

**NÚMERO DE INDIVÍDUOS DO ALGORITMO NEAT PARA ESTIMATIVA DO VOLUME
COMERCIAL DA CAATINGA ARBÓREA**

**NUMBER OF INDIVIDUALS OF NEAT ALGORITHM FOR ESTIMATION OF THE COMMERCIAL
VOLUME OF ARBOREOUS CAATINGA**

Loic Aymard M'foumyt-M'foumyt¹ Poliana Camila Alves de Oliveira Adriana Leandra de Assis
Carlos Alberto Araújo Júnior Christian Dias Cabacinha

RESUMO

A Caatinga Arbórea é uma tipologia vegetacional do bioma Caatinga e o conhecimento do seu estoque volumétrico é fundamental na elaboração de planos de manejo. O objetivo deste trabalho foi definir o número mínimo de indivíduos para obter uma rede satisfatória para estimar o volume comercial da Caatinga arbórea. Dados de volume comercial com casca de 300 indivíduos foram obtidos para compor a base de dados para o processamento com as Redes Neurais Artificiais usando o algoritmo genético NEAT. O processamento foi composto por duas fases, na qual a primeira continha redes com 10 indivíduos até 100 em intervalos de 10 indivíduos e a segunda, redes de 100 indivíduos até 1000 em intervalos de 100 indivíduos. Realizou-se uma análise das estatísticas correlação e raiz quadrada do erro médio para cada fase e análise de variância de um critério e teste de médias de Scott-Knott para comparar as diferentes configurações que geraram resultados satisfatórios. As redes com número de indivíduos inferior a 300 foram descartadas e a partir de 400 indivíduos, com densidade de conexão inicial de 0,1, apresentaram as maiores correlações e as menores raízes quadradas do erro médio no treinamento das redes, sendo este o número mínimo recomendado de tamanho populacional para configuração de redes treinadas com o algoritmo NEAT para modelar o volume comercial em Caatinga arbórea. **Palavras-chave:** algoritmo genético; produção volumétrica; tamanho populacional.

ABSTRACT

The Arboreous Caatinga is a vegetation typology of the Caatinga biome and the knowledge of its volumetric stock is fundamental in the elaboration of management plans. The objective of this work was to define the minimum number of individuals to obtain a satisfactory network to estimate the commercial volume of the Arboreous Caatinga. Commercial volume data with bark of 300 individuals were obtained to compose the data base for processing with Artificial Neural Networks using the genetic algorithm NEAT. The processing consisted of two phases, in which the first contained networks with 10 individuals up to 100 in intervals of 10 individuals and the second, networks of 100 individuals up to 1000 in intervals of 100 individuals. An analysis of the correlation and root square of mean error for each phase and analysis of variance of a one way and Scott-Knott averages test were performed to compare the different configurations that generated satisfactory results. Networks with a number of individuals of less than 300 were discarded and from 400 individuals with initial connection density of 0.1, presented the highest correlations and the lowest roots square of the mean error in the training of the networks, this being the minimum number recommended population size for the configuration of trained networks with the NEAT algorithm to model the commercial volume in Arboreous Caatinga.

Keywords: genetic algorithm; volumetric production; population size.

INTRODUÇÃO

A Caatinga é uma vegetação que ocupa uma área de 844.453 km², 10% do território brasileiro, distribuída nos estados do Ceará, Alagoas, Minas Gerais, Bahia, Maranhão, Paraíba, Pernambuco, Piauí, Rio Grande do Norte e Sergipe. Subdivisão deste bioma, a Caatinga Arbórea é composta por diversas espécies de interesse medicinal, alimentício (humano e animal) e econômico (RIZZINI, 1997).

Diante da importância e uso da Caatinga Arbórea, faz-se necessário realizar estudos para obtenção do estoque volumétrico presente, realizando a supressão dos indivíduos para valoração e quantificação do que será explorado (SANQUETTA et al., 2014). Comumente, usa-se equações volumétricas ajustadas para estimar o volume da vegetação, porém os modelos matemáticos encontrados na literatura nem sempre conseguem explicar

¹ Estudante de Engenharia Florestal, Graduando, Instituto de Ciências Agrárias, Universidade Federal de Minas Gerais, Avenida Universitária, 1000, CEP 39404-547, Montes Claros (MG), Brasil. E-mail: loicmfoumyt@gmail.com

o comportamento dessa variável de forma satisfatória devido à alta variabilidade das mesmas (REZENDE et al., 2006). Como saída, pode-se fazer uso de outras ferramentas como as Redes Neurais Artificiais (RNA).

