

## CONFIGURAÇÃO DE REDES NEURAIS ARTIFICIAIS PARA ESTIMAÇÃO DO VOLUME DE ÁRVORES

Daniel Henrique Breda Binoti<sup>1</sup>, Mayra Luiza Marques da Silva Binoti<sup>2</sup>, Helio Garcia Leite<sup>3</sup>

**Resumo:** Objetivou-se nesse trabalho avaliar diferentes configurações do treinamento de Redes Neurais Artificiais (RNA) para a estimação do volume de árvores. Como método comparativo, realizou-se a quantificação volumétrica de maneira tradicional utilizando o modelo de Schumacher e Hall. Os dados utilizados neste estudo foram provenientes de cubagens de 2.307 árvores de povoamentos de eucalipto (*Eucalyptus grandis* x *Eucalyptus urophylla*). As variáveis utilizadas para treinamento das RNA foram: variáveis numéricas (quantitativas) consideradas foram: idade em meses, diâmetro com casca a 1,30 m de altura (*dap*) em cm, altura total (*Ht*) em m e volume da árvore (*V*) em m<sup>3</sup>. Enquanto que as variáveis categóricas (qualitativas) foram: projeto, espaçamento e clone. Para treinamento das RNA utilizou-se o sistema NeuroForest. A estimação do volume de árvores por meio de redes neurais artificiais pode ser feita com diferentes configurações, variando o número de neurônios na camada oculta, as funções de ativação, os algoritmos e os parâmetros de treinamento. Redes neurais simples com apenas um neurônio na camada oculta podem ser utilizadas com eficiência para estimação do volume de árvores.

**Palavras-chave:** NeuroForest; Schumacher e Hall; eucalipto.

## CONFIGURATION OF ARTIFICIAL NEURAL NETWORKS FOR ESTIMATING THE VOLUME OF TREES

**Abstract:** The aim of this study was to evaluate different configurations of training of Artificial Neural Network (ANN) to estimate the volume of trees. As the comparative method, a traditional volumetric quantification using the model of Schumacher and Hall was performed. The data used in this study were from the cubages of 2307 eucalyptus trees (*Eucalyptus grandis* x *Eucalyptus urophylla*). The variables used for the training of the ANNs were, as to numeric variables (quantitative), age in months; in-shell diameter at the height of 1.30 m in cm (dbh);

---

<sup>1</sup> Pós-doutorando em Ciência Florestal, Universidade Federal de Viçosa. <danielhbbinoti@gmail.com>.

<sup>2</sup> Doutora em Ciência Florestal, Docente, Universidade Federal dos Vales do Jequitinhonha e Mucuri. <mayrabinoti@gmail.com>.

<sup>3</sup> Doutor em Ciência Florestal, Docente, Universidade Federal de Viçosa. <hglete@gmail.com>.

total height in m (Ht); and tree volume in m<sup>3</sup> (V), whereas the categorical variables (qualitative) were design, spacing, and clone. The NeuroForest system was used to train the ANNs. The estimation of the volume of trees through artificial neural networks can be made with different configurations by varying the number of neurons in the hidden layer, the activation functions, algorithms and the training parameters. Simple neural networks with only one neuron in the hidden layer can be used effectively to estimate the quantity of trees.

**Keywords:** NeuroForest; Schumacher and Hall; eucalypt.

## 1 INTRODUÇÃO

As variáveis que interferem no volume das árvores, como idade, espaçamento, índice de local, solo, apresentam na maioria das vezes relações complexas e tendências muitas vezes não lineares. Consequentemente, dificultando o ajuste de modelos volumétricos com inclusão dessas variáveis. As redes neurais artificiais (RNA) apresentam-se como uma ferramenta apropriada para serem utilizadas nessas situações, onde a inclusão de variáveis categóricas não é factível (BINOTI et al., 2009).

Uma rede neural artificial (ou simplesmente uma rede neural ou rede) é um modelo computacional biologicamente inspirado, constituído por elementos de processamento simples (neurônios artificiais) que aplicam uma determinada função matemática aos dados (função de ativação) gerando uma única resposta, são dispostos em camadas e ligados entre si, sendo estas conexões, geralmente, associadas a coeficientes denominados de pesos. O ajuste destes pesos é realizado por um processo chamado treinamento ou aprendizado, sendo responsável pela extração das características dos dados e armazenamento de conhecimento das redes. Enquanto que a aplicação de uma rede consiste no processo de generalização, ou seja, utilização de uma rede treinada em dar resposta para dados inéditos (BRAGA et al., 2007).

