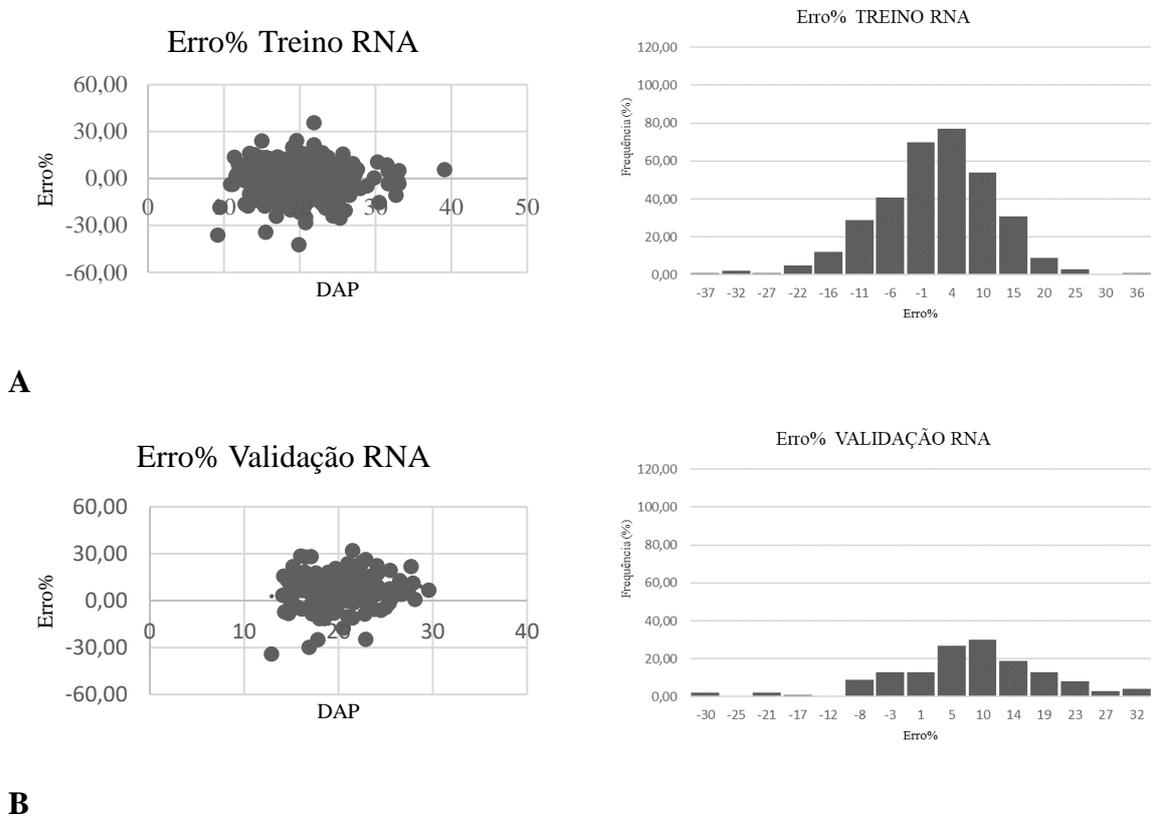
Para a RNA, ao avaliar graficamente sua dispersão residual, é possível verificar que essa abordagem apresenta tendenciosidade em superestimar a altura (maiores valores de erros negativos) (Figura 3 F). Essa tendenciosidade também pode ser verificada no histograma de frequências que apresentou uma assimetria a esquerda devido aos altos valores de erros negativos observados para algumas árvores. O mesmo comportamento foi verificado para os resultados de treinamento e validação da RNA (Figura 4).

As RNA, de acordo com a literatura, são mais tolerantes a ruídos, no entanto para a base de dados em tela, verifica-se que a rede apresenta resultados inferiores ao modelo de Curtis, ajustado por regressão para todos os métodos.

Um detalhe relevante, considerado apenas para a geração das RNA, é que diferentemente das demais abordagens, as RNA em todos os experimentos foram treinadas com o DAP como variável de entrada e a HT como variável de saída. Não houve o controle de

parcela como já mencionado anteriormente. O controle da parcela atua como uma estratificação da base de dados, homogeneizando as variâncias e conseqüentemente diminuindo o erro de estimativa, mas no contexto de RNA não faria sentido esta estratificação pois havia um número muito pequeno de observações em cada parcela, insuficiente para uma divisão adequada de dados de treinamento e validação.

Figura 4: Dispersão gráfica dos resíduos e histograma de frequência dos erros para treino e validação das RNA's.

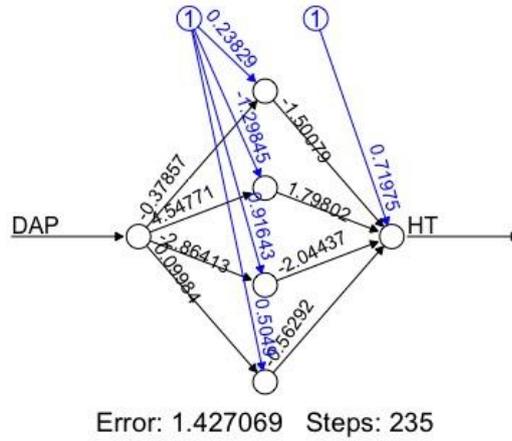


Onde: Etapa de treino das Redes Neurais Artificiais(A); Etapa de validação das Redes Neurais Artificiais(B);

Todas os métodos apresentaram erros percentuais concentrados, nos intervalos de -40 a 40%.

A Figura 5 apresenta a RNA treinada neste experimento com os respectivos pesos sinápticos

Figura 5 – RNA para o experimento 1

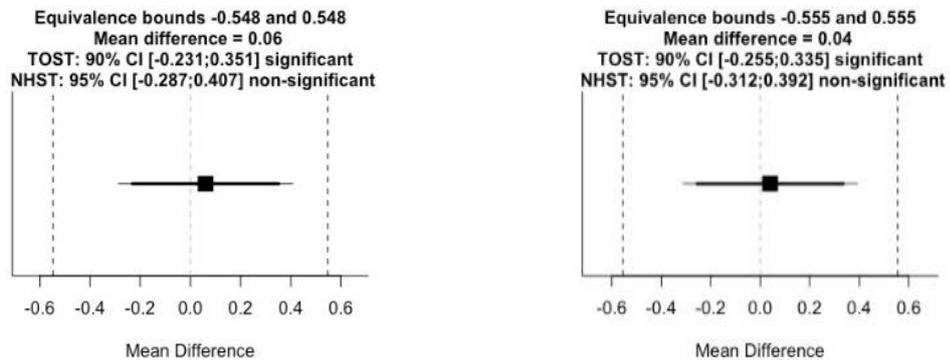


Na figura é possível observar que foi utilizada apenas uma camada oculta, contando com quatro neurônios, com a variável de entrada DAP, e a variável de saída HT. O comportamento da RNA contou com 235 passos e tendo sua parada quando alcançou um erro mínimo global igual a 1,4270.

Teste de equivalência para o experimento 1

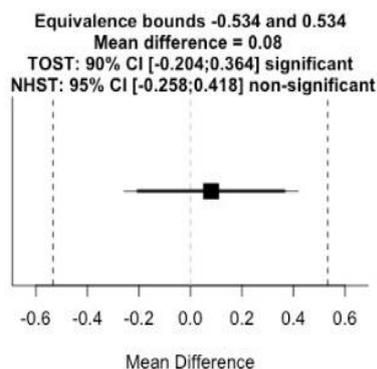
A Figura 6 apresenta o resultado do teste de equivalência para cada abordagem avaliada no experimento 1.

Figura 6 – Teste de equivalência para o experimento 1.

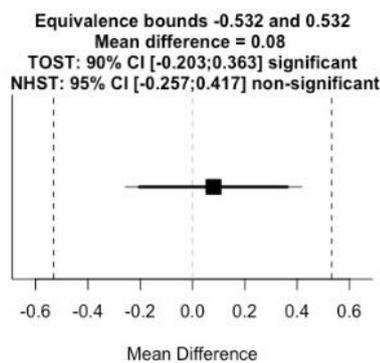


A

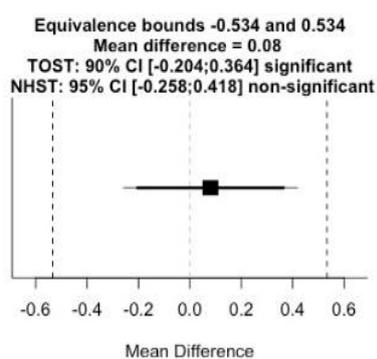
B



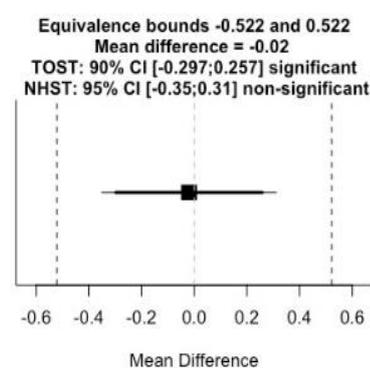
C



D



E



F

Onde: Método dos Mínimos Quadrados(A); Regressão Quantílica(B); Modelo Misto com efeito aleatório nos dois parâmetros(C); Modelo Misto com efeito aleatório no parâmetro b_0 (D); Modelo Misto com efeito aleatório no parâmetro b_1 (E); Redes Neurais Artificiais(F).

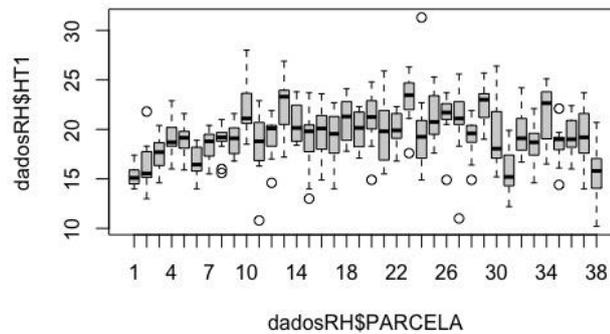
Os testes de equivalência partem da premissa que os dados comparados são diferentes, portanto quando seu comportamento permanece dentro dos limites de equivalência, significa que os parâmetros comparados são estatisticamente similares.