As RNA são sistemas complexos que, através da entrada de dados, detectam padrões e funcionam similarmente a um cérebro humano, capazes de fazer interpolações e extrapolações de forma linear e geram cenários distintos em diferentes arquiteturas que são algoritmos ligadas aos neurônios (HAYKIN, 2001; BRAGA et al., 2014). Essa metodologia tem se destacado pela sua capacidade de produzir excelentes resultados e pela tendência em gerar erros mínimos.

Um dos algoritmos que pode ser usado para o treinamento de redes neurais é o Neuroevolution of Augmenting Topologies (NEAT). O NEAT, é uma técnica que a partir de um algoritmo evolutivo, ajusta simultaneamente os pesos sinápticos e a topologia de uma RNA. A aplicação desse algoritmo facilita a vida do usuário de redes, já que elimina a etapa de busca da melhor arquitetura que geralmente é feita por tentativa e erro.

O processo de otimização utilizado pelo NEAT é iniciado a partir de um conjunto de configurações (população inicial), que é alterada a cada iteração, tendo como resultado um novo conjunto de configurações (nova população). Esse processo da geração de uma nova população a partir da população corrente é realizado a partir de operadores de seleção, crossover e mutação. Como a cada iteração são encontradas melhores configurações, é esperado que nesse processo de busca, a solução ou a configuração ótima global, seja encontrada (RODRIGUES, 2001; BUCKLAND, 2002; STANLEY; MIIKKULAIEN, 2004).

Contudo durante este processo, dependendo do número de indivíduos definido na fase de configuração inicial da rede neural, isso pode demandar um alto tempo de processamento para encontrar a melhor solução.

Desse modo, o presente trabalho teve por objetivo definir o número mínimo de indivíduos para obter redes satisfatórias a partir do treinamento de Redes Neurais Artificiais com algoritmo genético NEAT para a estimativa do volume comercial da Caatinga arbórea.

MATERIAL E MÉTODO

Caracterização da área

Os dados utilizados neste trabalho foram coletados na Floresta Nacional Contendas do Sincorá, unidade de conservação de uso sustentável localizada em Contendas do Sincorá-BA nas coordenadas 13°55'14,51"S e 41°6'54,79" com uma área de 11.215,78 hectares. O clima é definido como semiárido quente do tipo BSwh' de Köppen, com chuvas acumuladas entre novembro e janeiro, precipitação entre 500 a 1000 mm anuais, temperatura de 21 a 28 °C e umidade relativa de 60 a 70%. O solo é caracterizado por um solo Argiloso e de Latossolo Vermelho-Amarelo.

A vegetação da unidade de conservação é dividida em três tipologias: Caatinga arbórea arbustiva, Caatinga arbustiva fechada e complexo herbáceo-arbustivo, com predominância da Caatinga arbórea arbustiva e indivíduos entre 3,5 a 7,0 m de altura.

Cubagem rigorosa

Para obtenção do volume comercial com casca (VCc/m^3) foram cubadas rigorosamente 300 árvores distribuídas em classes diamétricas com amplitude de 5 cm, considerando todas as espécies necessárias para coleta dos dados. O método de cubagem usado foi de Smalian, com a tomada do diâmetro ao longo do tronco, nas alturas de 0,10; 0,30; 0,50; 0,70; 1,00; 1,30; 2,00 m e depois de um em um metro até o diâmetro comercial de 3 cm.

Processamento de dados

Utilizando o algoritmo genético Neuroevolution of Augmenting Topologies (NEAT) 30 redes neurais foram treinadas. Foram separados 70% da base de dados para treinamento que foi dividido em duas fases. Na primeira fase, as redes tiveram uma variação no número de indivíduos, 10 até 100, com intervalos de 10 indivíduos (Etapa 1). Já na segunda fase, o número de indivíduos variou de 100 até 1000, em intervalos de 100 indivíduos (Etapa 2). Utilizou-se como densidade de conexão inicial 0,1 (ARAÚJO, 2015) e como critério de parada, fixou-se o número de épocas em 3000 ou erro médio quadrático igual a 0,0001, ou seja, o treinamento de cada RNA era finalizado quando um dos dois critérios fosse atingido (BINOTI et al., 2014).

