

SYNTIA REGINA RODRIGUES DE SOUZA

**USO DE REDES NEURAIS NA ESTIMATIVA VOLUMÉTRICA DE
CLONES DE *EUCALYPTUS spp* NO PÓLO GESSEIRO DO ARARIPE,
PERNAMBUCO**

RECIFE-PE – JULHO/2015.



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

**USO DE REDES NEURAIS NA ESTIMATIVA VOLUMÉTRICA DE
CLONES DE *EUCALYPTUS spp* NO PÓLO GESSEIRO DO ARARIPE,
PERNAMBUCO**

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

Área de Concentração: Biometria e Estatística Aplicada

Orientador: José Antônio Aleixo da Silva

Co-orientador: Tiago Alessandro Espínola Ferreira

RECIFE-PE – JULHO/2015.

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

Dissertação julgada adequada para obtenção do título de Mestre em Biometria e Estatística Aplicada, defendida e aprovada em 31/07/2015 pela Banca Examinadora.

Orientador:

Prof. José Antônio Aleixo da Silva
Departamento de Ciências Florestais
Universidade Federal Rural de Pernambuco

Banca Examinadora:

Prof. Tiago Alessandro Espínola Ferreira
Pós-Graduação em Biometria e Estatística Aplicada
Universidade Federal Rural de Pernambuco

Prof Mêuser Jorge Silva Valença
Pós-Graduação em Ciências da Computação
Universidade de Pernambuco

Dedico este trabalho a minha família e amigos, em especial aos meus pais e minha avó por todo incentivo e confiança.

Agradecimentos

A Deus, por estar sempre presente em minha vida, por iluminar meu caminho com pessoas maravilhosas e muita paz. Pela saúde, força e discernimento em todos os momentos.

Aos meus pais, Marcos e Ana e irmãos, Paloma e Marcos Vinícius, pela paciência, incentivo e amor. Meus familiares por todas as palavras de incentivo e total paciência com minha ausência frequente. A minha avó Adelice, por me transmitir paz e muita vontade de evoluir.

Ao professor José Antônio Aleixo da Silva, pela orientação, confiança, paciência, incentivo e dedicação na elaboração deste trabalho. Foi para mim um exemplo de profissional e tem me inspirado muito.

Ao meu noivo, Sabi Yari Moïse Bandiri, por seu amor, carinho, incentivo, apoio e compreensão sempre.

A meu amigo doutorando Guera O.G. Mariel, agradeço pelo apoio, paciência, ajuda e total incentivo. A minhas amigas: Jainara, Jaciara, Geovana, Kelly, Simara, Rosangela, Aleonia, Maíra, Vanessa, por todo carinho, incentivo e mesmo com a distância fazerem parte de minha vida.

Aos meus colegas estatísticos: Hemílio, Hélio, Carlos, Juliana, Amanda. Aos meus colegas de programa em Biometria: Eneide, Ewerton, André, Isaac, Karla, Hérica e Luisa. Aos meus colegas de programa em Ciências Florestais: Joseane, Cintia, Tibério, Robson e Diego, pessoas que me apoiaram com suas amizades, conselhos e tirando minhas dúvidas. Foi uma ótima convivência e espero ter a oportunidade de estar sempre junto a vocês.

Em especial a Margarida, Dalka e Fábio, pessoas que me deram total incentivo nesse sonho. Obrigada por toda ajuda.

Ao Secretário Marco Santos pela ajuda, apoio e disponibilidade sempre. Ao Programa de Pós-graduação em Biometria e Estatística Aplicada e a todos os professores e coordenadores que me apoiaram. E, por fim, a CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, pelo apoio financeiro. E a todos que contribuíram para a conclusão deste trabalho.

“Talvez seja esse o segredo da vida: nada é impossível. Nós é que estabelecemos limites em nossas mentes, bloqueando nossa capacidade criativa. Nós deixamos de realizar obras que superam a normalidade, porque temos pouca fé em nosso potencial.”

Roger Bottini Paranhos.

Resumo

O Pólo Gesseiro do Araripe em Pernambuco é responsável por 97% da produção nacional de gesso e a principal fonte de energia para o processo de calcinação da gipsita, matéria prima para produção de gesso, é a lenha proveniente da Caatinga, vegetação natural da região. Devido aos altos custos de outras fontes de energia elevar a produção de gesso implica em aumentar o desmatamento da Caatinga. Uma solução econômica e ambiental para esse problema é a implantação e o manejo sustentado de povoamento de espécies nativas ou o reflorestamento com espécies florestais de rápido crescimento. Dentre as florestas de rápido crescimento, o gênero *Eucalyptus* se destaca por sua alta produção e adaptabilidade ao semiárido nordestino. O objetivo deste trabalho é estimar o volume de clones de *Eucalyptus spp* no Pólo Gesseiro do Araripe empregando a metodologia de Redes Neurais Artificiais (RNAs) comparando-a com os modelos volumétricos de Schumacher e Hall e Spurr. Os dados são referentes a um experimento implantado na Estação Experimental do Instituto Agrônomo de Pernambuco, onde foram testados 15 clones de *Eucalyptus spp* plantados em 2002 e com o corte final em 2009. A função de interesse estimada foi o volume da árvore (V) em relação do diâmetro a altura do peito (DAP), altura total d árvore (Ht) e tipo de clone. Também foi avaliado o ajuste dos melhores modelos por tamanho de amostra. Os resultados foram avaliados com o coeficiente de determinação ajustado (R^2_{aj}), raiz quadrada do erro médio percentual (RMSE%), o erro padrão da estimativa (Syx%) e análise dos gráficos de dispersão do resíduo. Os resultados obtidos no trabalho confirmaram a expectativa mostrando a eficiência dos ajustes independente do tamanho da amostra.

Palavras-chave: Manejo Florestal, Modelos Volumétricos, Redes Neurais Artificiais.

Abstract

The Araripe Gypsum Pole in Pernambuco is responsible for 97% of national production of plaster. The main source of energy for the gypsum calcination process, raw material for plaster production is the wood from the natural vegetation of Caatinga. Due to the high costs of other energy sources, increasing the gypsum production implies more deforestation of the Caatinga. An economic and environmental solution for that problem is the implementation and the sustainable management of native species or the reforestation with fast growing forest species. Among the fast growing forest species the genus *Eucalyptus* stands out for its productivity and adaptation to the Northeast semi-arid region. The objective of this study was to estimate the volume of the *Eucalyptus spp* clones in Gypsum Araripe Pole employing the methodology of Artificial Neural Networks (ANN) comparing it with the volumetric models of Schumacher and Hall and Spurr. Data came from an experiment implanted in the Experimental Station of the Agronomic Institute of Pernambuco, where 15 clones of *Eucalyptus spp* were planted in 2002, with final cut in 2009. The function of interest estimated was the volume of the tree in function of the diameter at the breast height (DBH), total height (Ht) and the clone type. It was also valued the adjustment of the best models for sample size. The results were evaluated with the adjusted coefficient of determination (R^2_{aj}), square root of the percentual mean error (RMSE%), standard error estimate (Syx%) and an analysis graphic of the residues. The obtained results confirmed the expectation showing efficiency of adjustments independent of the sample size.

Keywords: Forest Management, Volumetric Models, Artificial Neural Networks.

Conteúdo

Conteúdo	11
Lista de Figuras	12
Lista de Tabelas	13
Lista de Abreviaturas e Siglas	14
Capítulo 1: Introdução	15
1.1 Objetivo.....	17
Capítulo 2: Revisão da Literatura	18
2.1. Caracterização do Pólo Gesseiro do Araripe.....	18
2.2. Florestas energéticas: as plantações de <i>Eucalyptus</i>	19
2.3. Variáveis dendrométricas	20
2.4 Estimativas volumétricas	22
2.5. Redes neurais artificiais (RNA).....	27
Capítulo 3: Material e Método	37
3.1. Área experimental.....	37
3.2. Modelos	39
3.3. Validação e comparação dos modelos	41
3.4. Uso de diferentes tamanhos de amostras	44
Capítulo 4: Resultados	45
4.1. Resultado geral.....	45
4.2. Resultado por diferentes tamanhos de amostras	49
Capítulo 5: CONCLUSÃO	57
REFERÊNCIAS BIBLIOGRÁFICAS	58

Lista de Figuras

Figura 2.1: Tora ilustrada com as dimensões para o calculo do volume.....	21
Figura 2.2: Modelo de um neurônio artificial	28
Figura 2.3: Funções de ativação: (a) Limiar, (b) Logística (Sigmóide), (c) Linear, (d) Tangente hiperbólica.....	29
Figura 2.4: Estrutura básica de uma MLP	32
Figura 3.1: Árvore ilustrada com as suas dimensões.....	40
Figura 4.1: Gráficos de dispersão dos resíduos e histograma de freqüência de resíduos evidenciando o comportamento dos desvios encontrados entre volume observado e volume estimado.....	48
Figura 4.2: Gráficos de dispersão dos resíduos evidenciando o comportamento dos desvios encontrados entre volume observado e volume estimado por amostra.	53
Figura 4.3: Gráficos de Coeficiente de determinação (R^2) por amostra dos modelos analisados.	54
Figura 4.4: Gráficos do Erro padrão da estimativa ($S_{yx}\%$) por amostra dos modelos analisados.	55
Figura 4.5: Gráficos da Raiz quadrada do erro médio (RMSE%) por amostra dos modelos analisados	55

Lista de Tabelas

Tabela 3.1: Descrição dos tratamentos usados no experimento.	38
Tabela 3.2: Frequência dos tratamentos de cada clone de <i>Eucalyptus spp</i> do experimento.....	39
Tabela 3.3: Definição das amostras e erros amostrais da base de <i>Eucalyptus spp.</i> do experimento.....	44
Tabela 4.1: Estatísticas descritivas das variáveis contínuas dos <i>Eucalyptus spp.</i> do experimento.....	45
Tabela 4.2: Estatísticas do ajuste dos modelos volumétricos de <i>Eucalyptus spp.</i>	46
Tabela 4.3: Estatísticas de ajuste para o volume de <i>Eucalyptus spp</i>	46
Tabela 4.4: Definição das RNAs.....	47
Tabela 4.5: Estatísticas descritivas por amostra das variáveis contínuas dos <i>Eucalyptus spp.</i> utilizadas no estudo	50
Tabela 4.6: Estatísticas das RNAs para o volume de <i>Eucalyptus spp.</i> considerando como entrada a variável 'Clone'	51
Tabela 4.7: Definição das RNAs por amostra das RNAs com e sem variável categórica 'Clone'	52

Lista de Abreviaturas e Siglas

- *APL*: Arranjo produtivo local
- *RNAs*: Redes Neurais Artificiais
- *MCP*: Modelo de McCulloch e Pitts
- *MLP*: Modelos Multilayer Perceptron
- *CAP*: Circunferência à altura do peito
- *DAP*: Diâmetro à altura do peito
- *Ht*: Altura total

Capítulo 1: Introdução

O local de maior produção de gesso no Brasil, com destaque mundial por seu depósito de reservas e qualidade de sua gipsita está situado no Nordeste, no estado de Pernambuco. Uma parte dessas reservas que obtém condições mais favoráveis de aproveitamento econômico se encontra na região do Araripe, que se localiza na fronteira dos estados do Piauí, Ceará e Pernambuco (SOBRINHO *et. al.*, 2003) no arranjo produtivo local (APL) denominado de o Pólo Gesseiro do Araripe que é responsável por 97% da produção nacional de gesso. A principal fonte de energia para o processo de calcinação da gipsita, matéria prima para o gesso, que é submetida a altas temperaturas, é a lenha proveniente da Caatinga, vegetação predominante na região. (SILVA, 2008/2009)

Devido aos altos custos de outras fontes de energia, elevar a produção de gesso implica em aumentar o desmatamento da Caatinga. Uma solução econômica e ambiental é a implantação e o manejo sustentado de povoamento de espécies nativas ou o reflorestamento com espécies florestais de rápido crescimento.

Segundo Riegelhapt e Pareyn (2010), 80% da lenha consumida no Polo Gesseiro do Araripe tem origem ilegal, desta forma, estima-se que 11.229 ha são explorados, anualmente, sem consentimento dos órgãos ambientais. Segundo Campello (2011), o diagnóstico realizado pelas empresas calcificadoras de gipsita do Araripe revelou que os reflorestamentos energéticos representam apenas 0,1% da matriz energética sendo a maior parte das outras fontes ilegais e pouco viáveis do ponto de vista econômico.

Diante disto, as novas fontes energéticas renováveis foram consideradas como opções viáveis e umas das ações realizadas para concretizar essas opções foi o estabelecimento de experimentos com clones de *Eucalyptus spp* em 2002.

Inúmeras espécies de rápido crescimento são utilizadas nos planos de reflorestamento energético, porém as do gênero *Eucalyptus* se destacam devido à diversidade, o rápido crescimento, o diversificado aproveitamento de sua madeira e fácil adaptação a diversas condições ambientais.

É importante, destacar a necessidade de um bom planejamento na implantação das florestas energéticas, de modo a garantir os benefícios ambientais (diminuição da pressão exercida nas florestas nativas), econômicos (maior retorno

financeiro) e sociais (elevação do número de empregos diretos e indiretos na região de implantação). A estimativa mais precisa da produtividade das florestas energéticas de *Eucalyptus spp* contribuirá, sem dúvida, para melhorar o planejamento do abastecimento de energia na região.

Existem inúmeros métodos utilizados para estimar o volume de madeira em florestas. Os métodos tradicionais usam principalmente os modelos volumétricos ajustados por meio de análise de regressão linear e não linear. Nos últimos anos, resultados satisfatórios têm sido obtidos ao empregar como ferramenta Redes Neurais Artificiais (RNA) (BINOTI, 2012).

As RNAs são modelos matemáticos que solucionam problemas simulando o funcionamento do cérebro humano. A utilização de algoritmo de aprendizado para solução de um problema é uma das propriedades mais importantes nas RNA.

Uma característica importante nos modelos de RNA é não precisar de suposição sobre as relações das variáveis utilizadas no ajuste do modelo.

O manejo florestal passa por transformações com a evolução do tempo e com isto há a necessidade do aprimoramento de ferramentas, com resultados viáveis e mais precisos. Assim, o uso da inteligência artificial por meio de RNA na pesquisa florestal, apresenta-se como uma nova ferramenta, apresentando resultados favoráveis e sendo útil na tomada de decisões (LIMA, 2014).

Dentro dos principais trabalhos desenvolvidos, podem se citar: Uso de RNA na avaliação funcional de estradas florestais.(OLIVEIRA 2008); Uso de redes neurais artificiais como uma alternativa para mapeamento de biomassa e carbono orgânico no componente arbóreo de florestas naturais (Schoeninger et al. 2008); Ajuste do modelo de Schumacher e Hall e aplicações de redes neurais artificiais para estimar volume de árvores de eucalipto (Silva et. al. 2009); Estimativa do volume de árvores utilizando RNA (Gorgens et al. 2009); Discriminação de populações com diferentes graus de similaridade por RNA (Pereira 2009); Configuração de redes neurais artificiais para estimativa do volume de árvore (Binoti et al. 2014)

1.1 Objetivos

- a) Estimar o volume de clones de *Eucalyptus* spp no Pólo Gesseiro do Araripe empregando a metodologia de Redes Neurais Artificiais (RNAs) Comparando o resultado com os modelos volumétricos de Schumacher e Hall e Spurr.
- b) Verificar a eficiência das RNA em diferentes tamanhos de amostra.