Verifica-se (Figura 6) que as alturas médias estimadas para todas as abordagens foram estatisticamente equivalentes à altura média observada, ou seja, seu efeito é semelhante a zero (cruza a linha zero) e também não é relevante pois está dentro do intervalo de equivalência. A diferença média variou entre $-0,02$ (RNA) a $+0,08$ (Modelagem Mista). A RNA embora tenha apresentado os piores resultados estatísticos de modelagem, foi a abordagem que apresentou a média da altura estimada mais próxima da média da altura observada, o que pode ser efeito da lei dos grandes números.

4.2- Experimento 2

Observando-se a figura 7 é possível verificar o comportamento das parcelas quando *outliers* foram inseridos. De maneira geral, observa-se para as parcelas com a presença de *outliers* neste experimento, um sutil aumento na sua mediana, aumento do intervalo interquartílico, o que denota maior dispersão, quando comparado ao boxplot para os dados sem *outliers* (Figura 2). Ressalta-se que neste experimento foram somados a duas árvores de 14 parcelas aleatorizadas, um erro de 3 metros na altura total.

Figura 7: Boxplot do conjunto de dados com a presença de *outlier* no experimento 2.



A Tabela 4, mostra os resultados das estatísticas obtidas para cada abordagem de modelagem avaliada para os dados com a introdução de *outliers* (3 m).

Tabela 4: Estatísticas das diferentes abordagens avaliadas no experimento 2.

Modelos	RMSE	RMSE%	$r_{\hat{y}y}$
Métodos dos Mínimos Quadrados	1,612	8,306	0,839
Regressão Quantílica	1,661	8,556	0,828
Modelos Mistos (efeito aleatório nos 2 parâmetros)	1,653	8,517	0,832
Modelos Mistos (efeito aleatório no parâmetro b0)	1,707	8,795	0,819
Modelos Mistos (efeito aleatório no b1)	1,653	8,517	0,832
RNA	2,049	10,556	0,726

Onde: $r_{\hat{y}y}$ = correlação entre altura observada e altura estimada.

Verifica-se que todas as abordagens de regressão continuaram retornando RMSE% inferiores a 10% e correlações acima 0,80, exceto a RNA. O erro quadrático médio variou para essas abordagens entre 1,612 (método dos mínimos quadrados) e 1,707 m (modelagem mista

com efeito aleatório no b0). A RNA gerou erro maior (2,049 m) e correlação inferior ($\sim 0,7$), no entanto, há uma ligeira piora nas estatísticas quando comparadas as do experimento 1 (dados sem *outliers*). Isso denota o impacto que a presença de valores estranhos provoca na modelagem, independentemente da abordagem escolhida. Observa-se ainda que os métodos menos afetados foram a regressão quantílica e a RNA.

Os resultados relacionados às redes durante o treinamento e validação (generalização), estão apresentados na Tabela 5.

Tabela 5: Estatísticas do treino e validação das RNA no experimento 2.

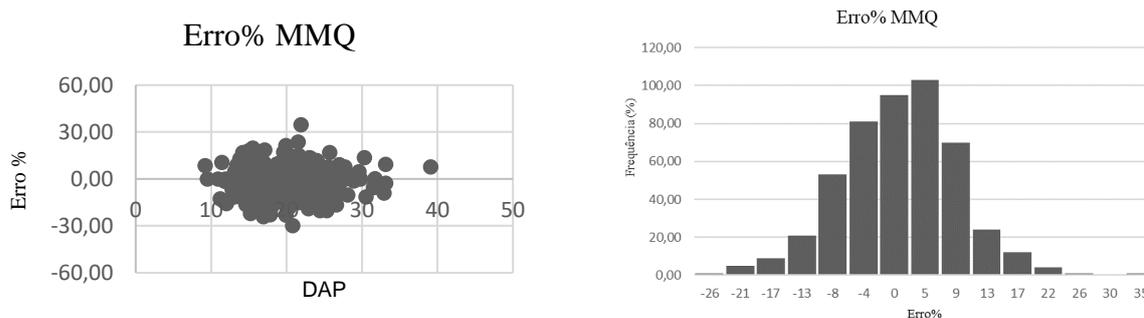
Etapas	RMSE	RMSE%	$r_{y\hat{y}}$
Treino (1)	2,599	13,288	0,723
Validação (1)	2,998	15,268	0,685

Onde: $r_{y\hat{y}}$ = correlação entre altura observada e altura estimada.

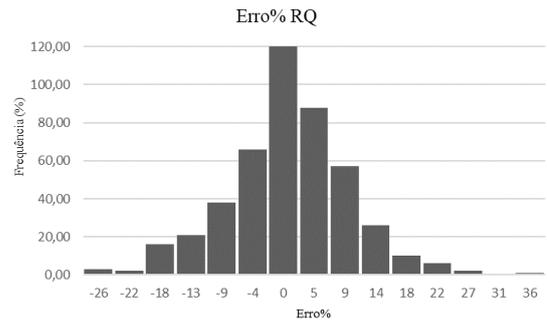
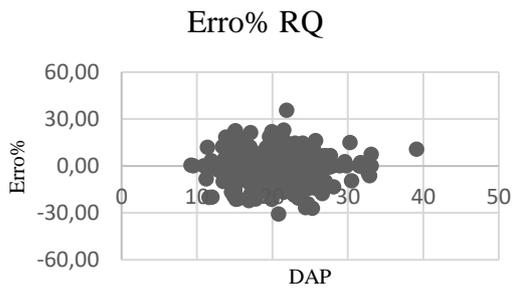
Observa-se para o treinamento e validação, RMSE% superior a 10%, tendo a generalização um erro maior que o treinamento. Esse erro para o treinamento e validação são de 2,599 e 2,998 m, respectivamente.

As figuras que seguem apresentam a dispersão gráfica dos resíduos e histograma de frequência dos erros para cada uma das abordagens (Figura 8) e para as etapas de treino e validação da RNA (Figura 9).

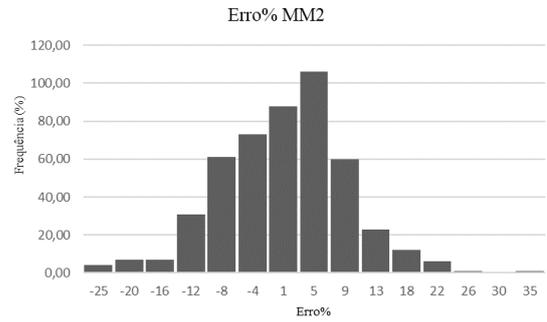
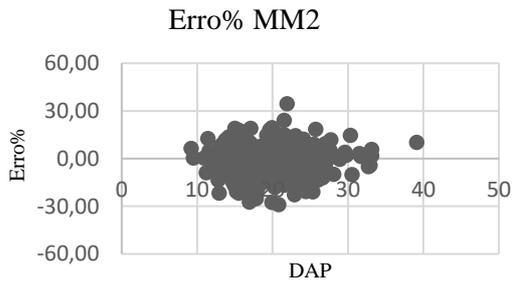
Figura 8: Dispersão gráfica dos resíduos e histograma de frequência dos erros.



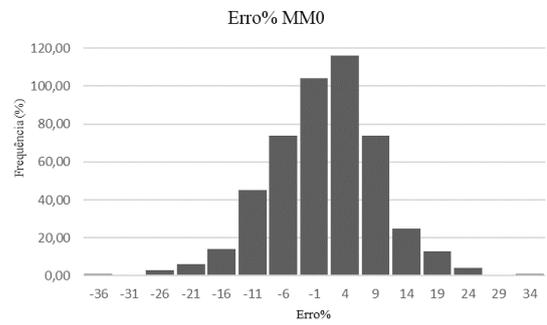
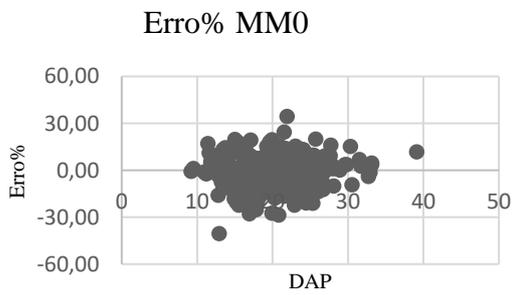
A



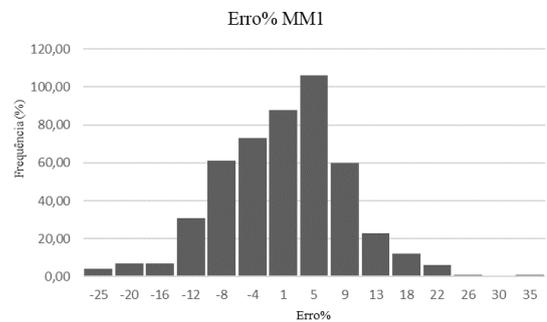
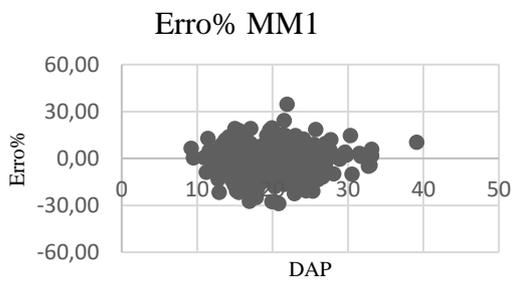
B



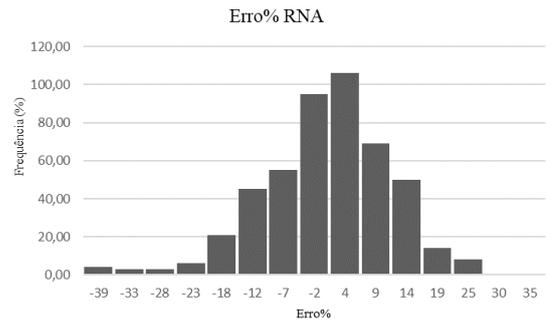
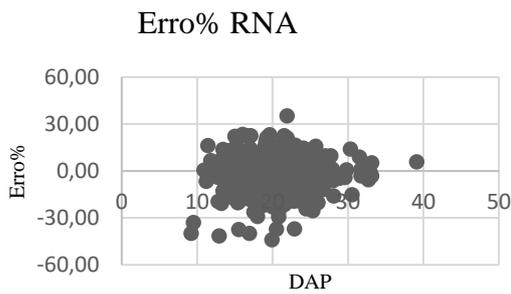
C