Definição do número mínimo de indivíduos

Para a escolha da melhor rede dentro de cada intervalo e definição do número ideal de indivíduos mínimo considerou-se as seguintes estatísticas: correlação entre os valores estimados e observados do volume ($r_{Y\hat{Y}}$) e raiz quadrada do erro médio em porcentagem (RQME%) de cada rede.

$$\text{Para a correlação: } r_{Y\hat{Y}} = \frac{\text{cov}(Y, \hat{Y})}{\sqrt{s^2(Y)s^2(\hat{Y})}}$$

$$\text{Para a raiz quadrado médio em porcentagem: } \text{RQME\%} = \frac{100}{\bar{Y}} \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}}$$

Em que: cov = covariância; s^2 = variância; n = número de observações; Y = volumes observados; \bar{Y} = volume observado médio; \hat{Y} = volumes estimados pela RNA.

Para avaliar de forma geral o comportamento para o treinamento em cada fase de processamento, gráficos do tipo boxplot para os resultados de correlação e para a RQME% foram gerados. Para comparar as RNA treinadas para cada fase do processamento, foi realizada para a correlação e RQME% uma análise de variância de um critério e teste de médias de Scott-Knott para o nível de significância de 5%. O treinamento das redes foi realizado no sistema computacional NeuroForest 4.0 (BINOTI, 2012) e as análises estatísticas no software R (R CORE TEAM, 2017).

RESULTADOS E DISCUSSÃO

Foram construídos gráficos boxplot, conhecido como um método alternativo ao histograma para representar os dados, fornecendo informações sobre as seguintes características do conjunto de dados: localização, medidas discrepantes (outliers), dispersão e assimetria. Dessa forma, foram analisados os valores obtidos durante o treinamento das RNA para a Correlação e para a RQME%.

Os gráficos boxplot para dos resultados de correlação e RQME% das redes treinadas em cada fase do processamento dos dados estão apresentados na Figura 1.

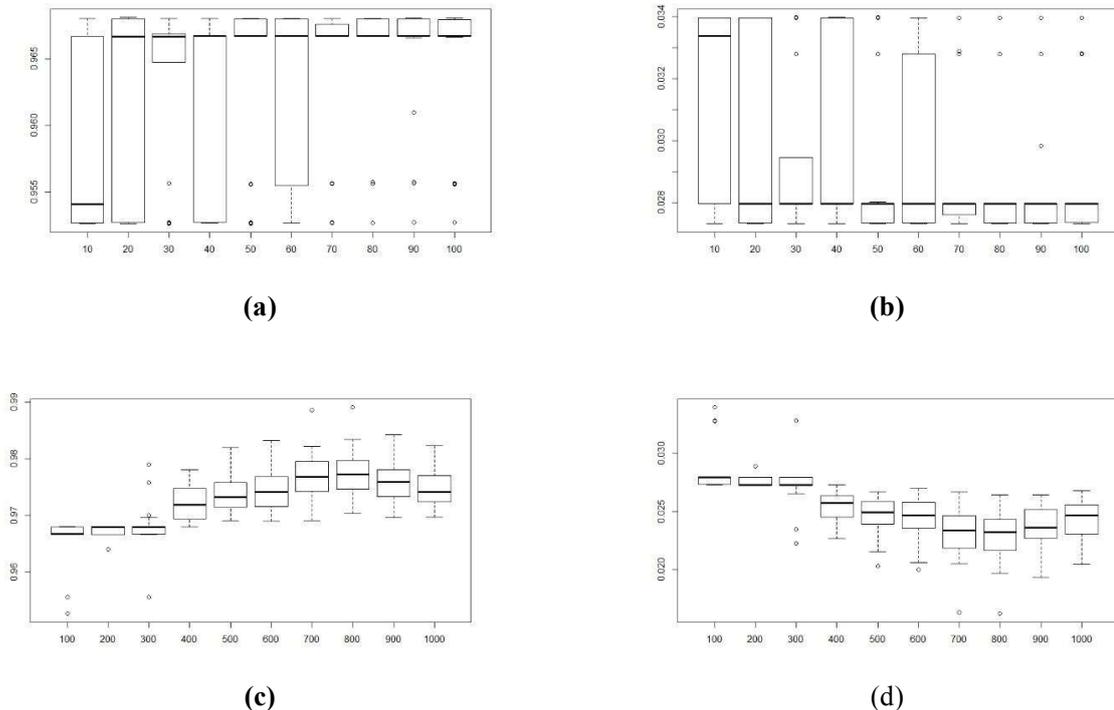


FIGURA 1: Boxplot para a correlação (a) e RQME% (b) para a fase 1 de processamento dos dados e para a correlação (c) e RQME% (d) para a fase 2 de processamento dos dados.