Capítulo 2: Revisão da Literatura

2.1. Caracterização do Pólo Gesseiro do Araripe

O Pólo Gesseiro do Araripe está localizado na região do Araripe e corresponde a 18% da área do estado de Pernambuco, fronteira com os estados do Ceará e Piauí. Compreende os municípios de Araripina, Bodocó, Exu, Ipubi e Trindade, que compõem a microrregião de Araripina.

A gipsita matéria prima na indústria do gesso, é explorada nessa microrregião sendo considerada a de melhor qualidade no mundo, com a pureza do minério variando entre 88% e 98% (SOBRINHO et al., 2003; BALTAR et al., 2003; ARAÚJO, 2004). A produção de gesso na região representa 97% da atual produção nacional.

A lenha proveniente da Caatinga é a principal fonte de energia para o processo de calcinação da gipsita e na maioria das vezes, é explorada de forma ilegal. Segundo Atecel (2006), a matriz energética do Polo Gesseiro do Araripe é muito diversificada e usa aproximadamente, 3% de energia elétrica, 5% de óleo diesel, 8% de óleo BPF (baixo poder de fusão), 10% de coque e 73% de lenha. Nas pequenas empresas gesseiras, a dependência da lenha é quase que total. Essas estatísticas relatam a alta dependência do Polo Gesseiro da energia proveniente das florestas. Se a isto se soma o vertiginoso aumento da produção de gesso nos últimos anos (5,6 milhões de toneladas em 2014), entende-se a aceleração da degradação da vegetação nativa (Caatinga). Por outro lado, o plantio de florestas energéticas pelo estabelecimento de espécies de rápido crescimento é uma forma de contribuir a atender a maior demanda energética (lenha) na região.

Há muitos anos a produção de gipsita no Araripe vem se destacando no Brasil. Os fatores que contribuíram para essa evolução são de ordem natural, como a facilidade de transporte e a facilidade de extração a céu aberto. Essa atividade econômica gera para a região 13,9 mil empregos diretos e 69 mil indiretos, resultantes da atuação de 174 indústrias de calcinação e 42 minas de gipsita que geram um faturamento anual na ordem de R\$ 1,4 bilhões/ano (SINDUSGESSO, 2014).

2.2. Florestas energéticas: as plantações de *Eucalyptus*

As florestas energéticas são estabelecidas com o objetivo de fornecer biomassa florestal, lenha e carvão de origem vegetal além de ter funções secundárias como evitar a pressão do desmatamento sobre as florestas naturais (SEBRAE, 2015).

As florestas energéticas atendem dez por cento (10%) da demanda em energia primária do Brasil (BRASIL, 2013). São numerosas as espécies direcionadas à produção de energia, substituindo a madeira oriunda de florestas nativas; dentro delas pode citar espécies como leucena, angico, bracatingas e espécies do gênero *Eucalyptus*.

As espécies do gênero *Eucalyptus* se destacam pelas suas características (elevada taxa de crescimento, facilidade de reprodução, qualidade da madeira e adaptação ao clima e solo da região) e são as mais utilizadas no estabelecimento das florestas energéticas no Brasil. O gênero *Eucalyptus* é originário da Austrália e Indonésia, chegou ao Brasil em 1825 e desde 1903, é utilizado em atividades econômicas. Oferece uma elevada variação latitudinal, possui mais de 700 espécies com genótipos adaptados às variadas condições de clima e solo (LIMA, 1996; SILVA, 2008/2009).

O manejo florestal de espécies ou clones desse gênero na região do Polo Gesso do Araripe é uma solução viável para diminuir a pressão sobre a Caatinga, vegetação nativa dessa região, possibilitando a produção de madeira para atender às necessidades da indústria do gesso com bases sustentáveis.

Habitualmente, equações de volume são usadas para prever o conteúdo volumétrico de árvores utilizando como variáveis preditoras o diâmetro à altura do peito (DAP) e altura total da árvore (Ht). O verdadeiro volume de uma árvore é calculado geralmente na prática com a árvore sendo abatida e são medidos os diâmetros e comprimentos relativos de pequenas secções que tem seus volumes calculados, utilizando geralmente a fórmula de Smalian, que envolve a média da área em corte transversal das extremidades grandes e pequenas, vezes o comprimento da sessão para adquirir o volume cúbico. O volume das secções é somado para adquirir o volume da árvore de interesse (BURKHART et. al., 2012).

2.3. Variáveis dendrométricas

O diâmetro e circunferência à altura do peito (DAP e CAP) são medidas fundamentais para estudos na Engenharia Florestal. Essas variáveis servem como referência para medições ou estimações da área basal e volume entre outros itens. A medida mais comum do diâmetro de uma árvore é o DAP o qual é medido a 1,30m de altura do chão, ponto de medida estabelecido internacionalmente. Quando a árvore na altura do DAP não é circular normalmente adotam-se duas medidas perpendiculares do diâmetro na mesma altura, considerando a parte mais larga e a mais estreita para obter a média estimando o valor do DAP. (ENCINAS et. al., 2002).

Qualquer corte transversal na superfície do fuste resulta na área transversal da árvore, se essa área for calculada em função do DAP é denominada a área basimétrica da árvore (g). A fórmula para o cálculo dessa variável é:

$$g = DAP^2 \cdot \frac{\pi}{4} \quad (2.1)$$

A altura total de uma árvore (Ht) é uma variável fundamental, junto com o DAP, para estimação do volume da madeira. A altura total (Ht) da árvore é a distância vertical desde o chão até o ápice da copa. Há vários métodos de medição para a altura, o de estimação é o que utiliza de instrumentos (hipsômetros), dependendo das condições de trabalho os erros nas estimativas variam entre -5% e +5% do valor real. (ENCINAS et. al., 2002). Recentemente esses aparelhos tem avançado contendo dispositivos eletrônicos o que estão agilizando as leituras (CAMPOS; LEITE, 2009)

É difícil identificar a forma do tronco de uma árvore. Árvores com altura e diâmetro iguais podem apresentar volumes diferentes se a forma dos troncos for diferente. Com o objetivo de simplificar o cálculo, foram estabelecidas na literatura fórmulas gerais com resultados satisfatórios no cálculo do volume. Há muitas opções para calcular o volume sólido da tora, uma opção bastante conhecida no meio florestal é o método de Smalian (Figura 2.1). A tora é dividida em seções e o volume total calculado pela soma dos volumes seccionados. A cubagem rigorosa dessa fórmula clássica deve ser feita com as árvores derrubadas (LEITE, 2008). Cada seção e o volume total são determinados pelas equações abaixo (2.2) e (2.3).

$$V_i = \left(\frac{g_1 + g_2}{2} \right) \cdot l \quad (2.2)$$

Em que:

V_i = Volume da seção

g_1 = área da maior base

g_2 = área da menor base

l = comprimento da seção

$$V_T = \sum_{i=1}^n V_i \quad (2.3)$$

V_T = Volume total

n = Número de seções

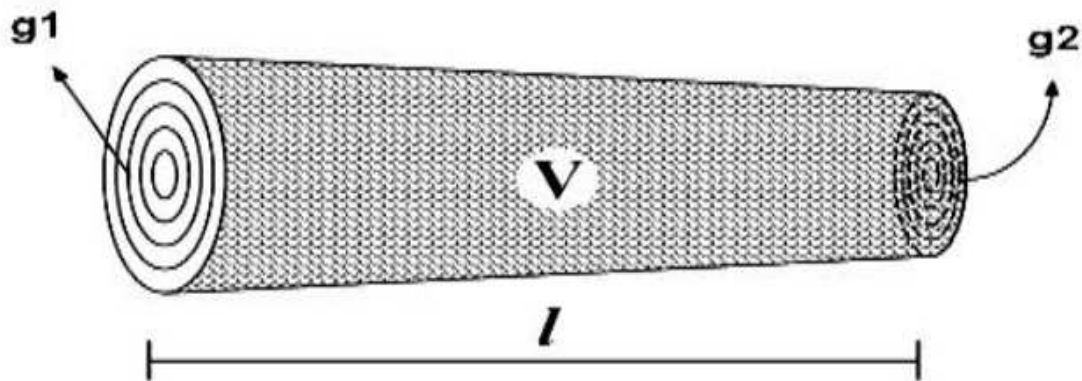


Figura 2.1: Tora ilustrada com as dimensões para o cálculo do volume. (Fonte: Lisboa et al., 2009).

Em mensuração florestal um dos principais objetivos é obter métodos que permita adquirir valores precisos do volume de madeira. A forma funcional para as equações de volume se apresenta como:

$$V = f(DAP, Ht) \quad (2.4)$$

Sendo, V o volume, da árvore.

2.4 Estimativas volumétricas

No manejo florestal, as predições do crescimento e produção têm tradicionalmente tomado a forma de tabelas comumente conhecidas como "tabela de volume". Essas tabelas expressam o volume de madeira da árvore esperado (geralmente em metros cúbicos) por combinações de características mensuráveis como DAP e Ht. Segundo Vuokila (1965), as origens de sistemas de simulação na área florestal encontram-se no final do século XVIII no desenvolvimento dessas tabelas de volume na Alemanha.

Segundo Rolim et al., (2006) na década de 80 era comum realizar o cálculo do volume com casca pela fórmula do volume cilíndrico, corrigido por um fator de forma igual a 0,7. Entretanto, quando se calcula o volume a partir de modelos volumétricos ajustados para o local da área de estudo, obtêm-se melhor precisão.

Na atualidade, a maioria dos sistemas de determinação da produção volumétrica é expressa como equações ou sistemas de equações matemáticas inter-relacionadas e não como tabelas de volume como antigamente. Um dos motivos dessa mudança é a propagação dos computadores, que permitiu a passagem de métodos gráficos que eram utilizados para a construção das tabelas de volume a métodos analíticos a partir de 1940 (ALDANA, 2010).

Embora a base seja a mesma e os mesmos modelos volumétricos tenham sido usados também na construção das tabelas de volume, é importante expressar que a mudança foi observada em dois planos: O primeiro foi na mudança do método de ajuste dos modelos, passando de métodos gráficos limitados a métodos analíticos de maior precisão, menor subjetividade e que possibilitam fazer prognoses para quaisquer combinações desejadas das variáveis de entradas. A segunda mudança foi na forma de apresentação dos resultados que passaram de tabelas baseadas em valores médios a equações genéricas que não podiam ser utilizadas pela complexidade de realização das estimativas (ALDANA, 2010).

A quantificação do volume de madeira em povoamentos florestais é imprescindível para a implementação de planos de manejo sustentáveis (LEITE; ANDRADE, 2002). As decisões de manejo florestal são baseadas em informações sobre condições de recursos atuais e futuros (AVERY; BURKHRT, 1995).

Os principais métodos de estimação de volume são o uso do fator de forma, equações de volume e funções de afilamento. (COCHRAN, 1965).

Das três ferramentas, as mais amplamente utilizadas nas estimativas e prognoses da produção madeireira são as equações volumétricas cujos parâmetros são determinados por regressão, constituindo um procedimento eficiente para a quantificação da produção em volume de um povoamento florestal (SANTANA; ENCINAS, 2004).

Ainda que hajam diferentes métodos de cálculos de volume as diferentes formas do fuste ou a grande quantidade do número de árvores para obter esse resultado pode dificultar a identificação precisa do volume. Uma alternativa é a estimação por meio de equações de regressão ou outras técnicas de previsão como as redes neurais artificiais (RNA).

A estimativa de equações volumétricas criadas especificamente para as áreas de manejo florestal sustentável constitui uma metodologia eficiente, econômica e com precisão aceitável para a quantificação da produção em volume da floresta. (SILVA, 2007).

A maioria das análises quantitativas, para estimação da variável volume está associada a modelos ajustados de regressão, podendo ser linear ou não linear. Obter informações confiáveis como o volume é importante para o planejamento florestal. Métodos de pesquisa surgem nesse cenário como importantes fontes de produção de conhecimento, principalmente para estimação do volume (SILVA et. al., 2009).

O fato dos modelos serem empíricos se faz necessário seu ajuste para obter equações adaptadas a diferentes espécies, idades, espaçamentos e regiões (PIMENTEL, 1993).

Segundo Machado et al. (2002), nessas equações volumétricas, o volume é a variável dependente, associado a variáveis independentes de fácil mensuração na floresta, como o DAP e a altura.

A regressão tem como objetivo estimar a variável dependente, por meio de uma hipótese a respeito de uma provável relação de causa e efeito entre variáveis explicativas, DAP e Ht. Uma forma geral de descrever um modelo de regressão linear simples é:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (2.5)$$

Em que Y_i é a variável de interesse (dependente), X_i a variável explicativa, β_0 e β_1 , são os parâmetros a serem estimados e ε é o erro aleatório.

No caso de um modelo possuir mais de uma variável explicativa, a expressão se define como modelo de regressão linear múltipla.

O modelo de regressão é baseado em algumas suposições. As principais suposições recaem sobre o componente erro do modelo, ε , cuja verificação é feita por meio dos resíduos, $\hat{\varepsilon}$.

i) $E(\varepsilon) = 0, \text{Var}(\varepsilon) = \sigma^2$

ii) Os erros são não correlacionados

iii) $\varepsilon \sim N(0, \sigma^2)$

Se i) e iii) ocorrem, então a variável dependente Y é uma variável aleatória com distribuição normal com variância σ^2 e média μ .

Os modelos lineares são frequentemente utilizados para aplicações em regressão, com o objetivo de relacionar uma variável resposta a uma ou mais variáveis explicativas. Entretanto, há uma classe de modelos que aplicando uma transformação adequada as variáveis, é possível escrever como função linear dos coeficientes permitindo a estimação dos parâmetros por meio do modelo de regressão linear. Esses modelos são conhecidos como linearizáveis (DRAPER; SMITH, 1998).

O principal objetivo dos modelos de regressão é a estimação dos parâmetros desconhecidos β_0 e β_1 . Esse procedimento é chamado de ajuste do modelo aos dados. O valor ajustado estimado de Y é dado pela equação seguinte:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \quad (2.6)$$

Resíduo é a diferença entre o Y observado e o \hat{Y} estimado.

$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i \quad (2.7)$$

$$\hat{\varepsilon}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i \quad (2.8)$$

As equações de volume estimam o resultado em função do DAP e a H_t , podendo ser expressa por uma dessas variáveis citadas ou a combinação delas. A função com duas entradas, em função do DAP e H_t , é a mais apresentada em estudos com esse objetivo.