D



E



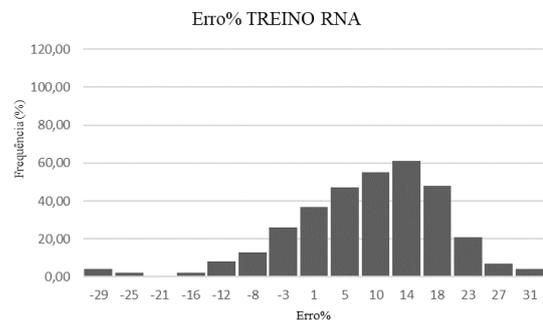
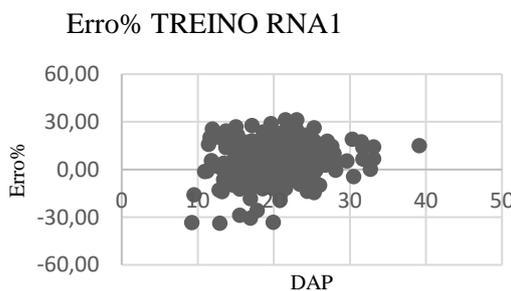
F

Onde: MMQ: Método dos Mínimos Quadrados(A); RQ: Regressão Quantílica(B);MM2: Modelo Misto com efeito aleatório nos dois parâmetros(C); MM0: Modelo Misto com efeito aleatório no parâmetro b0(D); MM1: Modelo Misto com efeito aleatório no parâmetro b1(E); RNA: Redes Neurais Artificiais(F).

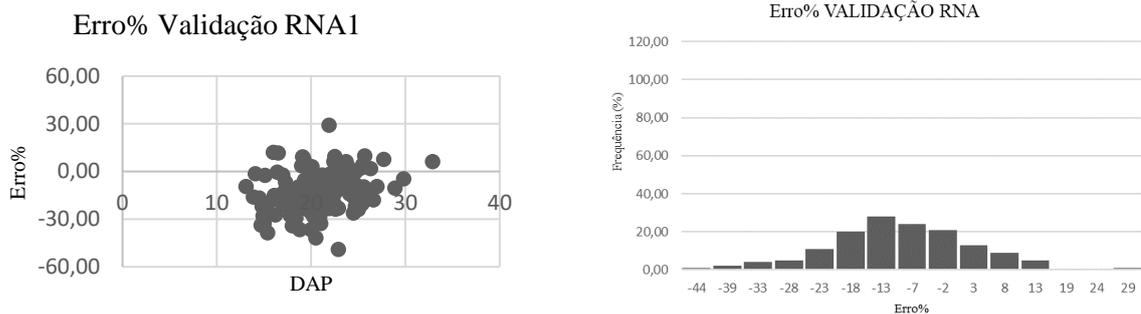
Mesmo inserindo o *outlier* de 3 metros, a abordagem dos Mínimos Quadrados continua apresentando os melhores resultados. Verificou-se RMSE% mais próximo de zero e uma correlação maior que as demais. A dispersões residuais foram semelhantes, ficando entre -40 e 40%.

Nota-se que ao inserir os *outliers* o histograma de frequência dos erros mostrou tendenciosidade (maiores erros negativos) para a maioria das abordagens, isso mostra que, com os *outliers* há uma tendência em superestimar as alturas, exceto na regressão quantílica.

Figura 9: Dispersão gráfica dos resíduos e histograma de frequência dos erros para treino e validação das RNA's para o experimento 2.



A



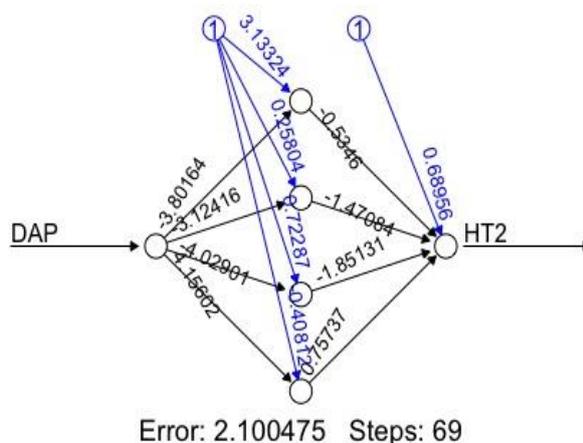
B

Onde: Treino RNA: Etapa de treino das Redes Neurais Artificiais(A);Validação RNA: Etapa de validação das Redes Neurais Artificiais(B);

Os erros se concentraram entre - 50% e 30% e houve uma grande diferença comportamental em comparação às frequências dos erros para treinamento e validação. De forma geral, as RNA do tipo Multilayer Perceptron (MLP), evidenciam as maiores tendenciosidades na fase de validação em comparação à fase de treinamento das redes.

Figura 10 – RNA para o experimento 2

Na Figura 10, pode-se observar a rede gerada na fase de treinamento para o experimento 2 com seus respectivos pesos sinápticos.

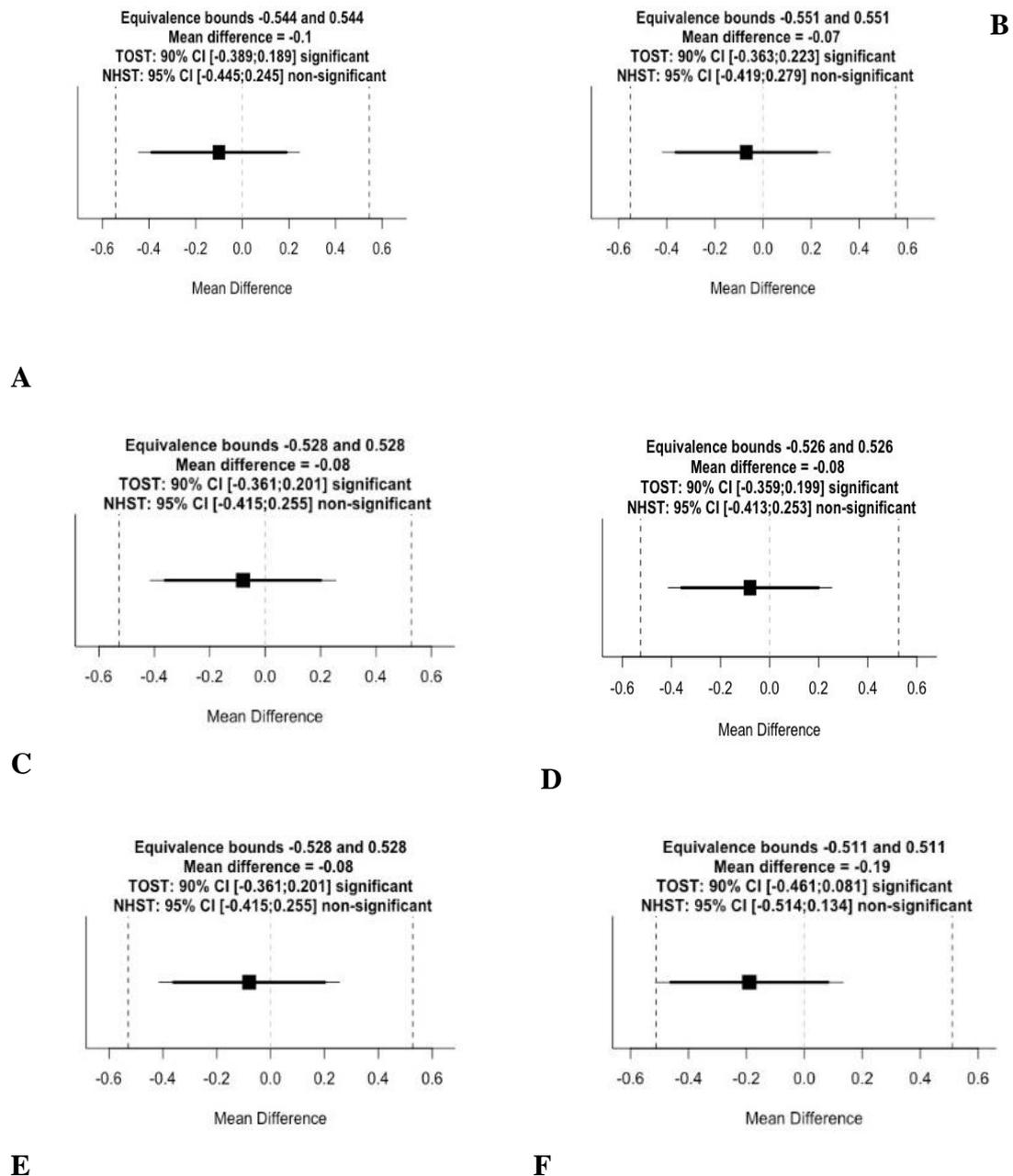


O comportamento da RNA contou com 69 passos e tendo sua parada quando alcançou um erro mínimo global igual a 2,1004.

Teste de equivalência para o experimento 2

A Figura 11 apresenta o resultado do teste de equivalência para cada abordagem avaliada no experimento 2.

Figura 11 – Teste de equivalência



Onde: Método dos Mínimos Quadrados(A); Regressão Quantílica(B); Modelo Misto com efeito aleatório nos dois parâmetros(C); Modelo Misto com efeito aleatório no parâmetro b_0 (D); Modelo Misto com efeito aleatório no parâmetro b_1 (E); Redes Neurais Artificiais(F).

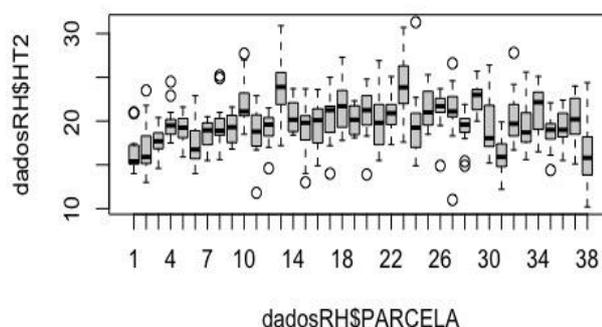
Ao se inserir os *outliers* (3 m) as alturas médias estimadas pelas abordagens se mostraram estatisticamente equivalentes à média da altura observada, ou seja, seu efeito é semelhante a zero (cruza a linha zero) e também não é relevante pois está dentro do intervalo de equivalência.