Atualmente, muitas formas de RNA são propostas e usadas com variações, principalmente no número de neurônios por camada, tipo de função de ativação dos neurônios, número de camadas ocultas e tipo de conexão entre os neurônios. Outro fator que interfere significativamente no tempo de treinamento e na eficiência de uma rede são os algoritmos utilizados, bem como os parâmetros e coeficientes utilizados em seu treinamento.

O número de neurônios na camada oculta é definido principalmente de forma empírica pela experiência do usuário, embora existam algumas metodologias propostas na literatura, como por Hirose et al. (1991), Arai (1993) e Fujita (1998). Além disso, no processo de determinação do número de neurônios, deve-se considerar que um número excessivo de

neurônios pode acarretar a memorização dos dados de treinamento, processo conhecido como *overfitting*. Por outro lado, um pequeno número de neurônios na camada escondida pode não ser suficiente para a realização da tarefa desejada, fenômeno conhecido como *underfitting* (SILVA et al., 2010).

A definição do algoritmo de treinamento a ser utilizado, bem como de seus parâmetros influenciam principalmente na fuga de mínimos locais, no desempenho da tarefa desejada e no tempo de treinamento. Por exemplo, taxas de aprendizado muito baixas acarretam um aprendizado muito lento, enquanto que taxas de aprendizado muito altas geram oscilações no processo de treinamento, dificultando e até mesmo impedindo a convergência do algoritmo de treinamento *error backpropagation*, pois a busca no espaço de soluções “pula” o erro mínimo global (HAYKIN, 2001; BRAGA et al., 2007; BULLINARIA, 2010; SILVA et al., 2010).

Objetivou-se nesse trabalho avaliar diferentes configurações do treinamento de RNA para a estimação do volume de árvores. Como método comparativo, ajustou-se o modelo de Schumacher; Hall (1933).

## 2 MATERIAL E MÉTODOS

### 2.1 Dados

Os dados utilizados neste estudo foram provenientes de cubagens de 2.307 árvores de povoamentos de eucalipto (*Eucalyptus grandis* x *Eucalyptus urophylla*) com produtividade média, aos 7 anos, de cerca de 40 m<sup>3</sup>ha<sup>-1</sup>ano<sup>-1</sup>.

As variáveis numéricas (quantitativas) consideradas foram: idade em meses, diâmetro com casca a 1,30 m de altura, em cm (*dap*), altura total, em m (*Ht*) e volume da árvore, em m<sup>3</sup> (*V*) (Tabela 1). Enquanto que as variáveis categóricas (qualitativas) foram: projeto, espaçamento e clone. O número de classes das variáveis categóricas foram 23, 7 e 15, respectivamente. As árvores-amostra foram cubadas em seções de 1 m, sendo o volume total com casca (*V*) obtido pela aplicação da fórmula de Smalian (CAMPOS; LEITE, 2009).

**Tabela 1.** Análise descritiva das variáveis numéricas.

**Table 1.** Descriptive analysis of numerical variables.

Variável	Mínimo	Máximo	Média	Desvio padrão
Idade (anos)	2,71	11,00	6,50	± 2,35
<i>dap</i> (cm)	10,82	32,47	18,27	± 4,37
Altura total (m)	17,94	43,80	28,22	± 5,10
Volume (m <sup>3</sup> )	0,100	1,580	0,379	± 0,23

## 2.2 Modelo Volumétrico

Ao estratificar os dados, foram definidos 53 estratos com base nas variáveis: projeto, espaçamento, idade e clone. Para cada estrato, ajustou-se o modelo de Schumacher e Hall (CAMPOS; LEITE, 2009), conforme a Equação 1.

$$V = \beta_0 \text{dap}^{\beta_1} \text{Ht}^{\beta_2} \quad (\text{Equação 1})$$

em que,  $V$  = volume,  $\text{m}^3$ ;  $\text{dap}$  = diâmetro a 1,3 m de altura, cm;  $\text{Ht}$  = altura total da árvore, m;  $\beta_i$  = parâmetros; e  $\varepsilon$  = erro aleatório,  $\varepsilon \sim \text{NID}(0, \sigma^2)$ .

## 2.3 Redes Neurais Artificiais

As variáveis numéricas foram normalizadas linearmente no intervalo de 0 a 1, e as variáveis categóricas foram normalizadas utilizando-se a metodologia 1-de-N (GOLDSCHMIDT; PASSOS, 2005; HEATON, 2010).