Ao longo do tempo, o nível geral de sofisticação desses modelos tem crescido devido a muitos fatores, incluindo técnicas estatísticas mais refinadas de estimativa, bases de dados em expansão, melhor compreensão da dinâmica da floresta, e o poder crescente e disponibilidade de computadores. (SHAO; REYNOLDS, 2006).

Muitas técnicas foram incorporadas nos modelos de crescimento e produção desenvolvidos para a atividade de manejo florestal. Dentro dos principais trabalhos desenvolvidos, podem se citar: (1) Modelos ao nível de povoamento (LEAK 1964, MOSER; HALL 1969, MURPHY; FARRAR 1985); (2) Equações e sistemas de equações volumétricas (MOSER 1972, 1974, LEARY et al. 1979, McTAGUE; STANSFILD 1994); (3) Tabelas de volume não-lineares (ADAMS; EK 1974), (4) Cadeias de Markov (BRUNER; MOSER 1973, FRAZIER 1978); (5) modelos matriciais (BUONGIORNO; MICHIE 1980, SOLOMON; HOSMER 1986, FAVRICHON 1998), e (6) Técnicas de Redes Neurais Artificiais (GUAN; GERTNER 1995; KELLER et al. 1997).

Dois modelos muito utilizados para estimar os volumes das árvores na ciência florestal são o logarítmico de Schumacher e Hall e o da variável combinada de Spurr, destacando a precisão e ajustamento obtidos e a facilidade em utilizar (COUTO, et al., 1999).

A função freqüentemente mais usada para estimar o volume da árvore é a logarítmica, de Schumacher e-Hall (1933). Essa equação pode ser resolvida diretamente utilizando técnicas de regressão não-linear, mas, na prática, a forma funcional é geralmente convertida numa forma que é linear nos parâmetros por meio da técnica de transformação logarítmica, aplicando assim a regressão linear na sua solução (BURKHART et al., 2012).

Schumacher e-Hall (1933)

$$V_i = \beta_0 \cdot DAP_i^{\beta_1} \cdot Ht_i^{\beta_2} \cdot \varepsilon_i \quad (2.9)$$

A transformação logarítmica tem a vantagem de estabilizar a variância do volume (BURKHART et al, 2012).

$$\ln V_i = \ln \beta_0 + \beta_1 \cdot \ln DAP_i + \beta_2 \cdot \ln H_i t + \ln \varepsilon_i \quad (2.10)$$

Numerosos modelos de volume têm sido propostos, mas além do modelo Schumacher e-Hall uma das mais eficazes é a equação "variável combinada", que combina o diâmetro à altura do peito (DAP) e altura total (Ht) em um único preditor, o modelo conhecido por Spurr (1952) (BURKHART et al, 2012).

Spurr (1952)

$$V_i = \beta_0 + \beta_1 \cdot DAP_i^2 Ht_i + \varepsilon_i \quad (2.11)$$

Em que:

Ln = logaritmo neperiano

V_i = volume em m^3 da árvore i

DAP_i = diâmetro a altura do peito (1,3 m de altura), (cm) da árvore i

Ht_i = Altura total da árvore, (m) da árvore i

β_i = parâmetros do modelo

ε = erro aleatório, $\varepsilon_i \sim NID(0, \sigma^2)$

A estimativa volumétrica por meio de equações é a técnica mais utilizada por muitos anos no estudo de manejo florestal até os dias de hoje. Muitos autores contribuíram para a evolução científica da área estudando o comportamento desses modelos em diferentes espécies de plantas com destaque para os modelos de Shumarcher e-Hall e Spurr.

Silva (1977) comparou cinco tipos de modelos, três lineares e dois não lineares, concluindo que o modelo proposto por Shumachere e-Hall foi o mais apropriado para o estudo.

Guimarães (1986) concluiu que o modelo de Schumacher e-Hall dentre outras três alternativas demonstrou alta eficiência na estimativa do volume de árvores provenientes de doze povoamentos de *Eucalyptus grandis*.

Rolim et al. (2006) testaram treze modelos para estimar volume para a floresta nacional do Tapirapí-Aquiri na Serra dos Carajás e concluíram que o modelo Schumacher e-Hall apresentou melhor precisão.

Leite (2008) ao estimar o volume comercial de árvores em função do DAP e Ht, para uma área de floresta de terra firme comparou os modelos de Spurr e

Shumacher e-Hall identificando que o modelo de Spurr apresentou as melhores medidas de precisão e distribuição residual.

Lima et. al. (2014) ao estudarem a volumetria e classificação da capacidade produtiva para *Mora paraensis* (Ducke) no estuário amapaense, identificou que o modelo de dupla entrada de Schumacher e-Hall logarítmico foi mais adequado para a estimativa de volume do fuste com casca para a espécie.

Entretanto, uma nova metodologia tem se destacado na modelagem volumétrica, denominada de Redes Neurais Artificiais (RNA).

2.5. Redes neurais artificiais (RNA)

Redes Neurais Artificiais consistem em uma técnica para solucionar problemas de inteligência artificial, construindo um sistema que tenha circuitos que simulem o cérebro humano, aprendendo, errando e fazendo descobertas. São técnicas computacionais que apresentam um modelo inspirado na estrutura das redes biológicas e que adquirem conhecimentos por meio da experiência (BULLINARIA, 2014).

Segundo (BRAGA et al., 2012; HAYKIN, 2001) citados por Lima (2014), pode-se definir as RNAs como sistemas em paralelo distribuídos e compostos por unidades de processamento simples, neurônios ou nodos, e que computam determinadas funções matemáticas, lineares e não-lineares.

Uma rede neural artificial é uma técnica de modelagem semelhante ao cérebro, com capacidade de adquirir conhecimento experimental. A RNA deve seu bom desempenho a interligação de suas unidades de processamento (HAYKIN, 2001). As RNAs tentam reproduzir as redes biológicas, entretanto, existem diferenças entre essas redes. Analisar as semelhanças entre esses dois tipos de sistemas é importante para entender melhor o sistema nervoso e também para encontrar idéias para pesquisas na neurocomputação (VALENÇA, 2005).

A combinação de diversos neurônios artificiais forma uma rede neural artificial. As entradas, simulando uma área de obtenção de estímulos, podem ser conectadas em muitos neurônios, resultando, assim, em uma série de saídas, em que cada neurônio representa uma saída (TATIBANA; KAETSU, 2014).

Os neurônios são divididos em três partes: o corpo da célula, os dendritos e o axônio, cada um com funções específicas, porém complementares. Os dendritos têm por função receber as informações, impulsos nervosos, vindas de outros neurônios e conduzi-las até o corpo celular. As informações são processadas no corpo celular e novos impulsos são gerados. Os impulsos são transmitidos até outros neurônios por meio do axônio até os dendritos dos neurônios seguintes (BRAGA et al., 2012).

A sinapse é definida como o ponto de contato entre a terminação axônica de um neurônio e o dendrito de outro. É pelas sinapses que os nodos se unem funcionalmente formando redes neurais, as sinapses são capazes de controlar a transmissão de impulsos, isto é, o fluxo de informação entre os nodos na rede neural. (BRAGA et al., 2012).

Uma simplificação do neurônio biológico é o modelo de neurônio proposto por McCulloch e Pitts (MCP, 1943). Pode ser representado em um modelo com n terminais de entrada X_1, X_2, \dots, X_n que representam os dendritos e um terminal de saída Y que representa o axônio. A entrada do neurônio tem pesos de conexão W_1, W_2, \dots, W_n para simular o comportamento das sinapses, os valores desses pesos podem ser positivos ou negativos, dependendo das sinapses relacionadas serem inibitórias ou excitatórias e eles determinam em que grau o neurônio deve considerar sinais de disparo que ocorrem naquela conexão. O efeito de uma sinapse particular i é dado por $X_i W_i$. (BRAGA et al., 2012). A figura abaixo mostra um modelo de um neurônio artificial:

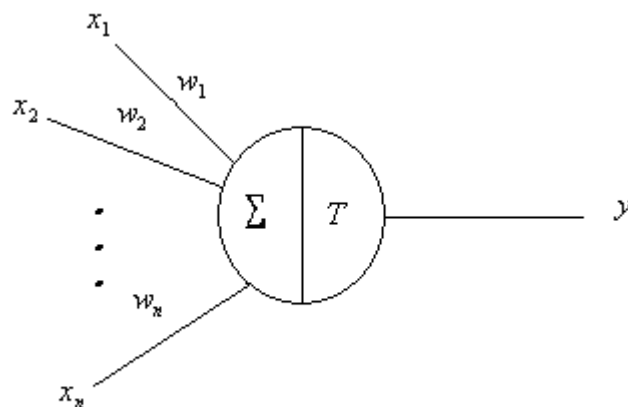


Figura 2.2: Modelo de um neurônio artificial, (Fonte: BRAGA et al., 2012).

Quando a soma dos impulsos que o neurônio biológico recebe ultrapassa o seu limiar de excitação ele é disparado. O corpo do neurônio é simulado por um mecanismo simples que faz a soma ponderada dos valores $X_i W_i$ recebidos pelo neurônio e com isso decide se o neurônio deve ou não disparar comparando a soma obtida ao limiar do neurônio. No modelo de McCulloch e Pitts (MCP), a ativação do neurônio é obtida por meio da aplicação de uma função de ativação, que ativa ou não a saída (BRAGA et al., 2012). Neste modelo, a função de ativação é dada pela função de limiar que terá então sua saída ativada quando:

$$\sum_{i=1}^n X_i W_i \geq T \quad (2.12)$$

em que n é o número de entradas do neurônio, W_i é o peso associado à entrada X_i e T é o limiar do neurônio.

A função de ativação é responsável por gerar a saída do neurônio a partir dos valores ponderados do peso e da entrada (UNICAMP, 2015). De acordo com BOCANEGRA (2002) citado por Lima (2014), essa função diferencia a decisão neural com base no potencial interno do neurônio e determina o valor a ser enviado aos demais neurônios.

. Abaixo estão dispostas algumas funções de ativação empregados na literatura para as arquiteturas do tipo MLP (Multilayer Perceptron).

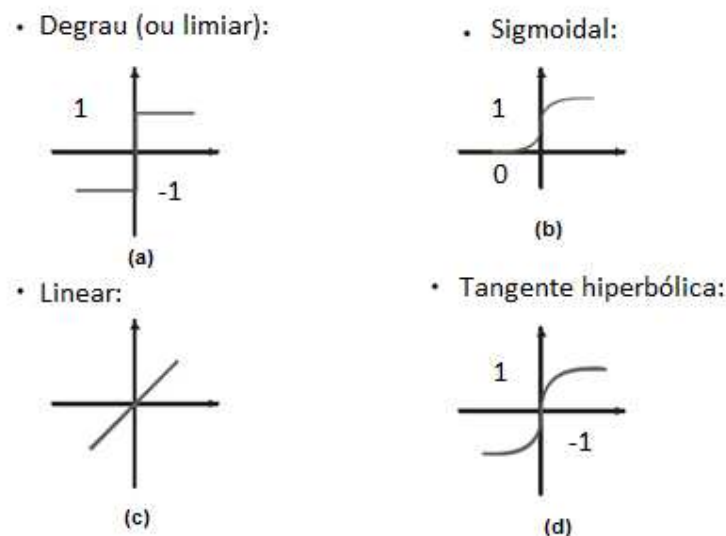


Figura 2.3: Funções de ativação: (a) Limiar, (b) Logística (Sigmóide), (c) Linear, (d) Tangente hiperbólica.

Existem varias funções utilizadas nas RNA dentre as quais se destacam as descritas abaixo:

✓ Função de ativação linear (identidade)

A função de ativação Linear é muito utilizada nas unidades que compõem a camada de saída das arquiteturas MLP. A expressão é dada por:

$$Y = \alpha X + \beta \quad (2.13)$$

em que α é um número real que define a saída linear para os valores de entrada, Y é a saída, X é a entrada e β é bias ($\beta \in \mathfrak{R}$). A saída linear com $\alpha = 1$ simplesmente repete o sinal que entra no neurônio na sua saída (UNICAMP, 2015).

✓ Função Logística (Sigmoidal)

A função de ativação sigmoide é muito comum nas RNA. Essa função assume valores contínuos em um intervalo entre 0 e 1, com uma característica muito importante para as RNA, ser diferenciável (UNICAMP, 2015).

$$Y = \frac{1}{1 + e^{-\alpha X}} \quad (2.14)$$

em que α é a suavidade da curva.

✓ Função Tangente Hiperbólica

A função logística apresenta apenas valores de ativação no intervalo (0, 1), em muitos casos ela é substituída pela função tangente hiperbólica, que preserva a forma sigmoide da função logística, porém assume valores positivos e negativos (-1, 1) (UNICAMP, 2015). A função tangente hiperbólica é dada pela expressão:

$$Y = \frac{e^{\alpha X} - e^{-\alpha X}}{e^{\alpha X} + e^{-\alpha X}} \quad (2.15)$$

✓ Função Arco-Tangente

Esta função pode ser apresentada como uma alternativa à função tangente hiperbólica para a implementação computacional, pois requer menos cálculos para sua elaboração. Ela possui valores de ativação no intervalo $(-\pi/2, \pi/2)$ (UNICAMP, 2015). A função arco tangente é dada pela expressão abaixo:

$$Y = a \tan(\alpha X) \quad (2.16)$$

No final da década de 1950, Roseblatt criou uma rede de múltiplos neurônios disposto em camadas do tipo discriminadores lineares e chamou essa rede de Perceptron, essa foi considerada o primeiro modelo de redes neurais. Foi um modelo matemático capaz de aprender por meio de exemplos. Inspirado no princípio Hebbiano de treinamento, os pesos de suas conexões são ajustados de acordo com os padrões apresentados dos mesmos. As arquiteturas das Redes perceptrons são organizadas em camadas. Na camada de entrada os estímulos são apresentados a rede. Nas camadas ocultas ou intermediárias ocorre a maior parte do processamento. Nas camadas de saída é a resposta dos padrões que são apresentados a rede. (BRAGA et al., 2012).