A regressão quantílica apresentou a menor diferença (-0,07 m) e a RNA a maior (-0,19 m). Este resultado denota que a regressão quantílica na presença de *outliers* parece ser menos afetada quando comparada às demais abordagens avaliadas.

4.3- Experimento 3

Observando-se a Figura 12 é possível verificar o comportamento das parcelas quando *outliers* foram inseridos. Observa-se para as parcelas com a presença de *outliers* neste experimento, um aumento na sua mediana, aumento do intervalo interquartílico, o que denota maior dispersão, quando comparado ao boxplot para os dados sem *outliers* (Figura 2). Ressalta-se que neste experimento foram somados a duas árvores de 14 parcelas aleatorizadas, um erro de 6 metros na altura total.

Figura 12: Boxplot do conjunto de dados com a presença de *outlier* no experimento 3.



A Tabela 6, mostra os resultados das estatísticas obtidas para cada abordagem de modelagem avaliada para os dados neste experimento.

Tabela 6: Estatísticas das diferentes abordagens avaliadas no experimento 3.

Modelos	RMSE	RMSE%	$r_{y\hat{y}}$
Método dos Mínimos Quadrados	1,697	8,745	0,823
Regressão Quantílica	1,707	8,794	0,818
Modelos Mistos (efeito aleatório nos 2 parâmetros)	1,710	8,813	0,823
Modelos Mistos (efeito aleatório no parâmetro b0)	1,758	9,058	0,810
Modelos Mistos (efeito aleatório no b1)	1,710	8,813	0,822
RNA	2,185	11,258	0,706

Onde: $r_{y\hat{y}}$ = correlação entre altura observada e altura estimada.

As correlações diminuíram, isso se deve ao fato de os valores estranhos terem aumentado de magnitude afetando a média das alturas o que consequentemente impactou as estatísticas de modelagem.

A abordagem de melhor resultado foi a dos Mínimos Quadrados com RMSE de 1,697 metros e uma correlação de 0,823.

A RNA gerou erro maior (2,185 m) e correlação inferior (~0,7). Comportamento similar aos outros experimentos.

Os resultados relacionados às redes durante o treinamento e validação (generalização), estão apresentados na Tabela 7.

Tabela 7: Estatísticas do treino e validação das RNA no experimento 3.

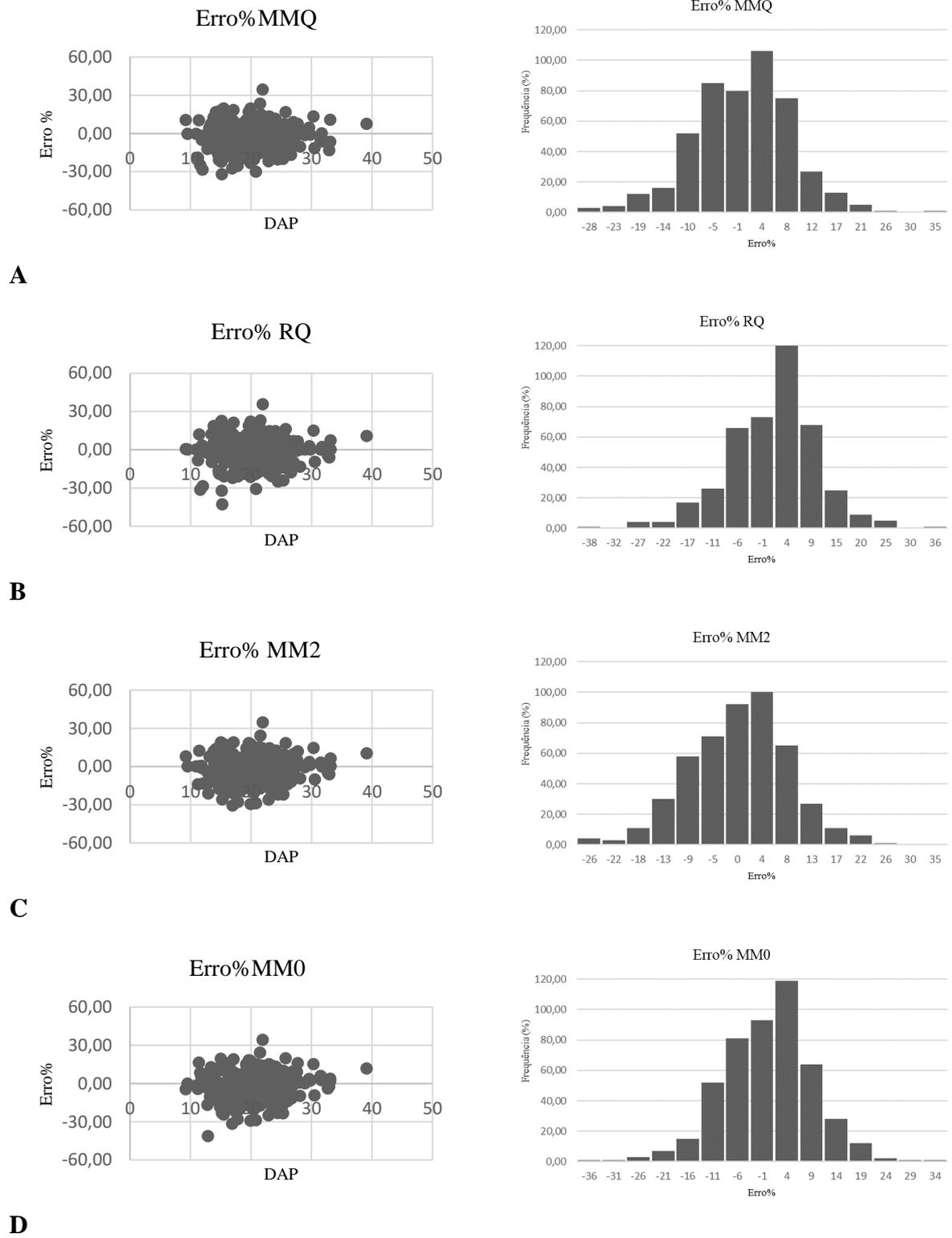
Etapas	RMSE	RMSE%	$r_{y\hat{y}}$
Treino	2,5204	12,7427	0,6650
Validação	2,4673	12,5136	0,6399

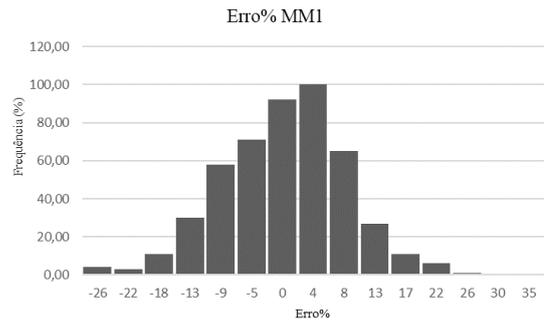
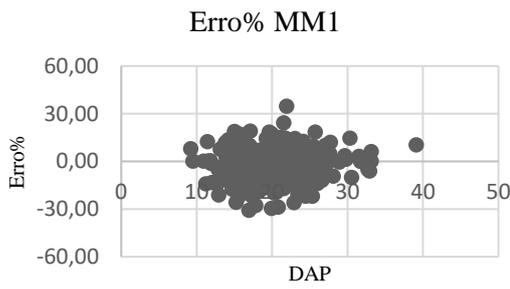
Onde: $r_{y\hat{y}}$ = correlação entre altura observada e altura estimada.

Observa-se para o treinamento e validação, RMSE% superior a 10%, tendo a generalização um erro maior que o treinamento. Esse erro para o treinamento e validação são de 2,5204 e 2,4673 m, respectivamente.

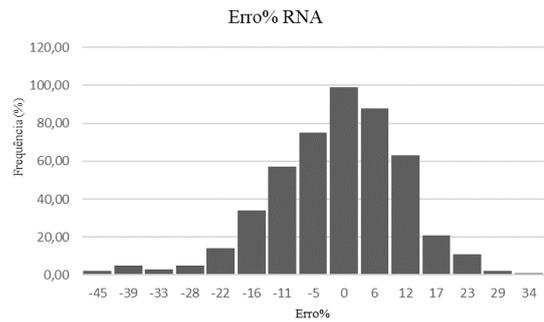
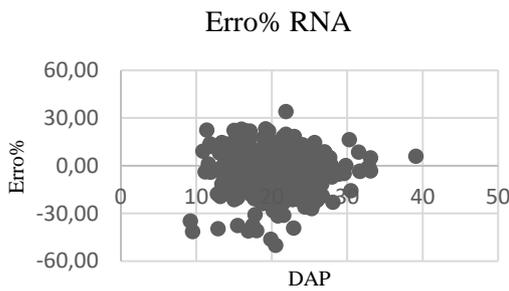
As Figuras 13 e 14 apresentam a dispersão gráfica dos resíduos e histograma de frequência dos erros para cada uma das abordagens e para as etapas de treino e validação da RNA, respectivamente.

Figura 13: Dispersão gráfica dos resíduos e histograma de frequência dos erros.