A camada de entrada foi constituída de 48 neurônios sendo um neurônio para cada variável numérica e classe de variável categórica, ou seja: *dap* (1), altura (1), idade (1), projeto (23), espaçamento (7), clone (15). Como saída, utilizou-se o volume total da árvore. As redes foram constituídas de apenas uma camada oculta, sendo que o número de neurônios nessa camada variou de 1 a 48 (que corresponde ao número de neurônios na camada de entrada). Como funções de ativação da camada oculta e de saída, foram usadas as funções: linear, log (logarítmica), tangente hiperbólica, seno e sigmoide (HEATON, 2010).

Os algoritmos de treinamentos utilizados foram *error backpropagation*, *resilient propagation*, *Manhattan update rule*, *scaled conjugate gradient*, *levenberg marquardt*, *quick propagation*, *simulated annealing*, e algoritmos genéticos.

Para o algoritmo *backpropagation*, utilizou-se variações no valor da taxa de aprendizado de 0,001 a 1,000 (a cada 0,001), bem como variações no parâmetro *momentum* de 0,0000 a 1,000 (a cada 0,0005), totalizando 1000 valores de taxas de aprendizado e 2001 valores de *momentum*. Os valores de taxa de aprendizado e taxas de *momentum* foram combinados totalizando 2.001.000 combinações. Além dessas combinações, utilizou-se a estratégia de auto ajuste dos valores de taxa de aprendizado e *momentum*, processo que reduz os valores das taxas de aprendizado e *momentum* com a evolução do algoritmo resultando em 8.004.000 configurações do algoritmo *backpropagation*.

Para o algoritmo *resilient propagation*, considerou-se quatro variações básicas denominadas RPROP+, RPROP-, iRPROP+, iPROP (HEATON, 2010). No algoritmo

*Manhattan update rule*, que por definição necessita de pequenas taxas de treinamento, restringiu-se a utilização de somente um valor de taxa de aprendizado: 0,00001. Os algoritmos *scaled conjugate gradient*, *levenberg marquardt* apresentam somente uma configuração não dependendo de valores de parâmetros de treinamento.

O algoritmo de treinamento conhecido como *quick propagation* é baseado no método de Newton utilizado para obtenção de raízes de uma função. Esse algoritmo é tolerante a altas taxas de aprendizado, por isso avaliou-se taxas variando de 0,1 a 10,0 (a cada 0,1), totalizando 100 configurações desse algoritmo.

A metaheurística *Simulated annealing* é baseada no processo de recozimento de metais utilizado em processos metalúrgicos. O processo consiste em aquecer um metal e verificar como esse metal se resfria e congela numa estrutura cristalina de energia mínima. Como principais parâmetros desse algoritmo tem-se o número de ciclos por iteração e os valores de temperatura máxima e mínima. Os valores de números de ciclos avaliados foram variados de 100 a 1000 (a cada 100), com temperatura mínima de dois e máxima de dez, totalizando 10 configurações.

Outra metaheurística utilizada foi o Algoritmo Genético (AG). O AG inspira-se no processo de evolução natural das espécies e nos princípios de reprodução genética. O sistema possui como principais parâmetros configuráveis o tamanho da população, a probabilidade de mutação e a taxa de cruzamento. Testou-se valores de população variando de 50 a 1000 (a cada 50) totalizando 20 valores de população, valores de mutação variando de 0,1 a 0,5 (a cada 0,1) totalizando 5 valores de mutação, valores de cruzamento variando 0,05 a 0,70 (a cada 0,05) totalizando 14 valores. Foram testados um total 114 configurações de AG.

Foram avaliadas todas as possíveis combinações entre números de neurônios na camada oculta, tipo de funções de ativação na camada oculta e de saída, número de ciclos de treinamento e configurações dos algoritmos. Os pesos iniciais de todas as redes foram gerados aleatoriamente. Para cada combinação avaliada gerou-se 100 redes, sendo considerada uma simulação. A escolha da melhor configuração e o processo de treinamento das redes neurais artificiais foi realizado por meio de avaliação do erro quadrático médio e do desvio padrão dos erros quadráticos médios de cada simulação, sendo escolhidas as configurações mais simples e consistentes. Como critério de parada dos algoritmos de treinamento, utilizou-se: o número total de ciclos (com os valores variando de 500 a 3000 (a cada 100), totalizando 26 valores) ou o erro quadrático médio inferior a 1%. Portanto, finalizou-se o treinamento quando um dos critérios foi atingido.