Os problemas mais complexos são resolvidos pelos modelos Multilayer Perceptron (MLP). As características do modelo são: tem pelo menos três camadas (entrada; escondida e saída); camada de entrada apenas propaga as entradas; cada unidade nas demais camadas atua como uma unidade perceptron com algumas modificações. (VALENÇA, 2011). A estrutura descrita acima está representada pela figura que se segue:

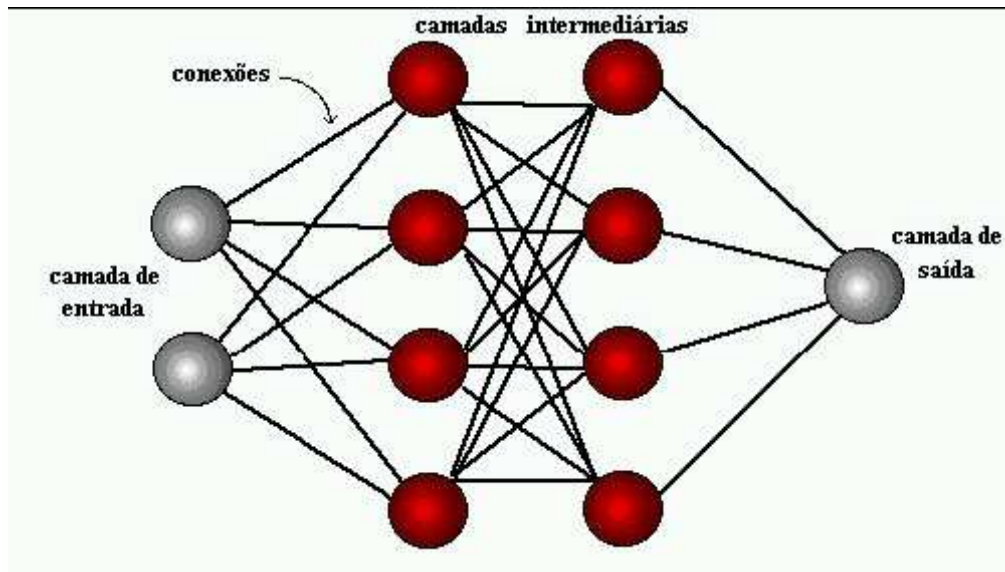


Figura 2.4: Estrutura básica de uma MLP, (TATIBANA; KAETSU, 2014).

Para o entendimento das redes neurais artificiais a arquitetura é um parâmetro importante, é por ela que se limita o tipo de problema que pode ser tratado pela rede. A arquitetura de uma RNA consiste no número de camadas da rede, número de neurônios em cada camada, tipo de conexão em cada camada e tipo de conexão entre os neurônios. A definição desses parâmetros esclarece o tipo de problema que pode ser tratado pela rede (VALENÇA, 2011).

A utilização de camadas ocultas permitem as RNAs implementar superfícies de decisão mais complexas. Essas camadas permitem que seus elementos se organizem de forma que cada elemento aprende a reconhecer características diferentes da entrada, assim, o algoritmo de treinamento decide quais características devem ser extraídas do conjunto de treinamento. Existem uma desvantagem em utilizar camada escondida é que o aprendizado se torna muito mais difícil (VIEIRA; ROISENBERG, 2015).

A propriedade mais importante das redes neurais é a capacidade de aprender e com isso melhorar seu desempenho. Isto é obtido pelo treinamento que é um processo iterativo de ajustes aplicado a seus pesos. O aprendizado ocorre quando a rede neural alcança uma solução generalizada para uma classe de problemas. (HAYKIN, 2001).

Diamantopoulou (2006) definiu aprendizagem como o processo pelo qual os parâmetros de uma rede neural são ajustados por meio de uma forma ininterrupta de estímulo pelo ambiente no qual a rede está atuando, sendo o tipo específico de

aprendizagem realizada determinado pela maneira como ocorrem os ajustes realizados nos parâmetros. No processo de aprendizado da rede há um algoritmo que consiste num conjunto de procedimentos para adaptar os parâmetros de uma RNA para que a mesma possa aprender uma tarefa específica (HAYKIN, 2001; VALENÇA, 2011).

O aprendizado acontece quando a rede adquire uma solução generalizada para um conjunto de problemas. Essa generalização é obtida por meio de um grupo de regras bem definidas para a solução de um problema, na qual a rede aprende uma determinada função. Essas regras são denominadas de algoritmo de treinamento ou aprendizado. (HAYKIN, 2001).

Uma característica da aprendizagem supervisionada é a utilização do erro de saída, diferença entre a resposta desejada e a resposta obtida, como orientador do ajuste dos pesos. O objetivo da aprendizagem é aproximar a saída desejada da obtida ajustando os parâmetros da rede para encontrar uma ligação entre os pares de entrada e saída fornecidos. Uma desvantagem para esse aprendizado é que na ausência do supervisor externo, o erro de saída, a rede não conseguirá prever para situações não cobertas pelos exemplos do treinamento. No aprendizado não-supervisionado não há o supervisor externo, ou seja, a rede não utiliza o erro de saída como orientação (BRAGA et al., 2012; CASTRO et al., 2013).

Apesar da afinidade entre o aprendizado supervisionado e o aprendizado dos seres humanos, muitos dos sistemas biológicos ocorrem por meio de aprendizado não-supervisionado, dois exemplos são os estágios iniciais dos sistemas de visão e audição (VALENÇA, 2011). A escolha de um tipo de aprendizagem e da rede neural artificial é determinada pelo tipo de problema a ser resolvido. Assim, os problemas tratáveis por meio de redes neurais se enquadram nas tarefas de aprendizagem: agrupamento de dados, classificação de padrões, otimização, recuperação de conteúdo e controle, aproximação de função, predição (séries temporais), dentre outros (LIMA et. al., 2014).

Os dados coletados para o treinamento da rede são divididos em três categorias: dados de treinamento, que serão utilizados para o treinamento da rede, os dados de validação usados para verificar a eficiência da rede quanto a sua capacidade de generalização durante o treinamento, e podendo ser empregado como critério de parada do treinamento e dados de teste, que serão utilizados para

verificar seu desempenho sob condições reais de utilização (TATIBANA; KAETSU, 2014).

O algoritmo mais conhecido para o treinamento de redes do tipo multicamadas é a retropropagação do erro. Esse processo tem como objetivo melhorar o desempenho da rede por meio do ajuste de pesos visando minimizar o erro. Essa variação é calculada para os pesos conectados às unidades de processamento da camada de saída. Prosseguindo, usa-se a informação do erro para mudar os pesos chegando às unidades da penúltima camada e assim por diante até chegar à primeira camada, o treinamento ocorre em duas fases, *forward* e *backward* (BULLINARIA, 2014). Como padrão, incorpora-se o limiar diferente de zero a função de propagação.

$$\sum_i W_{ji} X_i > T_i \Rightarrow \sum_i W_{ji} X_i - T_i > 0 \quad (2.17)$$

Em que W_{ji} são os pesos, X_i os padrões de entrada e T_i o limiar.

O erro de um neurônio de saída j na iteração n é definido por:

$$e_j(n) = Y_d^j(n) - Y_j(n) \quad (2.18)$$

Sendo a soma dos erros quadráticos de todos os neurônios de saída na iteração n definida pela Equação (2.15) (BRAGA et. al., 2012).

$$\varepsilon(n) = \frac{1}{2} \sum_j e_j^2(n) \quad (2.19)$$

As principais etapas da operação rede MLP consistem em: apresenta-se um padrão de entrada e determina-se a saída da rede; comparar a saída obtida com a desejada na camada de saída; a regra de aprendizagem ajusta os pesos aferentes a cada camada de saída; propaga-se esse ajuste, em função do erro, no sentido saída-entrada; repete-se o processo até se atingir o critério de parada (BRAGA et al., 2012).

O interesse da RNA tem seu motivo nas variadas propriedades úteis e sua habilidade à adaptação podendo ser retreinada a partir de modificações ajustando seus pesos em tempo real, aprendizagem generalizada com os dados de treinamento, não linearidade e identidade neurobiológica (BRAGA et al., 2012; HAYKIN, 2001; BULLINARIA, 2014).

Uma alternativa que vem sendo implantada recentemente aos modelos volumétricos e com resultados satisfatórios é o uso da técnica de redes neurais artificiais (RNA) no manejo florestal (BINOTI, 2013).

Com a evolução do tempo, a Ciência Florestal acaba passando por mudanças quanto a seus aspectos de análises gerando com isso novas tendências e necessidades de aprimoramentos a fim de obter resultados mais precisos e viáveis.

Ainda é recente o uso de RNA na pesquisa florestal, entretanto se apresenta como uma nova ferramenta apresentando resultados interessantes e sendo importante na tomada de decisões. Há muitos estudos que comprovam a eficiência da aplicação da RNA na área florestal.

Oliveira (2008) mostrou que é possível utilizar técnicas de RNA, como ferramenta de previsão de defeitos em estradas florestais, concluindo que o uso de RNAs garantiu resultados mais precisos e confiáveis nos trechos de unidades avaliados.

Schoeninger et al. (2008) apresentaram a viabilidade do uso de RNA como alternativa para o mapeamento de biomassa e carbono em grandes extensões florestais. Concluindo que a automatização e a integração de dados para a quantificação de dados de biomassa e carbono, utilizando como metodologia de apoio imagens de satélite e RNAs, podem minimizar os custos e melhorar a precisão e exatidão dos mesmos.

Silva et. al. (2009) ao ajustar o modelo de Shumacher e-Hall e aplicar redes neurais para estimar o volume de árvores de eucaliptos concluiu que o modelo de Shumacher e-Hall apresenta bons resultados e pode ser usado na sua forma linear. Além disso afirmou que as RNAs apresentou ao problema boa adequação sendo uma boa alternativa.

Gorgens et al. (2009) utilizaram a RNA para estimar o volume de árvores, e concluíram que sua capacidade de generalização permitiu a utilização apenas uma rede para realizar a previsão de volume de árvores de locais e de espécies diferentes.

Silva et al. (2009) avaliaram a aplicação de RNA para estimação do volume de madeira de eucalipto comparando com o modelo de Schumacher e-Hall, concluíram que o uso de redes neurais artificiais apresentaram bons resultados.

Pereira (2009) publicou um estudo comparativo entre as funções discriminantes de Fisher e de Anderson e a RNA em relação ao número de

classificações erradas de indivíduos pertencentes a diferentes populações, com distintos níveis de dissimilaridade. Concluindo que a RNA se mostraram uma técnica promissora em problemas de classificação.

Soares et al. (2012) estimaram diâmetros das árvores utilizando uma rede neural Perceptron de Múltiplas Camadas a partir de apenas três medidas na base da árvore. Concluindo resultados satisfatórios por meio do erro da estimativa.

Binoti et al. (2014), concluíram que a estimação do volume de árvores por RNA pode ser feita com configurações diferentes, alterando o número de neurônios na camada oculta, as funções de ativação, os algoritmos e os parâmetros de treinamento. Concluíram também que redes neurais simples podem ser utilizadas com eficiência para estimação do volume de árvores.

Capítulo 3: Material e Método

3.1. Área experimental

A região do Araripe, que está localizada no Oeste de Pernambuco, é formada pelos municípios de Araripina, Bodocó, Cedro, Dormentes, Exu, Granito, Ipubí, Moreilândia, Oricuri, Parnamirim, Santa Cruz, Santa Filomena, Serrita, Terra Nova e Trindade. Esses municípios correspondem a 18% da área total do estado de Pernambuco. O arranjo produtivo local (APL) do gesso, denominado de Pólo Gesseiro do Araripe está situado nessa região.

A introdução de clones de eucaliptos na região iniciou no ano de 2002, com o projeto denominado “Experimentação Florestal no Pólo Gesseiro do Araripe” financiada pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Foram testados 15 clones de *Eucalyptus* spp doados pela Companhia Suzano na Estação Experimental do Instituto Agrônomo de Pernambuco (IPA), na Chapada do Araripe, município de Araripina, com coordenadas geográficas são 07°27'37”S e 40°24'36”W com altitude 831(SILVA, 2008/2009).

Os tratamentos foram distribuídos em parcelas de dimensão 14m x 21m com espaçamento 2m x 3m, e 49 plantas por parcelas sendo consideradas 25 plantas na área útil das parcelas que foram mensuradas a cada 6 meses, iniciando em março de 2002, e terminando o corte do experimento em setembro de 2009, quando houve a cubagem rigorosa de todas as árvores. Os dados utilizados para realização deste estudo são referentes a 1.862 árvores, nas quais foram coletos dados de volume cubado pelo método de Smalian, altura total e CAP, sendo que essa medida foi convertida em DAP. A descrição dos tratamentos estão nas tabelas 3.1 e 3.2.

Tabela 3. 1: Descrição dos tratamentos (tipos de clones) usados no experimento.

Tratamento	Descrição do clone
T01	Híbrido de <i>Eucalyptus tereticornis</i> (cn)
T02	Híbrido de <i>E. urophylla</i> x <i>E. tereticornis</i> de Laura (pc)
T03	Híbrido de <i>E. urophylla</i> x <i>E. tereticornis</i> de Laura (pc)
T04	Híbrido de <i>E. urophylla</i> x <i>E. tereticornis</i> de Laura (pc)
T05	Híbrido de <i>E. urophylla</i> x <i>E. tereticornis</i> de Laura (pc)
T06	Híbrido de <i>E. urophylla</i> x <i>E. tereticornis</i> x <i>E. pellita</i> (pc)
T07	Híbrido de <i>E. urophylla</i> (cn)
T08	Híbrido de <i>E. brassiana</i> (cn)
T09	Híbrido de <i>E. urophylla</i> x <i>E. tereticornis</i> de Laura (pc)
T10	Híbrido de <i>E. urophylla</i> x <i>E. tereticornis</i> de Laura (pc)
T11	Híbrido de <i>E. urophylla</i> (cn)
T12	Híbrido de <i>E. brassiana</i> (cn)
T13	Híbrido de <i>E. brassiana</i> (cn)
T14	Híbrido de <i>E. urophylla</i> (cn)
T15	Híbrido de <i>E. brassiana</i> (cn)

Em que: pc = polinização controlada e cn = cruzamento natural.

Tabela 3.2: Freqüência dos tratamentos de cada clone de *Eucalyptus spp* do experimento.

Tratamento	Qtd	%
T01	147	8,2%
T02	156	8,7%
T03	137	7,2%
T04	151	8,7%
T05	104	4,7%
T06	135	6,8%
T07	126	7,1%
T08	113	6,0%
T09	149	8,1%
T10	148	7,9%
T11	105	5,1%
T12	69	3,2%
T13	121	6,6%
T14	134	8,1%
T15	67	3,3%
Total geral	1862	100%

Qtd = número de árvores de cada tratamento

3.2. Modelos

Para estimar o volume dos *Eucalyptus spp* foi utilizado o método de redes neurais e os resultados foram comparados com os modelos Schumacher e Hall e Spurr. A função estimada foi:

$$V = f(\text{DAP}, \text{Ht}) \quad (3.1)$$

O treinamento da RNA foi realizado visando estimar o volume das árvores utilizando como neurônios na camada de entrada o DAP e a Ht. Além disso, foi estimado modelos considerando a utilização de uma variável indicadora (tipo de clone) como entrada para avaliar o desempenho da rede, considerando uma

característica importante das redes neurais artificiais que é não precisar de suposição sobre as relações das variáveis utilizadas no ajuste do modelo.

A figura 5 ilustra as medições na árvore.

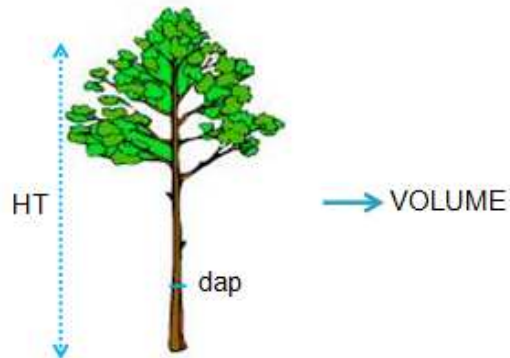


Figura 3.1: Árvore ilustrada com as suas dimensões. (Fonte: Campos; Leite, 2009).