E

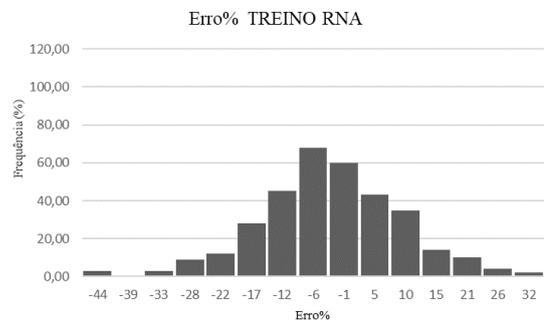
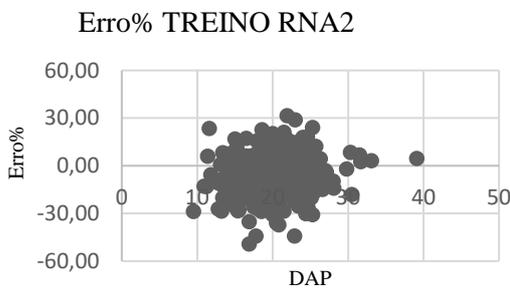


F

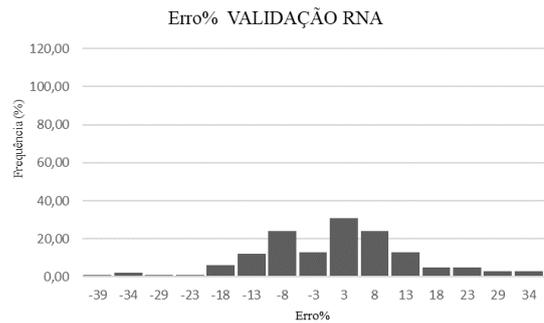
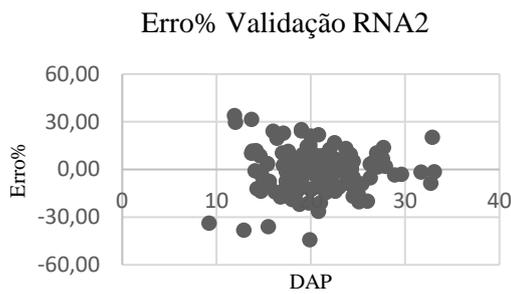
Onde: MMQ: Método dos Mínimos Quadrados(A); RQ: Regressão Quantílica(B); MM2: Modelo Misto com efeito aleatório nos dois parâmetros(C); MM0: Modelo Misto com efeito aleatório no parâmetro b0(D); MM1: Modelo Misto com efeito aleatório no parâmetro b1(E); RNA: Redes Neurais Artificiais(F).

O método que gerou os melhores resultados foi o dos Mínimos Quadrados, apresentando RMSE% mais próximo de zero e uma correlação maior que os demais.

Figura 14: Dispersão gráfica dos resíduos e histograma de frequência dos erros para treino e validação no experimento 3.



A



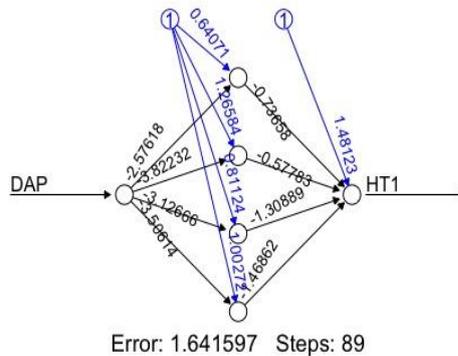
B

Onde: Treino RNA: Etapa de treino das Redes Neurais Artificiais(A);Validação RNA: Etapa de validação das Redes Neurais Artificiais(B);

A validação apresenta esse comportamento por utilizar apenas 30% da base de dados. É possível observar na etapa de treino o comportamento da camada de entrada para a camada de saída, passando por uma camada oculta (Figura 15).

Figura 15 – RNA treinada para o experimento 3

Na Figura 15, pode-se observar a rede gerada na fase de treinamento para o experimento 3 com seus respectivos pesos sinápticos.

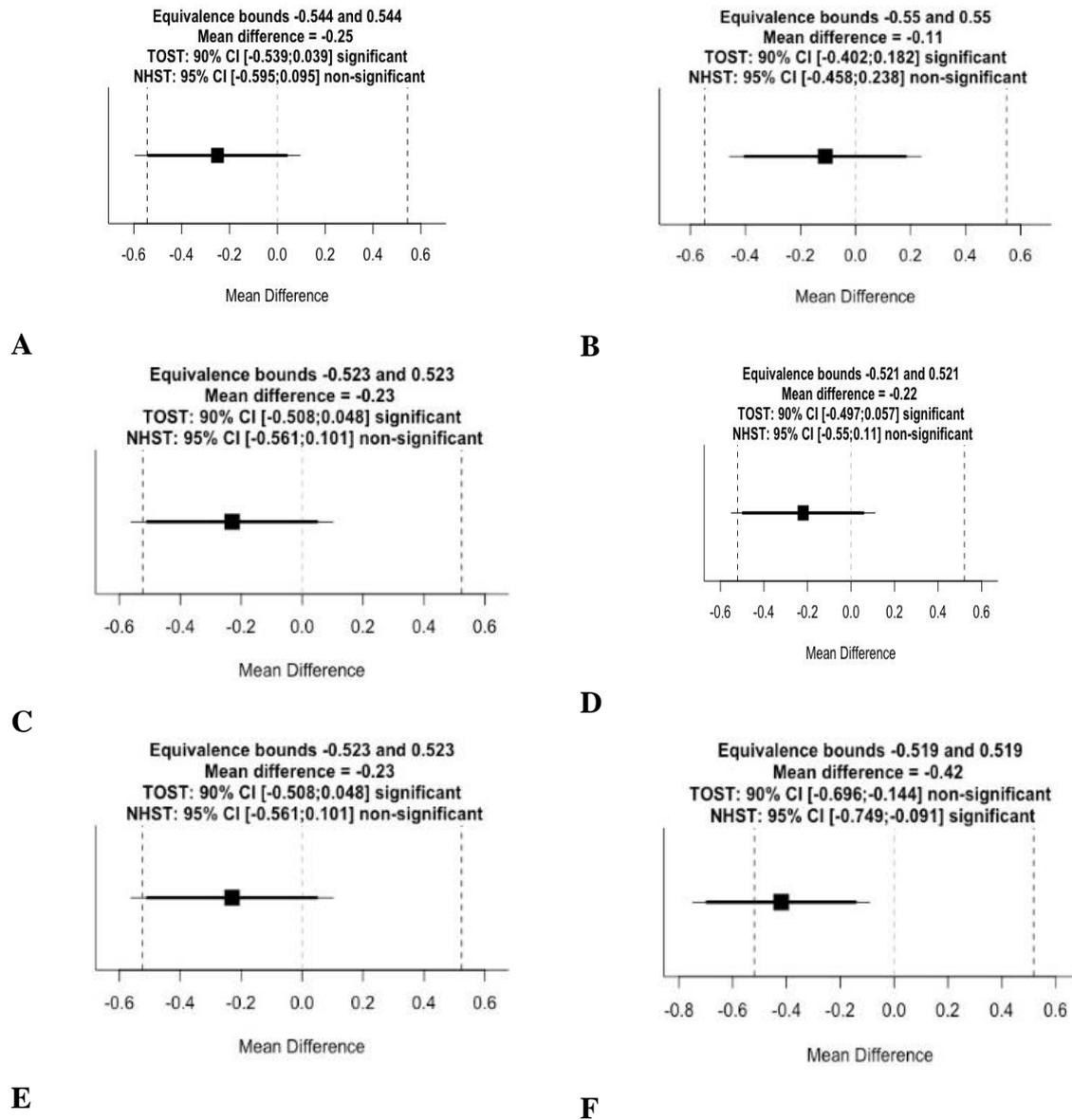


O comportamento da RNA contou com 89 passos, tendo sua parada quando alcançou um erro igual a 1,6415.

Teste de equivalência para o experimento 3

Em relação ao teste de equivalência entre os métodos, foi confrontada a média de cada abordagem com a média da altura observada.

Figura 16 – Teste de equivalência



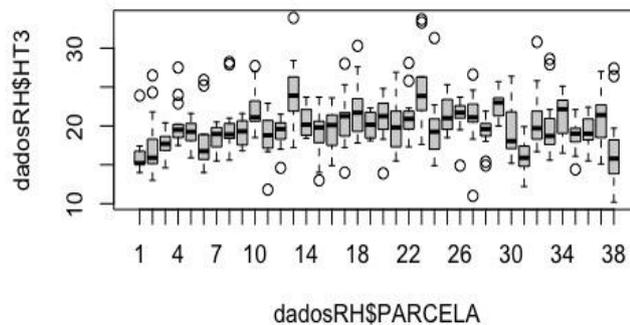
Onde: Métodos dos Mínimos Quadrados(A); Regressão Quantílica(B); Modelo Misto com efeito aleatório nos dois parâmetros(C); Modelo Misto com efeito aleatório no parâmetro b0(D); Modelo Misto com efeito aleatório no parâmetro b1(E); Redes Neurais Artificiais(F).

Com o *outlier* todas as médias das alturas estimadas por todas as abordagens se mostraram estatisticamente similares à altura observada, exceto a RNA, onde se pode observar a figura fora do linear aceito como estatisticamente igual. Isso pode ter ocorrido devido a não estratificação dos dados para as redes neurais, o que acabou acarretando em um maior erro.

4.4- Experimento 4

Observando a figura 17 é possível verificar o comportamento das parcelas quando os *outliers* foram inseridos. Observa-se para este experimento, um aumento no número de círculos vazios no boxplot, o que denota maior número de *outliers*, quando comparado ao *boxplot* para os dados sem *outliers* (Figura 2). Ressalta-se que neste experimento foram somados a duas árvores de 14 parcelas aleatorizadas, um erro de 9 metros na altura total.

Figura 17: Boxplot do conjunto de dados com a presença de *outlier* no experimento 3.



É possível avaliar que a medida que os valores de *outlier* foram aumentando mais os círculos se afastaram em relação às caudas da caixa, o que indica que os valores dos outliers presentes estão maiores que nos outros experimentos.

A Tabela 8, mostra os resultados das estatísticas obtidas para cada abordagem de modelagem avaliada para os dados do experimento 4.

Tabela 8: Estatísticas das diferentes abordagens avaliadas no experimento 4.

Modelos	RMSE	RMSE%	$r_{y\hat{y}}$
Método dos Mínimos Quadrados	1,8217	9,3849	0,7995
Regressão Quantílica	1,7478	9,0043	0,8088
Modelos Mistos (efeito aleatório nos 2 parâmetros)	1,7669	9,1027	0,8132
Modelos Mistos (efeito aleatório no parâmetro b0)	1,8102	9,3258	0,8023
Modelos Mistos (efeito aleatório no b1)	1,7669	9,1027	0,8132
RNA	2,3596	12,1548	0,6776

Onde: $r_{y\hat{y}}$ = correlação entre altura observada e altura estimada.