O treinamento e a avaliação das redes foram feitos no sistema NeuroForest – Volumetric, desenvolvido por Helio G. Leite, Daniel H. B. Binoti e Mayra Luiza M. da S. Binoti, disponível

gratuitamente na página <http://neuroforest.ucoz.com/>. A versão disponível no site foi alterada pelos próprios autores a fim de rodar scripts executando as configurações testadas nesse trabalho.

## 2.4 Avaliação dos Modelos

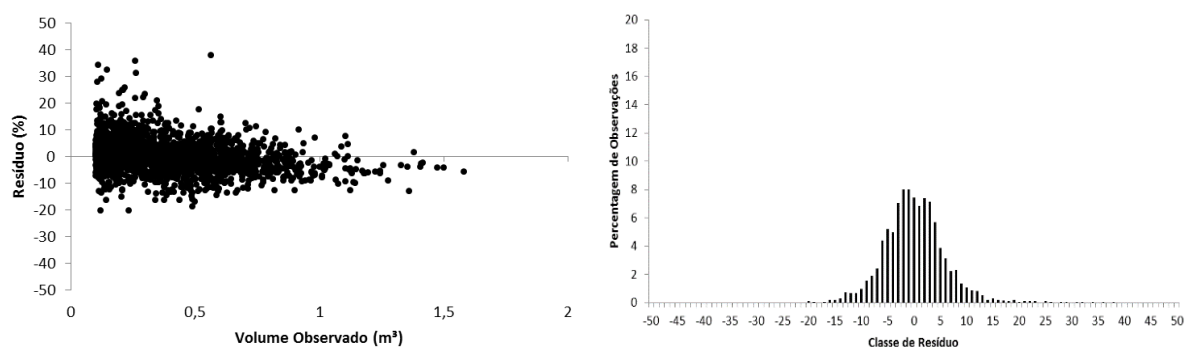
As estimativas das RNA e dos modelos de Schumacher e Hall foram avaliadas com base nos coeficientes de correlação com os valores observados e a raiz quadrada do erro médio (RMSE) (Equação 2), bem como na análise gráfica dos resíduos (erros percentuais) e histograma de erros percentuais.

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

Em que:  $Y$  = saída observada nos inventários;  $\hat{Y}$  = saída estimada;  $\bar{Y}$  = média da saída observada;  $n$  = número total de dados.

## 3 RESULTADOS E DISCUSSÃO

Foram estimados 53 equações volumétricas a partir do modelo de Schumacher e Hall. O coeficiente de correlação entre os valores observados e estimados pelas equações foi de 0,993 e o valor de RMSE estimado foi de 7,28%. As estimativas apresentaram-se sem tendenciosidade e livre de viés. A dispersão dos resíduos em função dos valores observados e o histograma de resíduos são apresentados na Figura 1.



**Figura 1.** Gráfico de resíduos e histograma de erros obtidos pelas estimativas realizadas pelo modelo de Schumacher e Hall.

**Figure 1.** Leavings graph and error histogram obtained through the estimates made by the model of Schumacher and Hall.

Foram avaliadas 1200 configurações diferentes de RNA referentes ao número de neurônios na camada oculta e combinações de funções de ativação. Como critério de parada, utilizou-se o número de ciclos totalizando 26 critérios. Essas configurações foram combinadas com as simulações de cada algoritmo *backpropagation* (8.004.000), *resilient propagation* (4), *Manhattan update rule* (1), *scaled conjugate gradiente* (1), *levenberg marquardt* (1), *quick propagation* (100), *simulated annealing* (10), e algoritmos genéticos (114). No total foram testadas  $2,49 \cdot 10^{11}$  configurações, sendo treinadas  $2,49 \cdot 10^{13}$  RNA.

Em função da enorme quantidade de redes treinadas, foram avaliadas apenas as configurações que apresentaram valores satisfatórios de erro quadrático médio, ou seja, valores menores que 1%.

Obtiveram-se resultados satisfatórios com configurações de apenas um neurônio na camada oculta, mostrando que redes de estruturas complexas não são necessárias para a estimação do volume de árvores. Provavelmente, isto se justifica pelo fato do número de neurônios na camada oculta ser responsável principalmente pela extração de características de não linearidade dos dados. Neste caso, um único neurônio foi capaz de extrair a relação entre as variáveis consideradas e o volume.

Recomenda-se que nas aplicações de RNA sejam escolhidas configurações mais simples, com o menor número de neurônios possível na camada oculta, a fim de evitar o *overfitting* e facilitar o processo de busca e otimização da configuração para uma determinada tarefa (BULLINARIA, 2010).