Os parâmetros foram estimados por meio do Método de Mínimos Quadrados Ordinários, com o auxílio do software Statística 8.0 (STATSOFT, 2012). Além disto, optou-se por utilizar as equações em suas formas lineares aplicando a transformação logarítmica no modelo Schumacher e-Hall, a fim de corrigir eventuais problemas de heterocedasticidade de variância e de normalidade dos erros. O modelo de Spurr é linear nos parâmetros se considerarmos $X = DAP^2H$ podendo ser estimado facilmente pelo método clássico de regressão.

O cálculo desse erro estimado é simples, corresponde à diferença entre o valor observado e o valor estimado.

$$\hat{\epsilon}_i = Y_i - \hat{Y}_i \quad i = 1, 2, \dots, N \quad (3.2)$$

Em que:

$\hat{\epsilon}_i$ = erro da estimativa para a árvore i .

Y_i = valor observado da árvore i .

\hat{Y}_i = valor estimado da árvore i .

N = número total de árvores

Como funções de ativação da camada oculta e de saída foram testadas as funções: exponencial, identidade, tangente hiperbólica e sigmoide. Foram treinadas 20 redes, sendo avaliado o desempenho das cinco melhores redes para posterior comparação entre os modelos, o critério de seleção das redes se baseou nos índices de treinamento da rede, fornecido pelo software na fase de treinamento e no coeficiente de determinação ajustado ($R^2_{aj.}$).

As RNAs treinadas foram do tipo *perceptrons* de múltiplas camadas, conhecida como MLP, de forma supervisionada, o algoritmo empregado foi a retropropagação do erro (*Back propagation*), a regra de aprendizado conhecida que minimiza do erro médio quadrático. Esta forma de aprendizagem ajusta os pesos das conexões entre os neurônios da rede de acordo com o erro e tem por objetivo encontrar um conjunto de pesos e polarizações que reduzam a função erro (HAYKIN, 2001).

As redes foram treinadas utilizando o *Software* Statistica 8 (STATISOFT, 2012).

A fim de comparação utilizou-se o modelo de os modelos Schumacher e-Hall (1933) (3.3) e Spurr 1952 (3.4).

$$V_i = \beta_0 \cdot DAP_i^{\beta_1} \cdot Ht_i^{\beta_2} \cdot \varepsilon_i \quad (3.3)$$

$$V_i = \beta_0 + \beta_1 \cdot DAP_i^2 Ht_i + \varepsilon_i \quad (3.4)$$

3.3. Validação e comparação dos modelos

O critério de avaliação de desempenho das redes neurais artificiais e dos modelos foi realizado com base o coeficiente de determinação ajustado ($R^2_{aj.}$) entre os volumes observados e estimados, a raiz quadrada do erro médio percentual (RMSE), o erro padrão da estimativa (Syx) e análise visual dos gráficos de dispersão dos resíduos por volume (m^3) e histograma de frequência de resíduos, sendo escolhidas as RNA que apresentaram maiores coeficientes de correlação e determinação e menores RMSE e Syx, com os gráficos de dispersão dos resíduos e histograma de resíduos não tendenciosos, com maior ocorrência de observações em torno de zero.

$$RMSE(\%) = 100 \sqrt{\left(\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n} \right)} / \bar{Y} \quad (3.6)$$

$$Syx(\%) = \frac{\sqrt{\left(\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-1} \right)}}{\bar{Y}} 100 \quad (3.7)$$

Em que:

Y = volume total observado.

\hat{Y} = volume total estimado.

\bar{Y} = média dos volumes totais observados.

n: número total de árvores.

A interpretação do coeficiente de determinação é a proporção de variação total do volume que é explicado pelo modelo. Seu resultado pode ser obtido pela razão da soma de quadrados da regressão (SQR) e a soma de quadrados total (SQT). A interpretação do coeficiente de determinação ajustado segue o mesmo modelo.

$$R^2 = \frac{SQR}{SQT} \quad (3.8)$$

$$R^2_{aj} = 1 - (1 - R^2) \left(\frac{n-1}{n-k} \right) \quad (3.9)$$

Em que:

k: número de coeficientes da equação;

n: número de árvores

A análise dos gráficos de dispersão dos resíduos dá uma idéia de quanto às estimativas pontuais estão próximas dos valores observados. Em uma estimativa perfeita, na qual cada uma das árvores tem o seu valor estimado perfeitamente, o gráfico representa uma reta no plano cartesiano com inclinação 1 (um). Portanto quanto mais o gráfico se aproxima da situação descrita, melhor a estimacão geral.

Para os modelos logarítmicos o erro padrão da estimativa foi corrigido na escala original da variável dependente, para possibilitar a comparação com o modelo aritmético.

Existem alguns métodos para descrever o quanto duas populações são diferentes com relação a alguma característica medida e observada. Uma medida de separação muito utilizada é o Kolmogorov-Smirnov (KS) e no presente contexto o método foi utilizado para descrever o quanto os valores observados e estimados são próximos. Quanto mais similar os conjuntos testados menor o valor da estatística, ou seja, quanto mais próximo o valor estimado do real menor o KS.

Esta estatística tem origem no teste de hipótese não paramétrico de Kolmogorv-Smirnov na qual se deseja testar se duas funções de distribuição associadas a duas populações são distintas ou não.

A hipótese a ser testada é:

$$\begin{cases} H_0 : \text{As distribuições são semelhantes} \\ H_1 : \text{As distribuições não são semelhantes} \end{cases}$$

Dada uma amostra observada X_1, \dots, X_n e um controle, amostra Y_1, \dots, Y_n , deve-se o quanto estão separadas as funções de distribuições. A estatística utilizada para o teste é a diferença máxima absoluta entre as funções $F_n(x)$ e $G_m(Y)$:

$$D_n = \sup_x |F_n(x) - G_m(Y)| \quad (3.10)$$

$F_n(x)$: representa a função de distribuição acumulada para x

$G_m(Y)$: representa a função de distribuição acumulada para y

3.4. Uso de diferentes tamanhos de amostras

Para verificar a eficiência do uso de RNA quando comparada com modelos tradicionais ajustados ou análise de regressão, usou-se diferentes tamanhos de amostras nas quais foram ajustadas as RNA e o melhor modelo de regressão ajustado, o Schumacher e-Hall. A metodologia empregada nesse objetivo foi a mesma empregada no modelo geral.

Foram utilizados os seguintes tamanhos de amostra de árvores sorteadas aleatoriamente entre 1862 árvores.

Tabela 3.3: Definição das amostras e erros amostrais da base de *Eucalyptus spp* do experimento.

Amostra	Quantidade	Erro _A %
Amostra 1	25 árvores	28,27%
Amostra 2	50 árvores	15,44%
Amostra 3	100 árvores	9,65%
Amostra 4	200 árvores	7,57%
Amostra 5	400 árvores	5,44%
Amostra 6	800 árvores	3,80%
Amostra 7	1600 árvores	2,66%
Base Completa	1864 árvores	-

O Erro amostral (**Erro_A %**) de cada amostra é dado pela formula abaixo:

$$E_A \% = \left(\frac{t_{\alpha/2} \cdot S_{\bar{Y}}}{\bar{X}} \right) \cdot 100 \quad (3.5)$$

Para critério de avaliação de desempenho das redes neurais artificiais nas diferentes amostras e do modelo foi realizado com base no coeficiente de determinação ajustado ($R^2_{aj.}$) entre os volumes observados e estimados, a raiz quadrada do erro médio percentual (RMSE), o erro padrão da estimativa (S_{yx}) e análise visual dos gráficos de dispersão dos resíduos por volume (m^3), além de análise visual de gráficos de comparação dos indicadores em função do tamanho da amostra para observar a evolução dos resultados.

Capítulo 4: Resultados

4.1. Resultado geral

O perfil métrico da amostra de *Eucalyptus spp.* pode ser observado na tabela 4.1, a seguir, em que, salientam-se as amplitudes do volume, diâmetro a altura do peito (DAP) e altura da árvore (Ht). Em complementação, vale ressaltar que as medidas apresentaram leve desvio comparado a magnitude das observações pontuais em torno da média, no entanto, o volume requer uma atenção especial, uma vez que seu alto desvio associado a moderada amplitude caracteriza prováveis unidades extremas, ou seja, *Eucalyptus* com volumes, ou mensurações de volume, acima do padrão aparentemente uniforme da amostra coletada. (extremos explicados pelo baixo DAP)

Tabela 4.1: Estatísticas descritivas das variáveis contínuas dos *Eucalyptus spp.* utilizadas no estudo. Em que: o volume m^3 , DAP cm é o diâmetro a altura do peito e Ht m é a altura da árvore.

Estatística Descritiva	Variáveis		
	Volume (m^3)	DAP (cm)	Ht (m)
Mínimo	0,01	4,50	7,29
Máximo	0,38	25,97	21,85
Mediana	0,07	11,40	15,53
Média	0,09	11,97	15,52
Desvio	0,05	2,76	1,91

A tabela 4.2 apresenta a estimativa dos modelos propostos para comparação dos resultados na predição de volume com a técnica de redes neurais artificiais.

Tabela 4.2: Estatísticas do ajuste dos modelos volumétricos de *Eucalyptus spp.*

Modelo	Parâmetro	Estimativa	Erro Padrão	T	p-valor
Schumacher e-Hall	β_0	0,000062	< 0,0000001	<0.001	<0.001
	β_1	1,682530	0,013684	<0.001	<0.001
	β_2	1,094109	0,028670	<0.001	<0.001
Spurr	β_0	0,011243	0,000471	<0.001	<0.001
	β_1	0,000031	< 0,0000001	<0.001	<0.001

Ao estudar o ajuste dos modelos de Schumacher e Hall e Spurr, como mostra a tabela 4.2, o p-valor < 0,001 confirma a adequacidade de ambos os modelos. Sendo assim, aceite o ajuste: $\text{Volume}_i = 0,000062 \times \text{DAP}_i^{1,682530} \times \text{Ht}_i^{1,094109}$, para o modelo de Schumacher e-Hall e, de forma equivalente, para o modelo de Spurr, $\text{Volume}_i = 0,011243 + 0,000031 \cdot \text{DAP}^2 \text{Ht}_i$.

Após a obtenção das estimativas, foram comparadas suas estatísticas (Tabela 4.3) e realizada análise visual dos gráficos de resíduos (Figura 4.1), tanto dos modelos de regressão quanto das RNA selecionadas com melhor capacidade em estimar o volume de árvores de eucalipto.

Tabela 4.3: Estatísticas de ajuste para o volume de *Eucalyptus spp.* Em que: R^2_{aj} . (Coeficiente de determinação de ajustado), RMSE% (raiz quadrada do erro médio percentual), Syx% (erro padrão da estimativa) e KS (Kolmogorv-Smirnov).

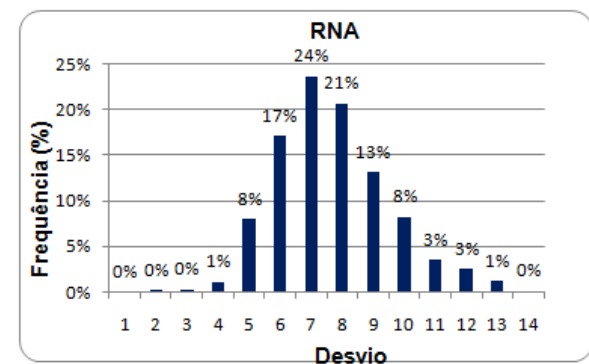
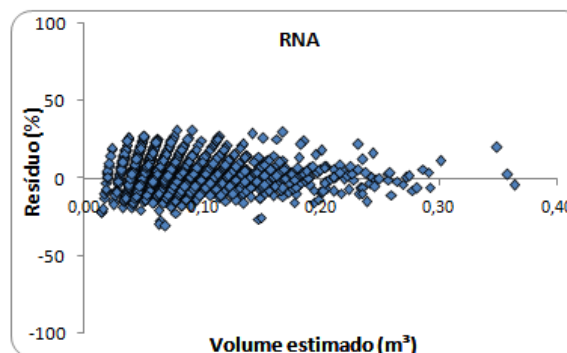
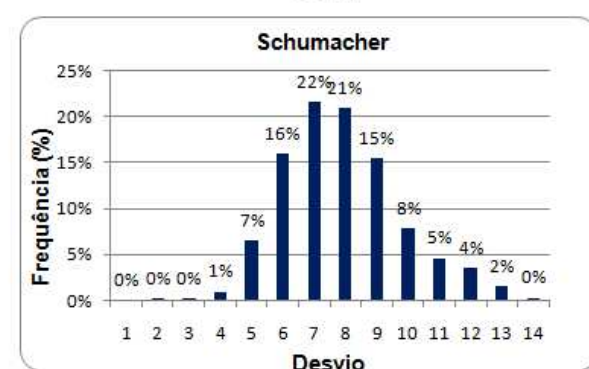
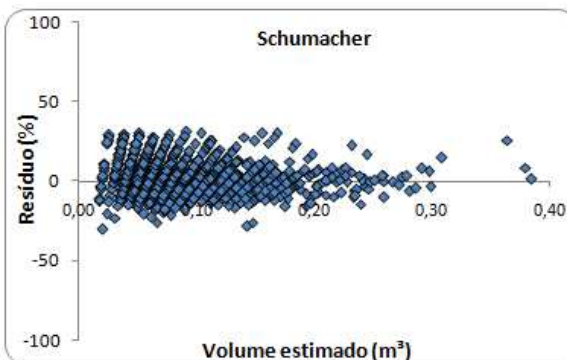
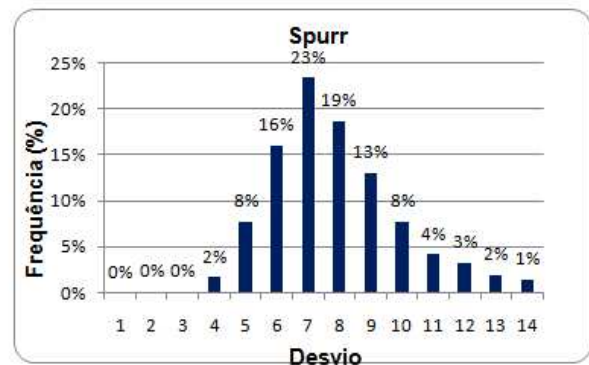
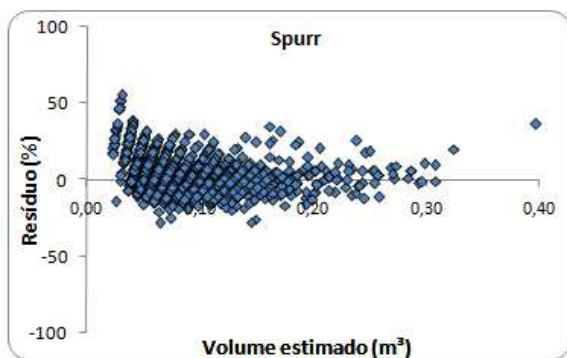
Validação	Spurr	Schumacher	RNA 1	RNA 2	RNA 3	RNA 4	RNA 5
		e Hall					
R^2	0,94926	0,95542	0,97721	0,97678	0,97709	0,97673	0,97703
R^2_{aj}	0,94920	0,95537	0,97718	0,97675	0,97706	0,97670	0,97700
RMSE%	3,59506	3,36966	3,30220	3,32110	3,29317	3,32326	3,30950
Sxy (%)	12,26156	11,46964	11,24014	11,30447	11,20939	11,31184	11,24096
KS%	11,10000	10,00000	8,40000	8,60000	8,30000	8,60000	8,40000

A tabela 4.4 apresenta as arquiteturas, camadas de entrada, intermediária ou oculta e camada de saída de cada rede neural avaliada e suas respectivas funções de ativação. Observa-se que a RNA 1 além de apresentar um bom desempenho foi utilizada uma configuração de neurônios na camada oculta mais simples para aplicação das RNAs.