Neste experimento o método dos mínimos quadrados que se destacou nos experimentos anteriores, apresentou o pior resultado dentre as diferentes técnicas de regressão.

A regressão quantílica se mostrou o melhor método com a presença de valores estranhos de maior magnitude, com RMSE de 1,7478m e uma correlação de 0,8088.

Nas etapas de validação é possível notar o desempenho das redes nas fases de processamento, que mostram uma forte piora na modelagem neste experimento.

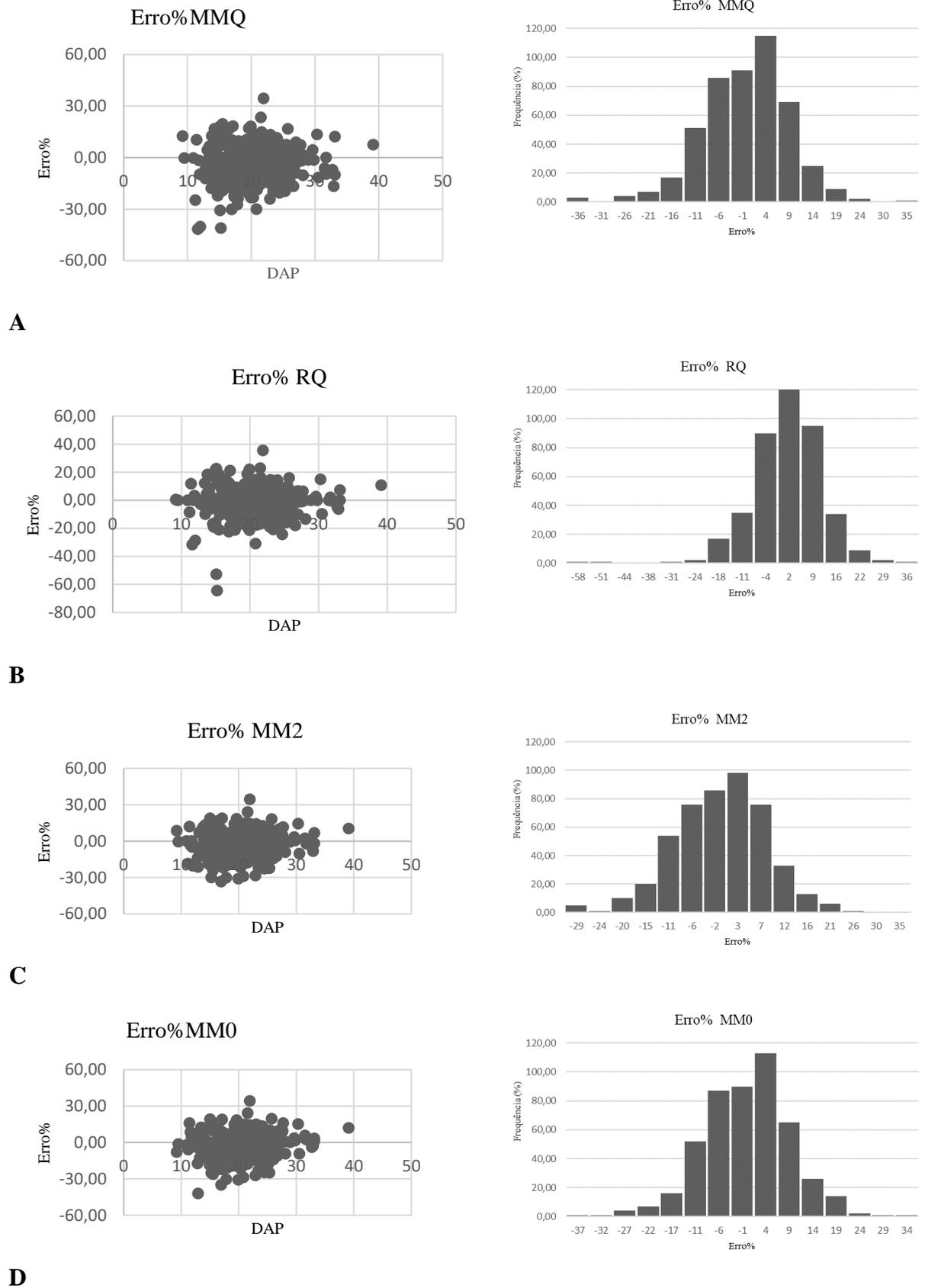
Tabela 9: Estatísticas do treino e validação das RNA no experimento 4.

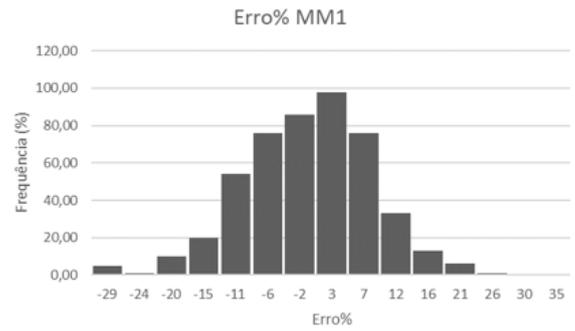
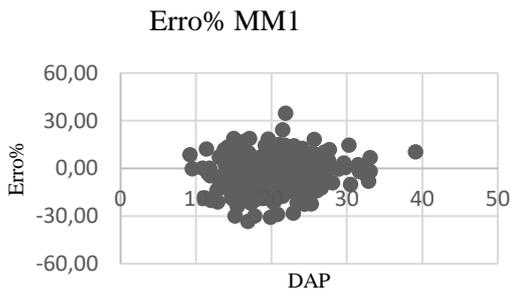
Etapas	RMSE	RMSE%	$r_{y\hat{y}}$
Treino	2,7988	14,1237	0,6100
Validação	2,9035	14,4078	0,5326

Onde: $r_{y\hat{y}}$ = correlação entre altura observada e altura estimada.

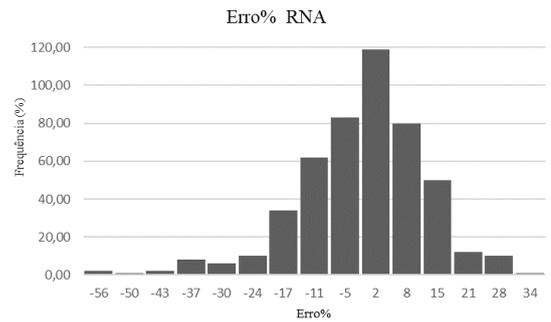
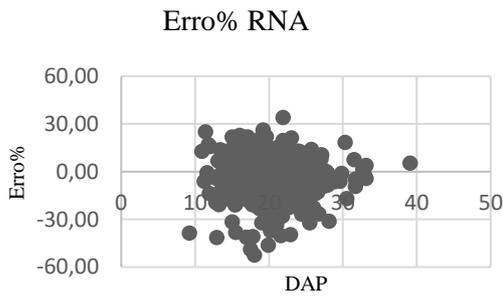
As figuras que seguem apresentam a dispersão gráfica dos resíduos e histograma de frequência dos erros para cada uma das abordagens (Figura 18) e para as etapas de treino e validação da RNA (Figura 19).

Figura 18: Dispersão gráfica dos resíduos e histograma de frequência dos erros.





E

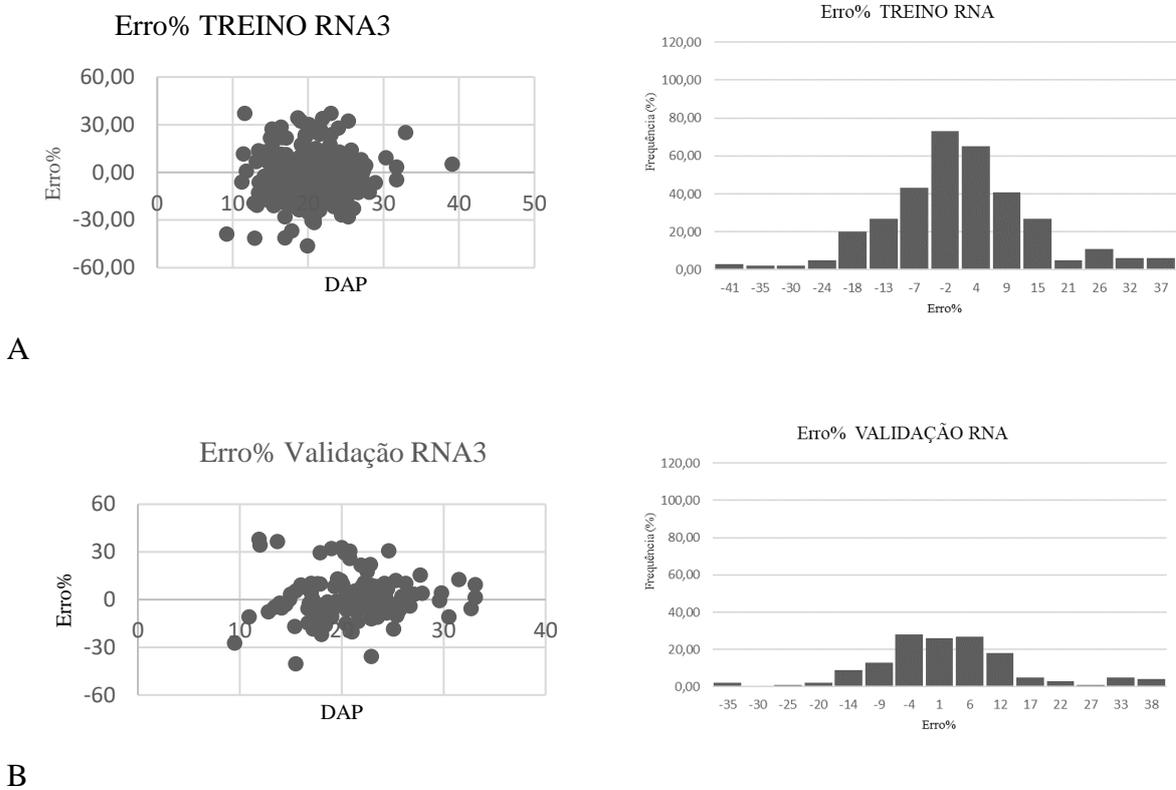


F

Onde: MMQ: Método dos Mínimos Quadrados(A); RQ: Regressão Quantílica(B); MM2:Modelo Misto com efeito aleatório nos dois parâmetros(C); MM0: Modelo Misto com efeito aleatório no parâmetro b0(D); MM1: Modelo Misto com efeito aleatório no parâmetro b1(E); RNA: Redes Neurais Artificiais(F).