Não houve diferença entre as funções de ativação na camada oculta avaliadas. Para as funções de ativação na camada de saída, a função linear, log seno e tangente hiperbólica apresentaram valores negativos para algumas estimativas volumétricas devido à normalização das variáveis numéricas ter sido feita entre valores de 0 a 1, enquanto que estas funções resultam em valores entre -1 e 1.

Não observaram-se redes satisfatórias com os algoritmos *scaled conjugate gradiente*, *Manhattan update rule*, *levenberg marquardt*, *simulated annealing* e algoritmos genéticos, para os dados desse estudo em nenhuma configuração testada.

Para taxas de aprendizado com valores acima de 0,1, o algoritmo *backpropagation* mostrou-se instável, não convergindo. Valores baixos de taxa de aprendizagem (menores que 0,01) e do *momentum* (menores que 0,01) apresentaram resultados satisfatórios, não influenciando, de forma perceptível, o tempo de treinamento do algoritmo, apesar de exigir um maior número de ciclos para a convergência.

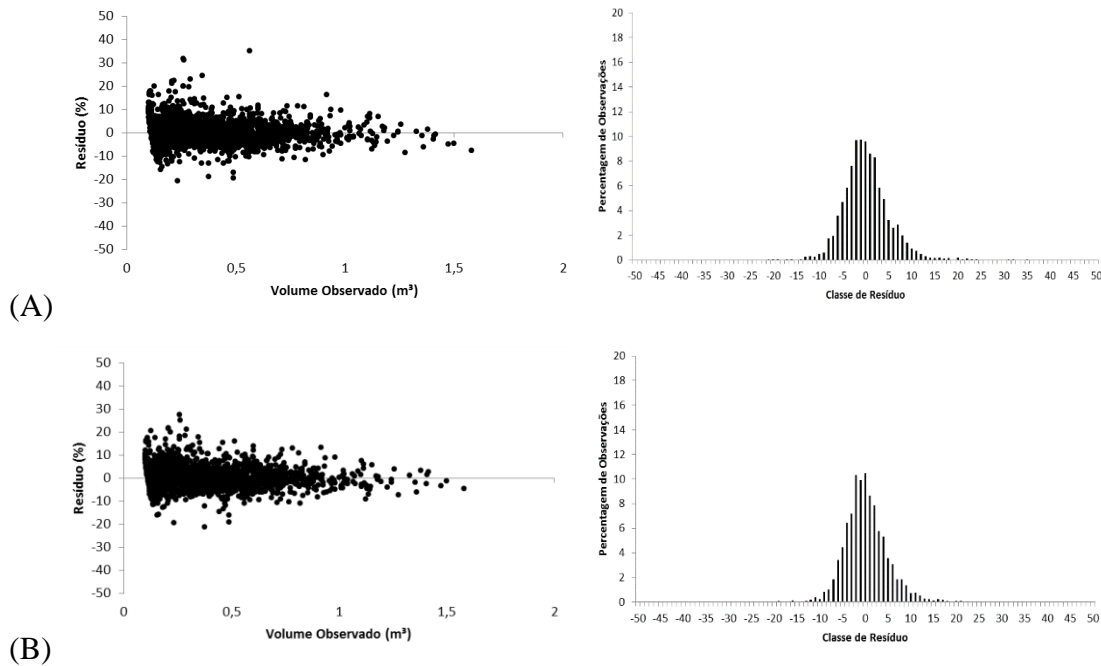
Não objetivou-se avaliar o tempo de processamento, visto que o sistema NeuroForest - Volumetric ser extremamente eficiente no processo de treinamento (cada simulação gastou aproximadamente 10 segundos), o que torna o tempo um fator irrelevante.

Para o algoritmo *resilient propagation* somente as configurações com mais do que três neurônios na camada oculta apresentaram resultados satisfatórios. Não verificou-se diferença entre as quatro variações básicas RPROP+, RPROP-, iRPROP+, iPROP.

O algoritmo *quick propagation* apresentou resultados satisfatórios somente para configurações que possuíam mais do que três neurônios na camada oculta e taxa de aprendizado inferior a 1,0. Taxas superiores a 1,0 causaram instabilidade no algoritmo, não apresentando convergência.