FIGURE 1: Boxplot for the correlation (a) and RQME% (b) for the data processing phase 1 and for the correlation (c) and RQME% (d) for the data processing phase 2.

Na fase 1 (Figura 1a e 1b) de processamento, observou-se para as duas estatísticas analisadas, uma grande variação nos resultados para as redes treinadas. Em geral as redes geraram valores de correlação e RQME% piores quando comparadas às redes da fase 2 de processamento. A mediana para a correlação aumenta

a partir da rede com 20 indivíduos e praticamente não se alterou até a rede com 100 indivíduos. Já para o RQME% a rede 10 apresentou a maior mediana e também a partir de 20 indivíduos a mediana diminuiu e praticamente não se alterou até a rede com 100 indivíduos. Para a dispersão verificou-se para as redes com 10, 20, 40 e 60 indivíduos, grande dispersão e uma forte assimetria nos resultados de correlação e RQME%. Para a rede 10, a assimetria é positiva, para as demais a assimetria é negativa. Embora as demais configurações tenham apresentado uma dispersão menor, verificou-se ainda assimetria, sendo para a rede 30 negativa e as demais configurações uma assimetria positiva, sendo que para as redes 50, 70, 80, 90 e 100 a mediana foi igual ao primeiro quartil, resultado que pode estar associado ao grande número de valores discrepantes identificados na análise do boxplot.

Diante destes resultados verificados as redes treinadas durante a fase 1 de processamento foram consideradas insatisfatórias e descartadas para a estimativa de volume comercial para Caatinga Arbórea.

Na fase 2 (Figura 1c e 1d) de processamento, observou-se para as duas estatísticas analisadas, resultados mais satisfatórios, exceto para as configurações considerando 100, 200 e 300 indivíduos que apresentaram ainda comportamento semelhante ao descrito para as redes da fase 1 de processamento. Estas redes geram valores médios de RQME% de 49,97, 48,05 e 47,61%, respectivamente. Esse comportamento era esperado, embora as redes com o maior número de indivíduos o custo computacional na etapa de treinamento seja maior, populações maiores toleram maior número de espécies potenciais, ou seja, maior quantidade de RNA testadas com alta diversidade topológica, aumentando as chances de obter bons resultados (STANLEY; MIIKKULAIEN, 2002).

A partir de 400 indivíduos a mediana da correlação aumenta até 800 e volta a diminuir para as redes 900 e 1000. Para o RQME%, como esperado, verificou-se o contrário, a partir de 400 os valores das mediana decrescem até o mínimo que ocorreu na rede com 800 indivíduos, voltando a subir para as redes 900 e 1000.

Realizou-se a análise variância (ANOVA) para comparar as estatísticas correlação e RQME%, somente para as configurações com 400, 500, 600, 700, 800, 900 e 1000 indivíduos, que foram consideradas as situações estudadas mais satisfatórias com RQME% inferiores a 45%.

A ANOVA para a correlação mostrou que haviam diferenças significativas (F: 43,919; p-value <0,001; CV: 0,37%) entre as diferentes configurações. O mesmo ocorreu para a RQME% (F: 38,919; p-value < 0,001; CV: 6,93%).

A Tabela 1 mostra o resultado do teste de médias Scott-Knott para as médias de correlação e de RQME% para as redes treinadas com 400 até 1000 indivíduos.

TABELA 1: Resultado do teste de médias de Scott-Knott para a correlação e RQME% para as melhores redes treinadas.

TABLE 1: Test result of Scott-Knott averages for correlation and RQME% for best trained networks.

Nº de indivíduos	Correlação*	RQME%*
400	0,9722 c	44,23 c
500	0,9737 b	43,00 b
600	0,9745 b	42,49 c
700	0,9768 a	40,57 a
800	0,9774 a	39,87 a
900	0,9760 a	41,09 a
1000	0,9748 b	41,78 b

(*) Médias seguidas pela mesma letra, na coluna, não diferem entre si pelo teste de Scott-Knott ao nível de 5% de significância.

Os valores médios de correlação para todas as redes foram muito próximos, sendo que o maior valor foi observado para a rede com 800 indivíduos que não diferiu estatisticamente das redes com 700 e 900 indivíduos. A rede com a menor média de correlação foi observada para rede de 400 indivíduos. Para o RQME% a rede com a menor média foi observada para a configuração com 800 indivíduos que também não foi estatisticamente diferente das redes com 700 e 900 indivíduos e a rede com a menor média de RQME% foi a com 400 indivíduos.