Tabela 4.4: Definição das RNAs

RNAs	Arquitetura	Função de Ativação
RNA 1	MLP 2-5-1	(Tanh - Identity)
RNA 2	MLP 2-6-1	(Logistic - Identity)
RNA 3	MLP 2-3-1	(Exponential - Exponential)
RNA 4	MLP 2-3-1	(Exponential - Identity)
RNA 5	MLP 2-7-1	(Logistic - Exponential)

Os gráficos das páginas seguintes (Figura 4.1) demonstram a distribuição dos resíduos por meio da dispersão e histograma. Essa visualização é importante para identificar a adequacidade dos modelos ajustados.



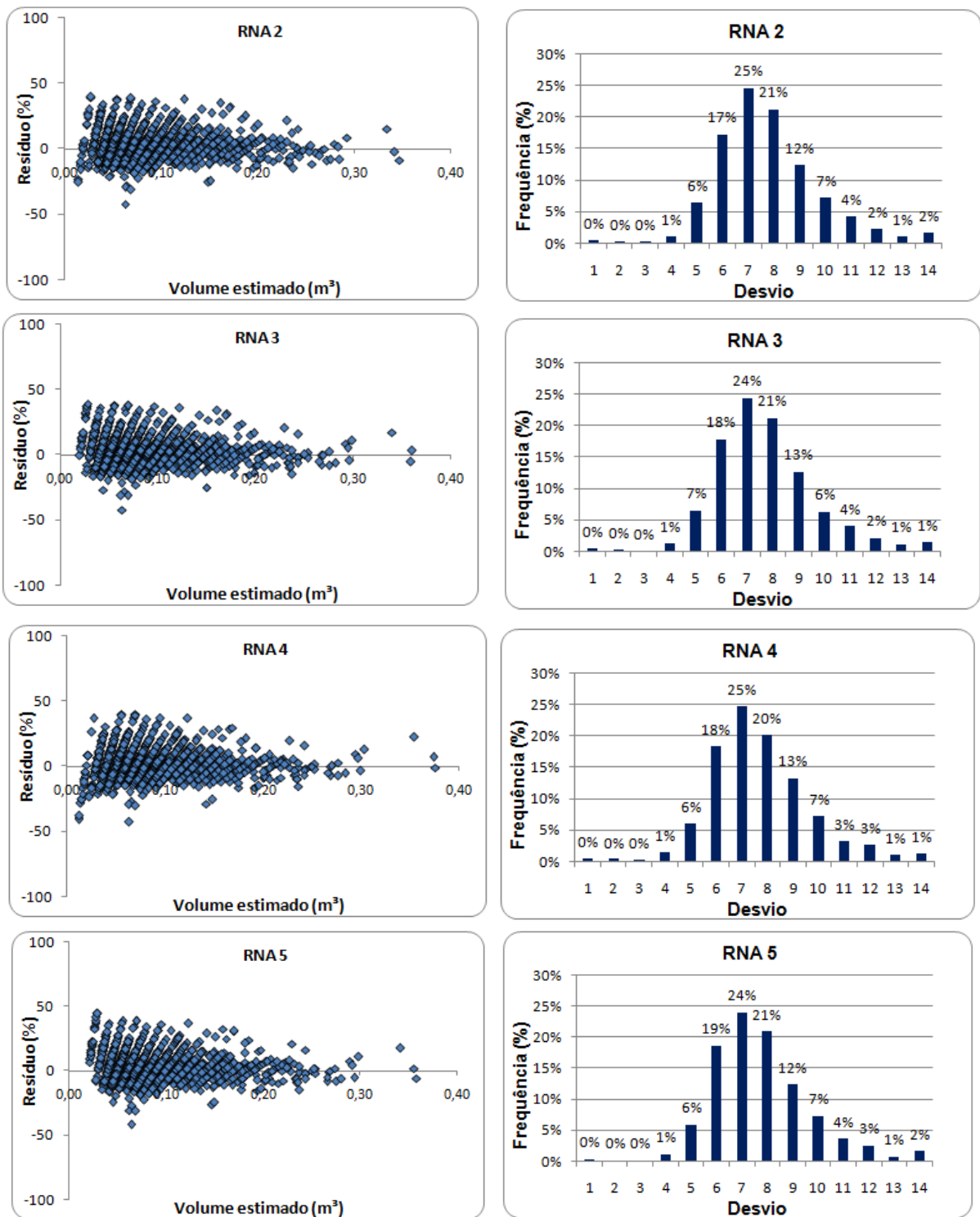


Figura 4.1: Gráficos de dispersão dos resíduos e histograma de freqüência de resíduos evidenciando o comportamento dos desvios encontrados entre volume observado e volume estimado.

Observando os resultados apresentadas pelo modelo de Schumacher e Hall, nota-se $R^2_{aj} = 95,5\%$ e $RMSE\% = 3,36\%$, evidenciando bom ajuste do modelo. O modelo de Spurr apresentou resultados próximos $R^2_{aj} = 94,9\%$ e $RMSE = 3,59\%$. De

acordo com estatísticas apresentadas pelas RNAs selecionadas, pode-se observar coeficientes de determinação ajustado com valores superiores aos modelos de regressão, em torno de 97%, e RMSE% em torno de 3%. Ao analisar visualmente os gráficos de dispersão dos resíduos e histograma de frequência dos resíduos para os modelos e RNA, pode-se notar que as estimativas pontuais estão próximas dos valores observados, não apresentando distribuições tendenciosas. Além disso, o valor do Syx% apresenta um bom desempenho.

Citando as estimativas que obteve resultados satisfatórios, destaca-se a RNA 1. Sendo esses resultados eficazes e eficientes em estimar o volume de árvores de eucalipto, com gráfico de dispersão dos resíduos sem apresentar tendências e dispersão dos resíduos distribuídos de forma independente, com maior ocorrência em torno de zero e distribuição normal, ou seja, não demonstrando tendências de subestimar ou superestimar os volumes.

Quando analisado o teste Kolmogorv-Smirnov apresenta uma boa concordância entre o valor estimado e o valor real nos modelos, aceita-se a hipótese nula, considerando todos os métodos válidos para estimar volume de eucaliptos spp.

Resultados satisfatórios nas estimativas das RNAs foram encontrados com configurações de apenas 3 a 7 neurônios na camada oculta, comprovando que não são necessárias estruturas muito complexas para estimação do volume de árvores. De acordo com Bullinaria (2014), recomenda-se então configurações de neurônios na camada oculta mais simples para aplicação das RNAs. As funções de ativação testadas, tangente, exponencial, identidade e sigmoideal, apresentaram resultados satisfatórios para estimação do volume, tanto na camada de oculta quanto na camada de saída.

4.2. Resultados por diferentes tamanhos de amostras

Uma análise em paralelo foi considerar adicionar a variável categórica no modelo, tipo de tratamento (Tipos de clone de *Eucalyptus spp.*), e avaliar se melhora o desempenho da rede neural. Foram comparadas as estatísticas das redes neurais (Tabela 4.5) e realizada análise visual dos gráficos de resíduos (Figura 4.2), as RNA selecionadas com melhor capacidade em estimar o volume de árvores de eucalipto. Além disso, consideramos sete (7) amostras aleatórias além da base completa para

avaliar o desempenho entre o melhor modelo de regressão ajustado, Schumacher e Hall, a rede neural artificial normal e a RNA considerando o tratamento (clone), com o objetivo de verificar se há mudanças na performance em relação ao tamanho da amostra.

O perfil métrico por amostra de *Eucalyptus spp.* pode ser observado na tabela 4.1, a seguir, em que, salientam-se as amplitudes do volume, diâmetro a altura do peito (DAP) e altura da árvore (Ht).

Tabela 4.5: Estatísticas descritivas por amostra das variáveis contínuas dos *Eucalyptus spp.* utilizadas no estudo. Em que: o volume m³, DAP cm é o diâmetro a altura do peito e Ht m é a altura da árvore.

Amostra / Variáveis		Mínimo	Máximo	Mediana	Média	Desvio
Amostra 1	Volume (m ³)	0,05	0,31	0,07	0,10	0,07
	DAP (cm)	9,28	21,58	11,46	12,64	3,29
	Ht (m)	12,18	20,70	16,00	16,06	1,77
Amostra 2	Volume (m ³)	0,02	0,26	0,08	0,09	0,05
	DAP (cm)	6,80	19,26	11,57	12,27	2,84
	Ht (m)	11,30	19,80	15,77	15,50	1,77
Amostra 3	Volume (m ³)	0,01	0,22	0,07	0,08	0,04
	DAP (cm)	5,09	19,23	11,49	11,74	2,58
	Ht (m)	7,50	19,80	15,48	15,45	1,85
Amostra 4	Volume (m ³)	0,01	0,38	0,07	0,09	0,05
	DAP (cm)	4,50	25,97	11,42	11,97	2,77
	Ht (m)	8,45	21,85	15,49	15,48	1,93
Amostra 5	Volume (m ³)	0,01	0,38	0,07	0,09	0,05
	DAP (cm)	4,50	25,97	11,42	11,97	2,77
	Ht (m)	8,45	21,85	15,49	15,48	1,93
Amostra 6	Volume (m ³)	0,02	0,38	0,07	0,09	0,05
	DAP (cm)	5,64	25,97	11,40	11,97	2,76
	Ht (m)	8,45	21,85	15,59	15,51	1,89
Amostra 7	Volume (m ³)	0,01	0,38	0,07	0,09	0,05
	DAP (cm)	5,09	25,97	11,40	11,97	2,74
	Ht (m)	7,29	21,85	15,53	15,53	1,93
Base Completa	Volume (m ³)	0,01	0,38	0,07	0,09	0,05
	DAP (cm)	4,50	25,97	11,40	11,97	2,76
	Ht (m)	7,29	21,85	15,53	15,52	1,91

Foram comparadas, por amostra, as estatísticas das estimativas (Tabela 4.6) e realizada análise visual dos gráficos de resíduos (Figura 4.2), do melhor modelo de regressão ajustado, Schumacher e Hall, das RNA considerando e não o tratamento, tipo de clone, selecionadas com melhor capacidade em estimar o volume de árvores de eucalipto.

Tabela 4.6: Estatísticas das RNAs para o volume de *Eucalyptus spp.* considerando como entrada a variável categórica 'Tratamento' (tipo de clone de *Eucalyptus*).

Estatística	Amostra	RNA com Clone	RNA	Schumacher
R²aj	25	0,99917	0,99413	0,98808
	50	0,99724	0,98403	0,95968
	100	0,98936	0,9862	0,96756
	200	0,99183	0,98378	0,97256
	400	0,99025	0,98379	0,97351
	800	0,98267	0,97838	0,95741
	1600	0,98455	0,98105	0,95484
	1862	0,98050	0,97718	0,95537
RMSE (%)	25	1,02867	2,18045	2,22681
	50	1,35342	2,79569	3,16725
	100	2,18120	2,41915	2,49448
	200	2,19968	2,60782	3,03082
	400	2,28010	2,63863	2,64500
	800	3,00855	3,26196	3,31740
	1600	3,05747	3,34009	3,38756
	1862	3,01761	3,3022	3,36967
Syx (%)	25	3,24001	6,86776	7,48992
	50	4,52146	9,33976	10,93522
	100	7,63573	8,46871	8,83714
	200	7,56937	8,97384	9,01809
	400	7,76382	8,98462	9,05536
	800	10,2546	11,1183	11,3236
	1600	10,3946	11,3554	11,5003
	1862	10,2715	11,2401	11,4696

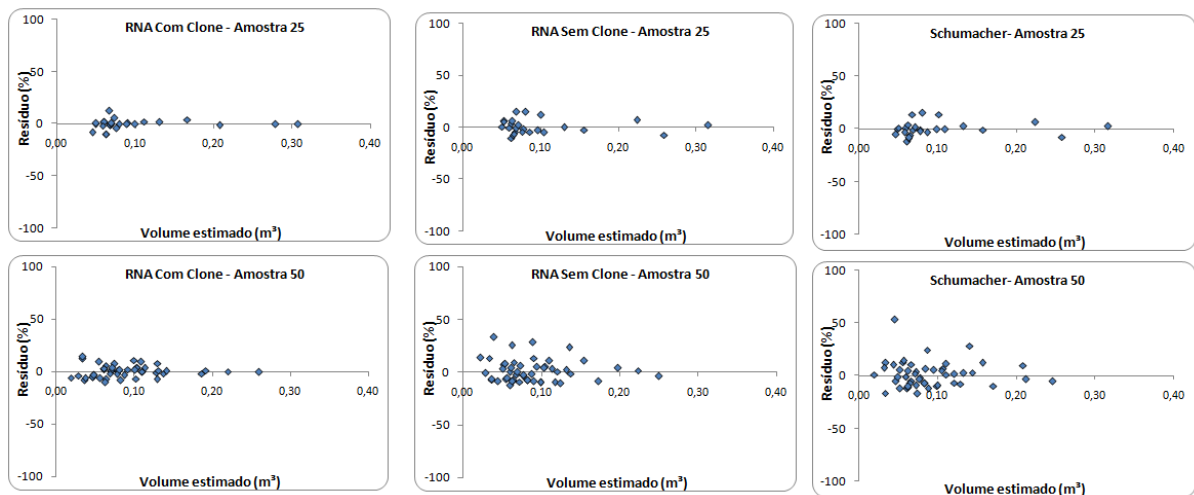
A tabela 4.7 apresenta as arquiteturas, camadas de entrada, intermediária ou oculta e camada de saída de cada RNA avaliada e suas respectivas funções de ativação para o estudo incrementando a variável categórica para estimativa do volume de *Eucalyptus spp.* A RNA apresentou um bom desempenho, porém sua

configuração de neurônios nas camadas ocultas foram perdendo a simplicidade para aplicação das RNAs devido a entrada de uma variável categórica..

Tabela 4.7: Definição das RNAs por amostra com e sem a variável categórica “Clone”.