Selecionou-se duas configurações que apresentaram resultados satisfatórios para demonstração gráfica (Figura 2): A) um neurônio na camada oculta, com função de ativação tangente hiperbólica na camada oculta e sigmoide na camada de saída, treinamento utilizando o algoritmo *backpropagation* com taxa de aprendizado igual a 0,001 e *momentum* nulo; e B) 12 neurônios na camada oculta, com função de ativação tangente hiperbólica na camada oculta e sigmoide na camada de saída, treinamento utilizando o algoritmo *resilient propagation* (RPROP+).





**Figura 2.** Gráfico de resíduos e histograma de erros obtidos pelas estimativas realizadas pelas RNA treinadas. A) um neurônio na camada oculta, com função de ativação tangente hiperbólica na camada oculta e sigmoide na camada de saída, treinamento utilizando o algoritmo backpropagation com taxa de aprendizado igual a 0,001 e momentum nulo; e B) 12 neurônios na camada oculta, com função de ativação tangente hiperbólica na camada oculta e sigmoide na camada de saída, treinamento utilizando o algoritmo resilient propagation (RPROP+).

**Figure 2.** Leavings graph and error histogram obtained through estimates made by RNA trained. A) a neuron in the hidden layer with hyperbolic tangent activation function in the hidden layer and sigmoid in the output layer, training using the backpropagation algorithm with learning rate equal to 0.001 and zero momentum, and B) 12 neurons in the hidden layer, with hyperbolic tangent activation function in the hidden layer and sigmoid in the output layer, training using the resilient propagation algorithm (RPROP+).

Observou-se pouca diferença nas estimativas com relação ao modelo de Schumacher e Hall. Deve-se salientar que as estimativas foram feitas utilizando 53 equações volumétricas (estratificação) contra somente uma única RNA para todos os dados. Fato que favorece a utilização das RNA bem como a aplicação operacional do método, por ser possível reduzir muito o tempo gasto com ajustes e avaliações de modelos volumétricos por estrato. Além disso, não observou-se diferenças perceptíveis entre a configuração A (pouca complexidade) e B (maior complexidade), fato que justifica a utilização de RNA com baixa complexidade, ou seja, poucos neurônios na camada oculta, o que pode levar a rede a memorizar os dados de treinamento (*overfitting*), ao invés de extrair as características gerais dos dados que permitirão a generalização.

## 4 CONCLUSÕES

A estimação do volume de árvores por meio de redes neurais artificiais pode ser feita com diferentes configurações, variando o número de neurônios na camada oculta, as funções de ativação, os algoritmos e os parâmetros de treinamento.

Redes neurais simples com apenas um neurônio na camada oculta podem ser utilizadas com eficiência para estimação do volume de árvores.

## 5 REFERÊNCIAS

ARAI, M. Bounds on the number of hidden units in binary-valued three-layer neural networks. **Neural Networks**, v.6, n.6, p. 855-860, 1993.

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.

BRAGA, A. de P.; CARVALHO, A. P. de L. F. de; LUDEMIR, T. B. **Redes Neurais Artificiais: Teoria e Aplicações**. 2 ed. Rio de Janeiro: LTC, 2007. 260p.

BULLINARIA, J. A. Introduction to Neural Computation. **Notas de aula**. 2008. Disponível em: <<http://www.cs.bham.ac.uk/~jxb/inc.html>>. Acesso em: 29 abr. 2012.

CAMPOS, J. C. C.; LEITE, H. G. **Mensuração florestal: perguntas e respostas**. 3ed. Viçosa: UFV, 2009. 548p.

FUJITA, O. Statistical estimation of the number of hidden units for feedforward neural networks. **Neural Networks**, v.11, n.5, p. 851-859, 1998.

GOLDSCHMIDT, R.; PASSOS, E. **Data mining: um guia prático**. Rio de Janeiro: Elsevier, 2005. 261p.

HAYKIN, S. **Redes neurais: princípios e prática**. 2ed. Porto Alegre: Bookman, 2001.

HEATON, J. **Programming Neural Networks with Encog3 in Java**. 2ed. Heaton Research, Incorporated, 2011. 240p.

HIROSE, Y.; YAMASHITA, K.; HIJIYA, S. Back-propagation algorithm which varies the number of hidden units. **Neural Networks**, v.4, n.1, p. 61-66, 1991.

SCHUMACHER, F. X.; HALL, F. S. Logarithmic expression of timber-tree volume. **Journal of Agricultural Research**, v.47, n.9, p. 719-734, 1933.

SILVA, I. N.; SPATTI, H. D.; FLAUZINO, R. A. **Redes Neurais Artificiais: para engenharia e ciências aplicadas**. São Paulo: Artliber, 2010. 399p.