A Regressão Quantílica apresentou a melhor dispersão. Com a inserção dos *outliers* deste experimento, os modelos apresentaram tendência em superestimar as alturas, porém a RQ conseguiu manter boas estatísticas.

Figura 19: Dispersão gráfica dos resíduos e histograma de frequência dos erros para treino e validação das RNA's com a presença de *outliers*.

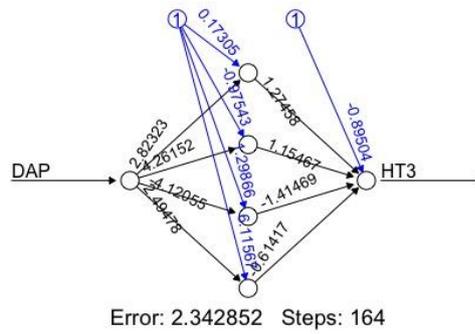


Onde: Treino RNA: Etapa de treino das Redes Neurais Artificiais(A);Validação RNA: Etapa de validação das Redes Neurais Artificiais(B);

Houve uma grande diferença comportamental em comparação às frequências dos erros para treinamento e validação. De forma geral, as RNA do tipo Multilayer Perceptron (MLP), apresentaram maiores tendências na fase de treinamento em comparação à fase de validação das redes. Isto pode ser observado pelo caráter dos histogramas.

É possível observar na etapa de treino o comportamento da camada de entrada para a camada de saída, passando por uma camada oculta a rede gerada na fase de treinamento para o experimento 4 com seus respectivos pesos sinápticos (Figura 20).

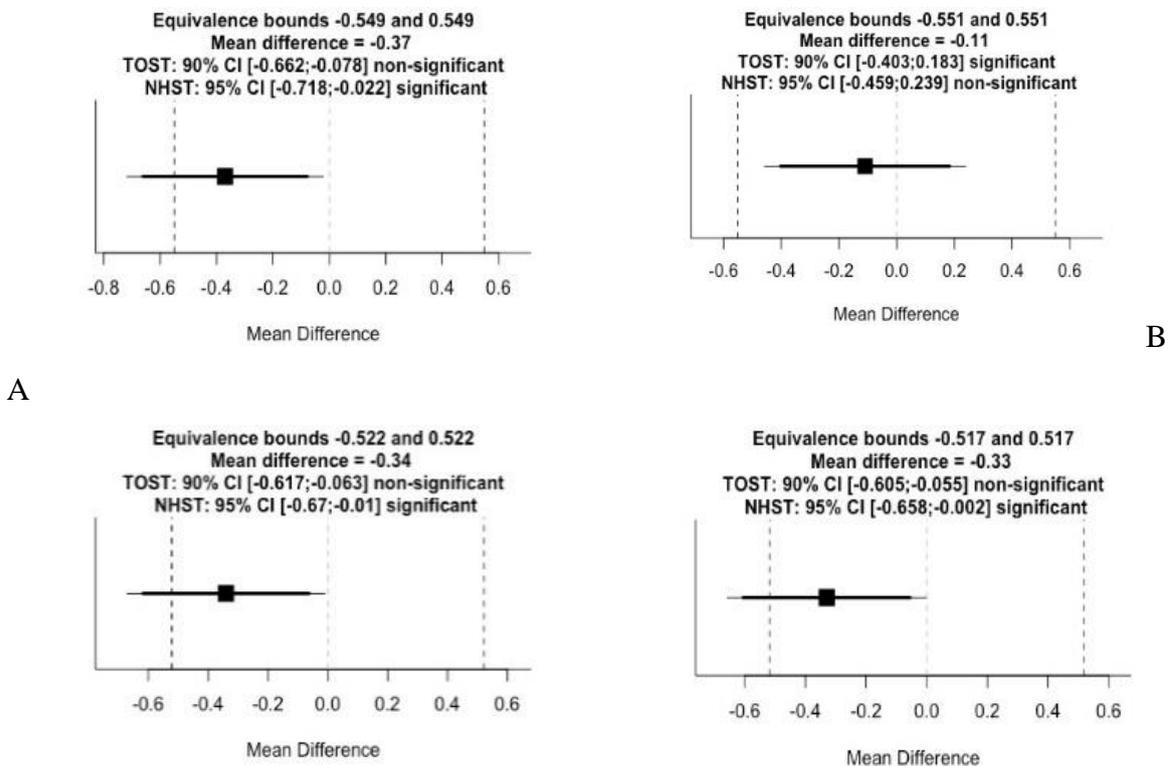
Figura 20 – RNA treinda para o experimento 4



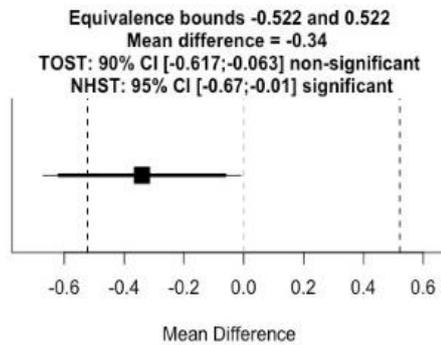
O treinamento da RNA contou com 164 passos e tendo sua parada quando alcançou um erro mínimo global igual a 2,342. É possível também observar os valores dos pesos sinápticos, que variam de acordo o comportamento das redes.

Teste de equivalência para o experimento 4

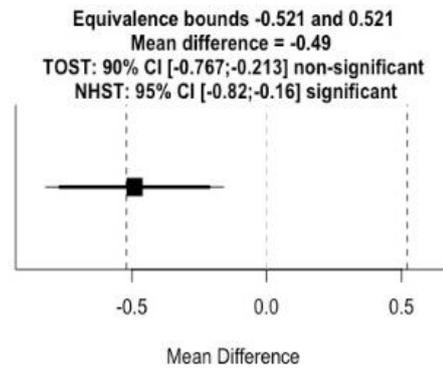
Em relação ao teste de equivalência entre os métodos se foi confrontado o resultado de cada ajuste com a altura observada.



C



D



E

F

Onde: Método dos Mínimos Quadrados(A); Regressão Quantílica(B); Modelo Misto com efeito aleatório nos dois parâmetros(C); Modelo Misto com efeito aleatório no parâmetro b_0 (D); Modelo Misto com efeito aleatório no parâmetro b_1 (E); Redes Neurais Artificiais(F).

Todos as abordagens apresentaram médias para as alturas estimadas não equivalentes à altura média observada, exceto a regressão quantílica, que mesmo no Experimento 3, com *outliers* inseridos de grande magnitude, manteve suas estatísticas de modelagem praticamente inalterada. Isso se deve ao fato de a abordagem trabalhar com as medianas, não com as médias, como nas outras abordagens por regressão.

A RNA não apresentou bom desempenho, isso pode ter ocorrido devido a base de dados apresentar variações que não foram captadas pela rede configurada.

5- Conclusão

Os resultados deste estudo permitem que as seguintes conclusões sejam formuladas:

A abordagem de ajuste por Regressão Quantílica foi a que gerou as menores tendenciosidades, nos menores e maiores valores de *outliers*, portanto essa abordagem se mostra precisa e menos sensível à presença de ruídos na base de dados.

Todos as abordagens possuem capacidade de generalização equivalentes sem a presença de *outliers*, o que mostra também precisão para estimativas de altura nesta condição. Na presença de erros em até dois desvios padrão dos valores observados de altura as abordagens de regressão centradas para a média (método dos mínimos quadrados e modelagem mista) também são precisas.

6- Referências

- ABPMA. Associação Brasileira dos Produtores de Mogno Africano.
- ARAÚJO, B. H. P. de. Modelagem da altura, volume e afilamento de *Calycophyllum spruceanum* Benth. empregando regressão e redes neurais artificiais. Manaus-AM: INoPA, 2015, 53 p. Dissertação (Mestrado em Ciências de Florestas Tropicais) – Instituto Nacional de Pesquisas da Amazônia.
- BARTOSZECK, A. C. P. S.; MACHADO, S. A.; FIGUEIREDO FILHO, A.; OLIVEIRA, E. B. Dinâmica da relação hipsométrica em função da idade, do sítio e da densidade inicial de povoamentos de Bracatinga da Região Metropolitana de Curitiba, PR. *Revista Árvore*, v.28, n.4, p.517-533, 2004.
- BAYAT, M.; GHORBANPOUR, M.; ZARE, R.; JAAFARI, A.; PHAM, B.T. Application of artificial neural networks for predicting tree survival and mortality in the Hyrcanian forest of Iran. *Computers and Electronics in Agriculture*, v.164, 2019.
- BINOTI, D. H. B.; BINOTI, M. L. M. da S.; LEITE, H. G. Configuração de redes neurais artificiais para estimação do volume de árvores. *Ciência da Madeira*, v. 5, n. 1, p. 58-67, 2014.
- BINOTI, M. L. M. Emprego de redes neurais artificiais em mensuração e manejo florestal. Viçosa-MG: UFV, 2012, 130 p. Tese (Doutorado em Ciências Florestais) – Universidade Federal de Viçosa.
- BINOTI, M. L. M. Redes neurais artificiais para prognose da produção de povoamentos não desbastados de eucalipto. Viçosa-MG: UFV, 2010, 54 p. Dissertação (Mestrado em Ciências Florestais) – Universidade Federal de Viçosa.
- BINOTI, M.L.M.S. et al. Aplicação de redes neurais artificiais para estimação da altura de povoamentos equiâneos de eucalipto. *Revista árvore*, v.37, n.4, p.639-645, 2013.
- BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDERMIR, T. B. *Redes neurais artificiais: teoria e aplicações*. 1. ed. LTC, Rio de Janeiro, 2000, 250 p.
- BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDERMIR, T. B. *Redes neurais artificiais: teoria e aplicações*. 2. ed.- [Reimpr.]. LTC, Rio de Janeiro, 2014, 26 p.
- BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDEMIR, T. B. *Redes Neurais Artificiais: teoria e aplicações*. 2 ed. Rio de Janeiro: LTC, 2007.
- BULLINARIA, J.A. Introduction to neural computation. 2014. Disponível em: <http://www.cs.bham.ac.uk/~jxb/inc.html>. acesso em 10 de fevereiro de 2022.
- CALEGARIO, N.; MAESTRI, R.; LEAL, C. L.; DANIELS, R. F. Estimativa do crescimento de povoamentos de *Eucalyptus* baseada na teoria dos modelos não lineares em multinível de efeito misto. *Ciência Florestal*, v. 15, n. 3, p. 285-292, 2005.