RNAs	RNA com Clone		RNA sem Clone	
	Arquitetura	Função de Ativação	Arquitetura	Função de Ativação
RNA Amostra 1	MLP 17-12-1	(Logistic - Identity)	MLP 2-7-1	(Exponential - Identity)
RNA Amostra 2	MLP 17-15-1	(Exponential - Identity)	MLP 2-7-1	(Tanh - Identity)
RNA Amostra 3	MLP 17-9-1	(Exponential - Identity)	MLP 2-4-1	(Logistic - Identity)
RNA Amostra 4	MLP 17-14-1	(Exponential - Identity)	MLP 2-3-1	(Exponential - Identity)
RNA Amostra 5	MLP 17-12-1	(Exponential - Identity)	MLP 2-4-1	(Tanh - Identity)
RNA Amostra 6	MLP 17-12-1	(Exponential - Identity)	MLP 2-8-1	(Tanh - Identity)
RNA Amostra 7	MLP 17-5-1	(Logistic - Identity)	MLP 2-3-1	(Logistic - Identity)
RNA Base Completa	MLP 17-9-1	(Exponential - Logistic)	MLP 2-5-1	(Tanh - Identity)

Os gráficos abaixo (Figura 4.2) demonstram, por amostra, a distribuição dos resíduos por meio da dispersão das redes neurais considerando e não a variável categórica, tipo de tratamento ou clone, como mais uma entrada. Além do modelo de regressão de melhor desempenho na estimativa da base completa.



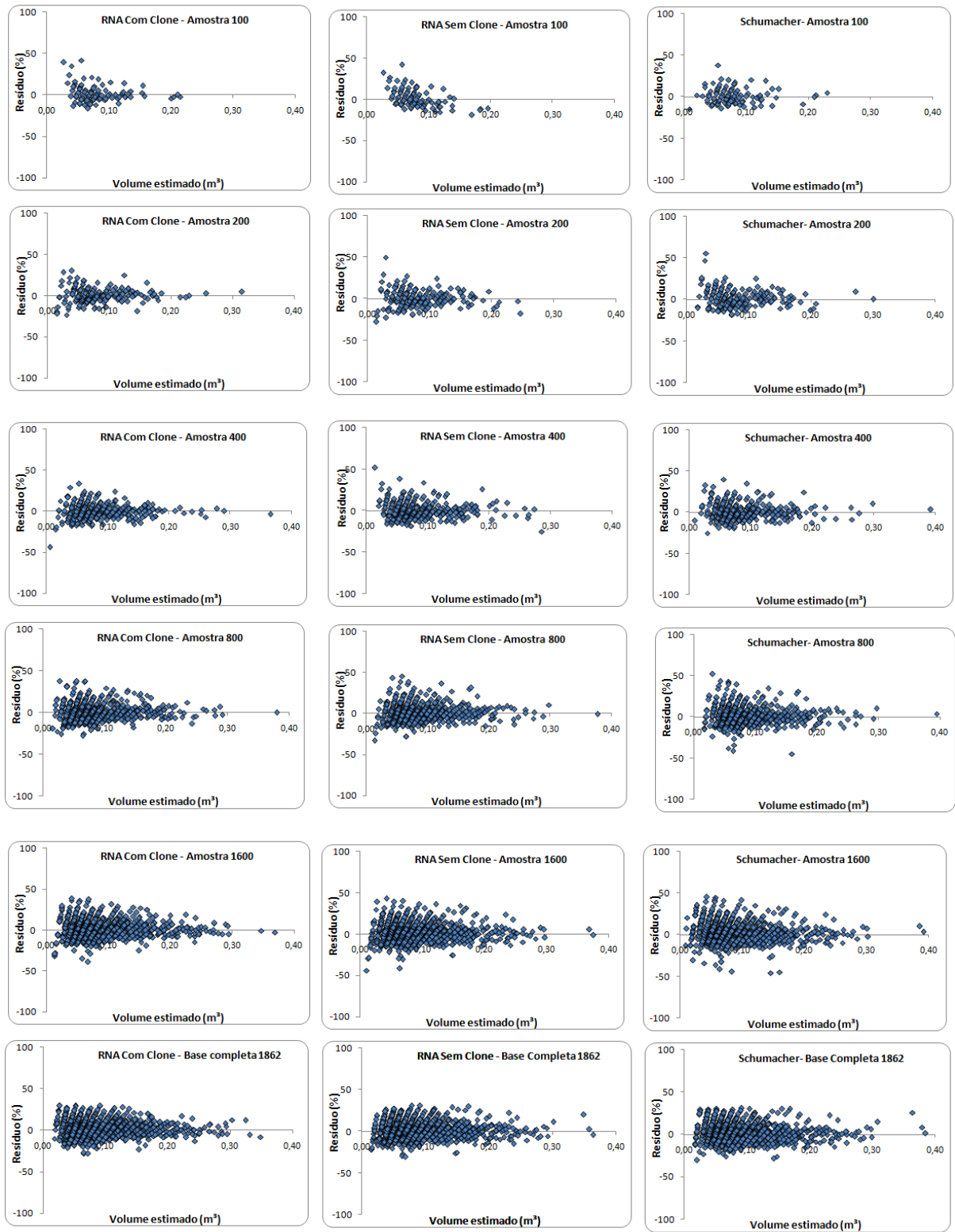


Figura 4.2: Gráficos de dispersão dos resíduos evidenciando o comportamento dos desvios encontrados entre volume observado e volume estimado por amostra.

Adicionando uma variável categórica (tratamento) a função a ser estimada os resultados das RNAs foram melhorados. De acordo com estatísticas apresentadas pelas RNA treinadas selecionadas, pode-se observar coeficientes de correlação ajustado com valores entre 98% e 99%, e RMSE entre 1% e 3%. Ao analisar visualmente os gráficos de dispersão dos resíduos, pode-se notar que apresentam uma melhora visível aos gráficos da RNA desconsiderando a variável categórica e também do modelo de regressão, com as estimativas pontuais próximas dos valores observados, não apresentando distribuições tendenciosas.

Os resultados foram satisfatórios nas estimativas das RNAs com a adição da variável categórica, porém as configurações na camada oculta aumentaram para 5 a 15 neurônios. De acordo com Bullinaria (2014), recomenda-se configurações de neurônios na camada oculta mais simples para aplicação das RNA.

Com o interesse de comparar a performance do modelo e das RNA em relação ao tamanho da base de dados, foram comparados os resultados das estatísticas dos modelos nos gráficos abaixo (Figura 4.3, 4.4 e 4.5).

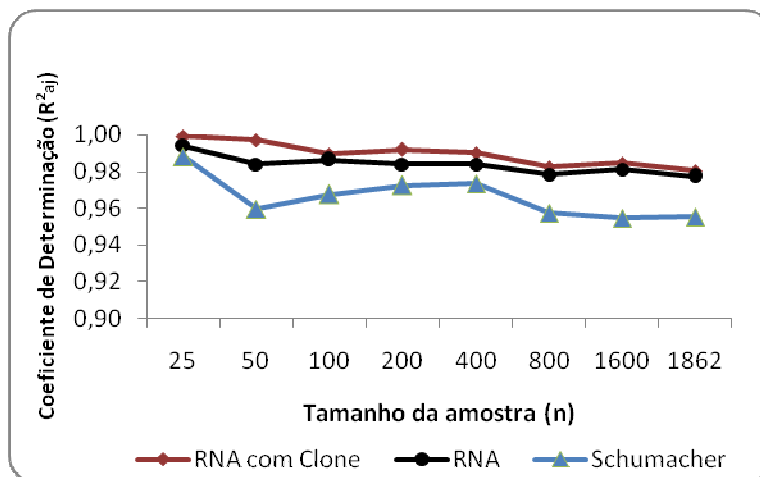


Figura 4.3: Gráficos de Coeficiente de determinação (R^2_{aj}) por amostra dos modelos analisados.

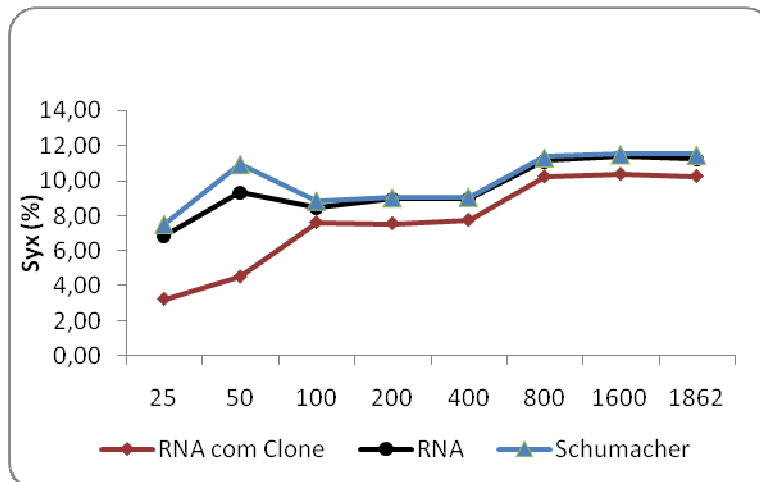


Figura 4.4: Gráficos de Erro padrão da estimativa (Syx%) por amostra dos modelos analisados.

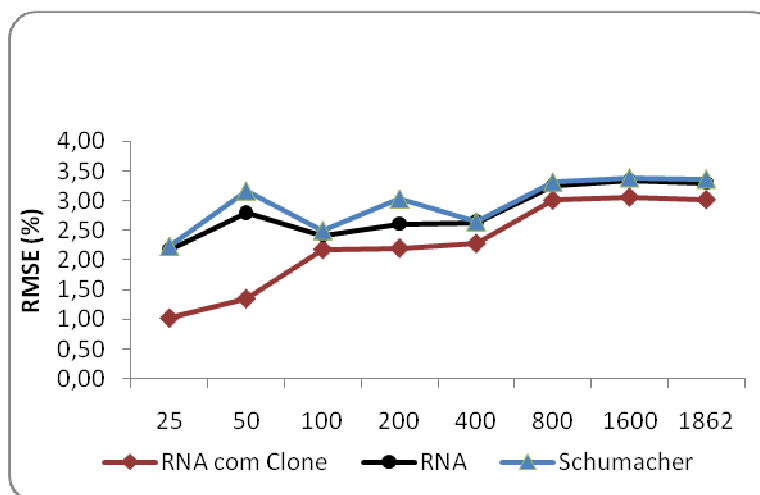


Figura 4.5: Gráficos da Raiz quadrada do erro médio (RMSE%) por amostra dos modelos analisados.

A avaliação da qualidade do ajuste dos modelos ao longo do tamanho da amostra utilizado para seu ajuste foi realizada por meio dos três indicadores contemplados nas figuras acima.

A análise do coeficiente de determinação ajustado, que indica basicamente quanto o modelo foi capaz de explicar os dados coletados na amostra, no presente estudo, favorece a RNA com a adição da variável categórica, tipo de clone. Os dois modelos de Redes Neurais Artificiais (com clone e sem clone) foram superiores ao modelo de Schumacher e-hall, o modelo que melhor se ajusta aos dados. A superioridade das Redes Neuras Artificiais com relação ao modelo de Schumacher

e-Hall e seu melhoramento com a inclusão do clone, foi confirmado com a análise do Erro padrão da estimativa (Syx%) e da Raiz quadrada do erro médio (RMSE) onde observamos uma leve superioridade na RNA com clone, em seguida a RNA sem clone e por ultimo o modelo de Schumacher-hall. Este resulta coincide com os de SILVA et al (2009). Para Hayklin (2001 citado por SILVA, 2009), esse resultado é de grande auxílio ao pesquisador. Com a disponibilidade dos modelos de RNAs com as três variáveis (DAP, Altura e Clone), não seria necessário o ajuste dos modelos volumétricos por clone já que o ajuste de RNA é representativo para todos os clones.

Enquanto á análise da influência do tamanho da amostra sobre os indicadores da performance de ajuste dos modelos, aqui avaliados, os resultados revelam que os indicadores não dependem apenas do tamanho da amostra mas também da variância, a tabela 3.3 que mostra que o erro amostral deixa evidente que quanto menor a amostra maior o erro consequentemente os resultados para essas bases podem não ser consistentes. Não se observa uma tendência de aumento do R^2_{aj} como esperado, nem uma tendência de diminuição da RMSE% e do Syx% com o aumento do tamanho da amostra. Os valores altos do R^2_{aj} nas pequenas amostras com seus correspondentes valores altos de Syx% e RMSE% induz o perigo de utilização de tamanho de amostras pequenas para obter modelos a serem usados na tomada de decisão. A utilização da base de dados maior permite abranger a maior variabilidade possível.

O uso de RNA para estimação de volume permite rapidez nos resultados e bom desempenho, além disso, a adição de variáveis categóricas ao modelo não são um problema ao treinamento da rede, isso pode ser uma característica que se destaca na escolha da metodologia para estimação de volume de *Eucalyptus*.

Capítulo 5: CONCLUSÃO

De acordo com os resultados expostos acima, comprova-se a capacidade em estimar volume de árvores de *Eucalyptus spp* utilizando o método de redes neurais artificiais (RNA) com eficiência. O ajuste dos modelos de Schumacher e-Hall e Spurr, confirmaram a adequacidade com resultados satisfatórios, o mesmo ocorreu com os resultados avaliados com o treinamento das redes neurais para estimar o volume de *eucalyptus spp* no Polo Gesseriso de Araripe – PE. O uso de RNA para estimação de volume permite rapidez nos resultados e bom desempenho. Concluindo assim, que a rede neural artificial é uma técnica eficiente para predição desta natureza.

A avaliação da qualidade do ajuste dos modelos por amostra demonstrou que a eficácia dos resultados independe do tamanho da base de dados, no presente estudo, pois foi observado certa instabilidade na evolução do desempenho dos resultados em relação ao tamanho da amostra. O modelo de regressão que melhor se ajustou aos dados foi o Shumacher e-Hall, este quando comparado com a metodologia de RNA por amostra apresenta resultados bons, porém a RNA ainda apresenta um acréscimo no resultado das estatísticas quando se adiciona uma variável categórica na rede, tipo de clone. A adição desse tipo de variável a rede neural artificial não é um problema a suposição da metodologia, a rede pode ser treinada normalmente e ganha um incremento aos resultados, essa característica pode ser um destaque na escolha da metodologia para estimar o volume de *Eucalyptus spp*.

REFERÊNCIAS BIBLIOGRÁFICAS

ADAMS, D.M., and EK, A.R., 1974, **Optimizing the management of uneven-aged forest stands**, Canadian Journal Forest Research 4: 274-287.

ALDANA, E. P., 2010. **MEDICIÓN FORESTAL**. Editorial Felix Varela, La Habana. 266 fl.

ARAUJO, S. M. S. **O polo gesseiro do Araripe: unidades geo-ambientais e impactos da mineração**. Universidade Estadual de Campinas – UNICAMP. Campinas, 2004.

ATECEL – Associação Técnica Científica Ernesto Luiz de Oliveira. **Diagnóstico energético do setor industrial do pólo gesseiro da meso região de Araripina-PE**. Campina Grande, 2006, 126p.

AVERY, T. E.; BURKHART, H. E. **Forest measurements**. New York: McGRAW-HILL, 1994. 432 p.

BALTAR, C. A. M.; BASTOS, F. F.; LUZ, A. B. **Diagnóstico do Pólo Gesseiro de Pernambuco (Brasil) com ênfase na produção de gipsita para fabricação de cimento**. 2003. Disponível em <http://www.minas.upm.es/catedra-anefa/Consultas/MAGALHAESCarlosIVIBERMAC.pdf>. Acesso em: 29.06.2015.

BINOTI, M. L. M. S.; BINOTI, D. H. B.; GLERIANI, J. M.; LEITE, H. G. **Ajuste do modelo de Schumacher e Hall e aplicação de redes neurais artificiais para estimar volume de árvores de eucalipto.** Revista *Árvore*, v. 33, n. 6, p. 1133-1139, 2009.

BINOTI, M. L. M. S. **Redes neurais artificiais para prognose da produção de povoamentos não desbastados de eucalipto.** 2010. 54 f. Dissertação (Mestrado em Ciência Florestal) – Universidade Federal de Viçosa, Viçosa, 2010.

BINOTI, M. L. M. S. **Emprego de redes neurais artificiais em mensuração e manejo florestal.** 2012. Tese (doutorado em Ciência Florestal) - Universidade Federal de Viçosa, Viçosa – MG, 2012, 136 p.

BINOTI, M. L. M. S.; BINOTI, D. H. B.; LEITE, H. G. **Aplicação de redes neurais artificiais para estimação da altura de povoamentos equiâneos de eucalipto.** Revista *Árvore* (vol.37 no.4). Viçosa, 2013.

BINOTI, D. H. B.; BINOTI, M. L. M. S.; LEITE, H. G. **Configuração de redes neurais artificiais para estimação do volume de árvores.** *Ciência da Madeira (Braz. J. Wood Sci.)*, v.5, n.1, p.58-67, 2014.

BOCANEGRA, C. W. R. **Procedimentos para tornar mais efetivo o uso de redes neurais artificiais em planejamento de transportes.** São Carlos - SP, 2002.

BULLINARIA, J. A. **Introduction to Neural Computation.** Notas de aula. Disponível em: <http://www.cs.bham.ac.uk/~jxb/inc.html>. 2014.

BUONGIOMO, J., and MICHIE, B.R., 1980, **A matrix model of uneven-aged forest management**, Forest Science 26: 609-625.

BURKHART, H. E. ; Tomé, M. **Modeling forest trees and stands**. New York: SPRINGER, 2012. 461 p.

BRAGA, A. DE P.; CARVALHO, A. P. DE L. F. DE; LUDERMIR, T. B. **Redes Neurais Artificiais: Teoria e Aplicações**. 2 ed. Rio de Janeiro: LTC, 2012.

BRASIL. Ministério de Minas e Energia. **Balço energético nacional 2013: Ano base 2012**, Empresa de pesquisa energética, Rio de Janeiro, 2013, 288 p.

BRUNER, H.D., and MOSER, J.W., Jr., 1973, **A Markov chain approach to the prediction of diameter distributions in uneven-aged forest stands**, Canadian Journal Forest Research 3: 409-417.

CAMPOS, J. C. C.; LEITE, H. G. **Mensuração florestal: perguntas e respostas**. 3.ed. Viçosa, MG: Universidade Federal de Viçosa, 2009. 543p.

CAMPELLO, F. C. B. **Análise do consumo específico de lenha nas indústrias gesseiras: a questão florestal e sua contribuição para o desenvolvimento sustentável da Região do Araripe – PE**. 2011. 61 f. Dissertação (Mestrado em Ciências Florestais). Universidade Federal Rural de Pernambuco, Recife. 2011.

CASTRO, R. V. O.; SOARES C. P. B.; LEITE, H. G.; SOUZA, A. L.; NOGUEIRA, G. S.; MARTINS, F. B. Individual Growth Model for Eucalyptus Stands in Brazil Using Artificial Neural Network. **ISRN Forestry**, v. 2013, p. 1- 12, 2013.

COCHRAN W. G.; **Técnicas de amostragem**. Rio de Janeiro: Fundo de Cultura. 1965.

COUTO, H. T. Z. e VETTORAZZO, S. C. **Seleção de equações de volume e peso seco comercial para Pinus taeda**. Departamento de Ciências Florestais da ESALQ-USP. Piracicaba, 1999.

DIAMANTOPOULOU, M. J. Tree-Bole Volume Estimation on Standing Pine Trees Using Cascade Correlation Artificial Neural Network Models. **Agricultural Engineering International: The CIGR EJournal**, VIII, 2006.

DRAPER, N. R.; SMITH, H. Applied regression analysis. 3^a. Edição. New York: John Wiley, 1998. 706p.

ENCINAS, J. I.; SILVA, G. F.; KISHI, I.T. **Comunicações técnicas florestais: Variáveis dendrométricas**. V.4, n.1. UnB. Brasília, 2002.

FAVRICHON, V., 1998, **Modeling the dynamics and species composition of a tropical mixed– species uneven–aged natural forest: Effects of alternative cutting regimes**, Forest Science 44: 113-124.

FRAZIER, J.R., 1978, **An investigation of the Markov model to predict individual tree DBH**, M.S. Thesis, Auburn Univ., Auburn, USA.

GADELHA, F., 2014. **Desempenho silvicultural e avaliação econômica de clones híbridos de eucaliptos plantados em diferentes regimes de manejo para fins energéticos.** Tese de Doutorado. Universidade Federal Rural de Pernambuco (UFRPE).

G. Shao & K.M. Reynolds (eds.): **Computer Applications in Sustainable Forest Management. Including Perspectives on Collaboration and Integration.** 2006 ISBN 1-4020-4305-8

GUAN, B.T., and GERTNER, G., 1995, **Modelling individual tree survival probability with a random optimization procedure: An artificial neural network approach,** AI Application 9: 39-52.

GUIMARÃES, D. P. **Variação de Shumacher-Hall para ajuste de equações volumétricas.** Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA. Planaltina, DF, 1986.

GÖRGENS, E. B.; LEITE, H. G.; SANTOS, H. 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. **Redes Neurais: princípios e prática.** Porto Alegre, 2001. Editora: Bookman.

KELLER, T., Guiot, J., and Tessier, L., 1997, **Climatic effect of atmospheric CO₂ doubling on radial tree growth in south eastern France,** Journal of Biogeography 24: 857-864.

LEAK, W.B., 1964, **An expression of diameter distribution for unbalanced, uneven-aged stands and forests**, Forest Science 10: 39-50.

LEARY, R.A., Holdaway, M.R., and Hahn, J.T., 1979, **Diameter growth allocation rule. in: A generalized forest growth projection system applied to the lake States region**, USDA Forest Service, Gen. Tech. Rep. NC-49. North Cent. Exp. Stat., Minnesota, pp. 22-26.

LEITE, H. G., ANDRADE, V. C. L. **Um método para condução de inventários florestais sem o uso de equações volumétricas**. R. Árvore, Viçosa-MG, v.26, n.3, p.321-328, 2002.

LEITE, F.S. **Estimativa do volume de madeira a partir do diâmetro da cepa em uma área explorada de Floresta amazônica de terra firme**. Universidade de Brasília – UnB, 2008.

LIMA, W. de. P. **Impacto Ambiental do Eucalipto**. 2ed. São Paulo: Editora da Universidade de São Paulo. 1996, 301p.

LIMA, R. B.; APARÍCIO, P. S.; FERREIRA, R. L. C.; SILVA, W. C.; GUEDES, M. C.; OLIVEIRA, C. P.; SILVA, D. A. S.; BATISTA, A. P. B. **Volumetria e classificação da capacidade produtiva para Mora paraensis (Ducke) no estuário amapaense**. Universidade Federal Rural de Pernambuco. Recife, 2014.

LISBOA, G. S. et al. Comparação do fator de empilhamento sob diferentes condições para madeira de Eucalyptus grandis. **Revista Ambiência**. Guarapuava, PR. v.5. n.1. p.81 - 91 jan./abr. 2009.

LUNDGREN, W. J. C. **Uso da geoestatística na estimativa volumétrica de clones de *Eucalyptus sp* no polo gesseiro do Araripe.** Universidade Federal Rural de Pernambuco – UFRPE. Recife, 2011.

MACHADO, S. A.; CONCEIÇÃO, M. B.; FIGUEIREDO, D. J. **Modelagem do volume individual para diferentes idades e regimes de desbaste em plantações de *Pinus oocarpa*.** Ciências Exatas e Naturais, Curitiba, v. 4, n. 2, p. 185-196, 2002.

MCTAGUE, J., and STANSFIELD, W.F., 1994, **Stand and tree dynamics of uneven-aged ponderosa pine,** Forest Science 40: 289-302.

MEUNIER, I. M. J. ; SILVA, J. A. A.; FERREIRA, R. L. C. **Inventário florestal programas de estudo.** Recife. 2001. 190 p.

MOSER, J.W., Jr., and HALL, O.F., 1969, **Deriving growth and yield functions for uneven-aged forest stands,** Forest Science 15: 183-188.

MOSER, J.W., Jr., 1972, **Dynamics of an uneven-aged forest stands,** Forest Science 18: 184-191.

MURPHY, P.A., and FARRAR, R.M., 1985, **Growth and Yield of Uneven-Aged Shortleaf Pine Stands in Interior Highlands,** USDA Forest Service Research Paper SO-218.

OLIVEIRA, R. J. **Uso de redes neurais artificiais na avaliação funcional de estradas florestais**. Universidade Federal de Viçosa, Viçosa, 2008.

PEREIRA, T. M. **Discriminação de populações com diferentes graus de similaridade por redes neurais artificiais**. Universidade Federal de Viçosa, Viçosa, 2009.

PIMENTEL–GOMES, F., HENRIQUE-GARCIA, C. **A determinação de equações volumétricas na Engenharia Florestal**. IPEF Série Técnica, Piracicaba, 9(26): 1-36, mar.1993. ISSN 100 – 8137.

Portal SEBRAE, o que são florestas energéticas. Disponível em <http://www.sebrae.com.br/sites/PortalSebrae/artigos/O-que-são-florestas-energéticas>. Acesso em 09 de julho de 2015.

Portal UNICAMP, redes neurais artificiais. Disponível em ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/theses/lnunes_mest/cap2.pdf. Acesso em 21 de abril de 2015.

Portal UEM. TATIBANA, C. Y. e KAETSU, D. K.. Introdução a redes neurais artificiais. Disponível em <http://www.din.uem.br/ia/neurais/>. Acesso em 21 de abril de 2015.

Portal UFSC. VIEIRA, R. C. e ROISENBERG, M. Redes neurais artificiais: um breve tutorial. Disponível em <http://user.das.ufsc.br/~gb/CIPEEL/TutorialRNA.pdf>. Acesso em 30 de junho de 2015.

RIEGELHAUPT, E. M.; PAREYN, F. G. C. **A questão energética e o manejo florestal da caatinga.** In: GARIGLIO, M. A.; SAMPAIO, E. V. S. B.; CESTARO, L. A.; KAGEYAMA, P. Y. Uso sustentável e conservação dos recursos florestais da caatinga. Brasília, DF: Serviço Florestal Brasileiro, 2010. p. 65-81.

ROLIM, S.G.; COUTO, H.Z.T. do; Jesus R.M. de; FRANÇA, J.T.; **Modelos volumétricos para a floresta Nacional do Tapirapé-Aquiri, Serra dos Carajas (PA).** Acta Amazonica, Vol. 36(1), p.107-114, 2006

SANTANA, O. A.; ENCINAS, J. I. **Equações volumétricas para uma plantação de Eucalyptus urophylla destinada a produção de lenha.** In: Simpósio latino-americano sobre manejo florestal, 3. Anais... Santa Maria: UFSM/PPGEF, 2004. p. 107-111.

SCHOENINGER, E. R.; KOEHLER, H. S.; WATZLAWICK, L. F.; OLIVEIRA FILHO, P. C. Uso de redes neurais artificiais como uma alternativa para mapeamento de biomassa e carbono orgânico no componente arbóreo de florestas naturais. **Revista Ambiência.** Guarapuava, PR v.4 n.3 p.529-549, 2008.

SCHUMACHER, F. X.; HALL, F. S. **Logarithmic expression of timber-tree volume.** Journal of Agricultural Research, v.47, n.9, p719-734, 1933.

SILVA, J. A. A. **Potencialidades de florestas energéticas de eucalyptus no pólo gesseiro do Araripe – Pernambuco.** Anais da Academia Pernambucana de Ciência Agrônômica. Recife. v. 5 e 6. p. 301 – 319. 2008a – 2009.

SILVA, J. A. A. **Análise de Equações Volumétricas para Construção de Tabelas de Volume Comercial para *Eucalyptus spp.*, Segundo a Espécie, Região e Método de Regeneração.** Universidade Federal de Viçosa. Minas Gerais, 1977.

SILVA, J. L. R. **Modelos volumétricos, fatores de forma e equação de afilamento para floresta de terra firme da região do rio aru município de Portel – Pará.** Universidade Federal Rural da Amazônia. Belém, 2007.

SILVA, M. L. M.; BINOTI, D. H. B.; GLERIANE, J. M.; LEITE, H. G. **Ajuste do modelo de Schumacher e Hall e aplicações de redes neurais artificiais para estimar volume de árvores de Eucalipto.** Revista *Árvore*, v. 33, n. 6, p. 1133-1139, 2009.

SINDUSGESSO (Sindicato das Indústrias do Gesso do Estado de Pernambuco). Disponível em <http://www.sindusgesso.org.br/sindusgesso-e-mppe-querem-revolucionar-sistema-produtivo-no-polo-gesseiro/>. Acesso em: 21.04.2015.

SOARES, C. P. B.; NETO, F. P.; SOUZA, A. L. **Dendrometria e inventário florestal.** Viçosa: Ed. UFV, 2012. 272p.

SOBRINHO, A.C.P. de L.; AMARAL, A.J.R.; DANTAS, J.O.C. Gipsita. Departamento Nacional de Produção Mineral. 2003, Disponível em: <http://www.dnpm-pe.gov.br/>, Acesso em: 01.03.2015.

SOLOMOM, D.S., Hosmer, R.A., and Hayslett, H.T., 1986, **A two-stage matrix model for predicting growth of forest stands in the Northeast**, Canadian Journal of Forest Research 16: 521-528.

STATSOFT. **STATISTICA** (data analysis software system), version 8. StatSoft, Inc., 2012.

SPURR, S.H. **Forest inventory**. New York: Ronald Press. 476p. 1952.

VALENÇA, M. **Fundamentos das redes neurais: Exemplos em Java**. 2ª. Edição, Revisado e ampliado. Olinda-PE: Livro Rápido, 2011. 386p.

VALENÇA, M. **Aplicando redes neurais: um guia completo**. Olinda-PE: Livro Rápido, 2005. 284p.

VOUKILA, Y., 1965. **Functions for variable density yield tables of pine based on temporary sample plots**. Commun. Inst. For. Fenn. 60:1-86.