- CAMPOS, J. C. C.; LEITE, H. G. Mensuração florestal: perguntas e respostas. 5.ed. Viçosa: UFV, 2017. 636p.
- CHARNET, R, FREIRE, C. A.; CHARNET E. M. R.; BONVINO H. Análise de modelos de regressão linear com aplicações. 2. ed. Campinas: Unicamp, 2008, 356p.
- CORDEIRO, G. M.; DEMETRIO, C. G. B. Modelos lineares generalizados e extensões. 2011. (à publicar).
- COULSON, R.N.; FOLSE, L.J.; LOH, D. K. Artificial intelligence and natural resource management. *Science*, v. 237, n. 4812, p. 262-267, 1987.
- CURTIS, R. O. Height diameter and height diameter age equations for second growth Douglas-fir. *Forest Science*, v.13, n.4, p.365-375, 1967.
- DRAPER, N. M.; SMITH, H. *Applied Regression Analysis*. 2.ed., New York: Wiley. 1981, 709p.
- FALESI, Italo Claudio; BAENA, A. R. C. Mogno-africano *Khaya ivorensis* A. Chev. em sistema silvipastoril com leguminosa e revestimento natural do solo. Embrapa Amazônia Oriental-Documentos (INFOTECA-E), 1999.
- FARIAS, Aline Araújo. *Uso de regressão quantílica na predição da produção de povoamentos de eucalipto*. 2018.
- FERREIRA, J.C.B et al. Altura de mudas da *Tibouchina granulosa* COGN. (Melastomaceae) estimada por redes neurais artificiais. *Revista da Sociedade Brasileira de Arborização Urbana*, v.9, n.1, p.151-160, 2014.
- GORGENS, E. B.; LEITE, H. G.; SANTOS, H. do N.; GLERIANI, J. M. Estimação do volume de árvores utilizando redes neurais artificiais. *Revista Árvore*, v. 33, n. 6, p.1141-1147, 2009
- HAYKIN, S. *Neural networks and learning machines*. 3. ed. Prentice Hall, New York, 2009, 906 p.
- HAYKIN, S. *Redes neurais: princípios e prática*. 2. ed. Bookman, Porto Alegre, 2001, 900 p.
- KOENKER, R.; BASSETT, G. Regression quantiles. *Econometrica*, v. 46, n. 1, p. 33-50. 1978.
- LEITE, H. G.; OLIVEIRA, F. H. T. Statistical procedure to test the identity of analytical methods. *Communications in Soil Science and Plant Analysis*, v. 33, p.1105- 1118, 2002
- LEITE, H.G.; ANDRADE, V. C. L. Uso do método da altura relativa em inventário florestal de um povoamento de *Pinus*. *Revista Árvore*, v.28, n.6, p.865-873, 2004.

LIMA, R. B. de. Distribuição diamétrica e volumetria em vegetação de caatinga arbustiva-arbórea, Floresta-PE. Recife-PE: UFRPE, 2014, 134 p. Dissertação (Mestrado em Ciências Florestais) – Universidade Federal Rural de Pernambuco.

MACHADO, S. A.; BASSO, S. F.; BEVILACQUA JR, V. G. Teste de modelos matemáticos para o ajuste da relação hipsométrica em diferentes sítios e idades para plantações de *Pinus elliottii* no Estado do Paraná. In: CONGRESSO FLORESTAL PANAMERICANO, 1.; CONGRESSO FLORESTAL BRASILEIRO, 7., Curitiba, 1993. Anais..., São Paulo: Sociedade Brasileira de Silvicultura, 1993. v.2. p.553-556.

MACHADO, S. A.; FIGUEIREDO FILHO, A. Dendrometria. 2 ed. Guarapuava: UNICENTRO, 2014.316p.

MACHADO, W, C.; FONSECA JÚNIOR, E. S. Redes Neurais Artificiais aplicadas na previsão do VTEC no Brasil. Boletim de Ciências Geodesicas, v.19, n.2, p. 227-246, 2013.

MAGGI, et al. Plano Nacional de Desenvolvimento de Florestas Plantadas. MAPA, Brasília, 2018.

McCULLAGH, P.; NELDER, J. A. Generalized linear models. 2 ed. London: Chapman e Hall. 1989. 511p

MEHTÄTALO, L; MALTAMO, M; KANGAS, A. The Use of Quantile Trees in the Prediction of the Diameter Distribution of a Stand. *Silva Fennica*, v. 40, n. 3, 2006.

NIED, A. Treinamento de redes neurais artificiais baseado em sistemas de estrutura variável com taxa de aprendizado adaptativa. 2007. Belo Horizonte, MG. Tese (Doutorado) – Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Minas Gerais.

OLIVEIRA, X.M. Equações volumétricas e de afilamento para mogno africano (*Khaya ivorensis* A.Chev.) nos estados de Minas Gerais e Pará, Brasil. Dissertação (Pós Graduação em Manejo Florestal) - Universidade Federal de Lavras. Lavras, MG, 2017.

PINHEIRO et al. Mogno africano - Boa alternativa para produção de madeira nobre. *Revista da Madeira* – edição N° 131, 2012. Disponível em <http://www.remade.com.br/br/revistadamadeira_materia.php?num=1605&subject=E%20mais&title=Mogno%20africano%20-%20Boa%20alternativa%20para%20produ%20de%20madeira%20nobre> Acesso em 02 de fevereiro de 2022.

PINHEIRO, Antônio Lelis et al. Ecologia, silvicultura e tecnologia de utilização dos mognos-africanos (*Khaya* spp.). Viçosa, MG: Sociedade Brasileira de Agrossilvicultura, 2011.

PLÁCIDO, A. C. et al. Dinâmica da relação hipsométrica em função da idade, do sítio e da densidade inicial de povoamentos de Bracatinga da Região Metropolitana de Curitiba, PR. *Revista Árvore*, v.28, n.4, p.517-533, 2004.

- POLTRONIERI, LUIZ S. et al. Detecção de Phamerochaete salmonicolor em mogno africano no estado do Pará. *Fitopatologia Brasileira*, v. 27, p. 321-321, 2002.
- RIBEIRO, Andressa; FERRAZ, Antonio Carlos; SCOLFORO, José Roberto Soares. O cultivo do mogno africano (*Khaya spp.*) e o crescimento da atividade no Brasil. *Floresta e Ambiente*, v. 24, 2017.
- RIEDMILLER, M.; BRAUN, H. A direct adaptive method for faster back propagation learning: The RPROP algorithm. *IEEE International Conference*, p.586-591, 1993.
- ROCHA, José Sales Mariano da. *Manual de projetos ambientais*. 1997.
- SANQUETTA, C. R.; WATZLAWICK, L. F.; DALLA CORTE, A. P.; FERNAND, L. V. *Inventários florestais: planejamento e execução*. 3 ed. Curitiba: Multi-Graphic Gráfica e Editora, 2014. 406p.
- SCOLFORO, J. R. S.; THIERSCHI, A. Estimativas e testes da distribuição de frequência diamétrica para *Eucalyptus camaldulensis*, através da distribuição Sb, por diferentes métodos de ajuste. *Scientia Forestalis*, Piracicaba, n. 54, p. 93- 106, dez. 1998.
- SCOLFORO, J.R.S. *Modelagem do crescimento e da produção de florestas plantadas e nativas*. Lavras: UFLA/FAEPE, 1998. 1p.
- SEDIYAMA, Gilberto Chohaku et al. Zoneamento agroclimático do cafeeiro (*Coffea arabica L.*) para o Estado de Minas Gerais. *Revista Brasileira de Agrometeorologia*, v. 9, n. 3, p. 501-509, 2001.
- SILVA et al. Qualidade da madeira de mogno africano para a produção de serrados. *Sci. For.*, Piracicaba, v. 44, n. 109, p. 181-190, mar. 2016.
- SILVA, E. N. Efeito número de filhos na distribuição condicional da renda familiar: uma aplicação de variáveis instrumentais para estimar o efeito quantílico de um tratamento. *Dissertação (Mestrado em Economia) - Universidade Federal de Pernambuco*. Recife, 2003.
- SILVA, G. F.; CURTO, R. A.; SOARES, C. P.B.; PIASSI, L. C.; Avaliação de métodos de medição de altura em florestas naturais. *Revista Árvore*, v.36, n.2, p.341- 348, 2012.
- SILVA, I. N. da; SPATTI, D.; FLAUZINO, R. *Redes neurais artificiais para engenharia e ciências aplicadas: curso prático*. Artliber, São Paulo, 2010, 399p.
- SOARES, C. P. B.; PAULA NETO, F.; SOUZA, A. L. *Dendrometria e inventário florestal*. Viçosa: Editora UFV, 2011. 272p.
- SOARES, T. S. et al. Uso de diferentes alternativas para viabilizar a relação hipsométrica no povoamento florestal. *Revista Árvore*, v.28, n.6, p.845-854, 2004.
- SOUSA, Glauce Taís Oliveira et al. Relações hipsométricas para *Eucalyptus urophylla* conduzidos sob regime de alto fuste e talhadia no Sudoeste da Bahia. *Scientia Plena*, v. 9, n. 4, 2013.

VENDRUSCOLO, D.G.S. et al. Estimativa da altura de árvores de *Tectona grandis* L.f. utilizando regressão e redes neurais artificiais. *Nativa*, v.5, n.1, p.52-58, 2017.

VENDRUSCOLO, D.G.S. et al. Estimativa da altura de eucalipto por meio de regressão não linear e redes neurais artificiais. *Revista Brasileira de Biometria*, v.33, n.4, p.556-569, 2015.