Embora do ponto de vista estatístico as redes tenham apresentado diferenças para as estatísticas correlação e RQME%, para a correlação a diferença entre a maior (800 indivíduos) e a menor correlação (400 indivíduos) é de 0,0052 e o menor (800 indivíduos) e o maior RQME% (400 indivíduos) tem uma diferença de 4,36%. Na prática tais diferenças são irrelevantes. O objetivo deste trabalho era analisar em que configuração de número de indivíduos o algoritmo NEAT gerava uma solução ótima para estimativas satisfatórias do volume comercial para a Caatinga Arbórea e obviamente com o menor custo computacional. Neste sentido, verificou-se

que a partir de 400 indivíduos as redes geradas estimavam o volume satisfatoriamente com um custo computacional bem inferior às redes de 700, 800 e ou 900 indivíduos.

Araújo (2015), quando avaliou redes NEAT para estimar o volume de formações florestais no estado de Minas Gerais, testou as seguintes configurações para o treinamento: número de indivíduos da população variando em 100, 300, 500, 600, 800 e 1000, e densidades de conexões iniciais de 0,1 e 0,5. A autora constatou que somente a partir de populações maiores que 600 foram geradas redes com bom desempenho e concluiu que a partir das duas densidades de conexões iniciais analisadas (0,1 e 0,5), é possível obter estimativas volumétricas com bom grau de exatidão para as onze tipologias florestais estudadas e diferentemente ao verificado neste estudo, recomendou populações iniciais com número de indivíduos maiores que 600, já que com populações inferiores não foram encontradas redes com bom desempenho para essa finalidade.

CONCLUSÕES

Os resultados mostram a potencialidade do uso de rede treinadas com o algoritmo Neuroevolution of Augmenting Topologies (NEAT) para estimar o volume na Caatinga Arbórea.

Recomenda-se para a configuração das redes durante a fase de treinamento densidade inicial 0,1 e populações iniciais com número mínimo de indivíduos igual a 400.

AGRADECIMENTOS

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).

REFERÊNCIAS BIBLIOGRÁFICAS

- ARAÚJO, A. C. S. C. **Redes neurais artificiais para a quantificação volumétrica de florestas nativas no estado de Minas Gerais**. 2015. 73p. Dissertação. (Mestrado em Ciência Florestal) – Universidade dos Vales do Jequitinhonha e Mucuri, Diamantina, 2015.
- BINOTI, D. H. B. **Sistemas computacionais aplicados ao manejo florestal**. 2012. 122 p. Tese. (Doutorado em Ciência Florestal) - Universidade Federal de Viçosa, Viçosa, 2012.
- BINOTI, M. L. M. S. et al. Redes Neurais Artificiais para estimação do volume de árvores. **Revista Árvore**, [s.1.], v.38, n.2, p.283-288, 2014.
- BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDERMIR, T. B. **Redes neurais artificiais: teoria e aplicações**. 2. ed.- [Reimpr.]. LTC, Rio de Janeiro, 2014.
- BUCKLAND, M. **AI Techniques for Game Programming**. Ohio: Premier Press, Inc., 2002.481p.
- HAYKIN, S. **Redes neurais: princípios e prática**. 2. ed. Bookman, Porto Alegre, 2001, 900 p.
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>. 2017.
- REZENDE, A. V. et al. Comparação de modelos matemáticos para estimativa do volume, biomassa e estoque de carbono da vegetação lenhosa de um cerrado sensu stricto em Brasília, DF. **Scientia Forestalis**, [s.1.], n. 71, p. 65-76, 2006.
- RIZZINI, C. T. **Tratado de fitogeografia do Brasil**. 2ª Edição. Âmbito Cultural Edições Ltda, Rio de Janeiro. 1997.
- RODRIGUES, F. L. **Metaheurística e Sistema de Suporte a Decisão no Gerenciamento de Recursos Florestais**. 2001. 239 p. Tese. (Doutorado em Ciência Florestal) - Universidade Federal de Viçosa, Viçosa, 2001.
- SANQUETTA, M. N. I. et al. Diversidade e estrutura fitossociológica da caatinga na região de Brumado – BA. **Enciclopédia Biosfera**, Centro Científico Conhecer - Goiânia, v.10, n.17; p. 2157 2014.
- STANLEY, K. O.; MIIKKULAINEN, R. Competitive Coevolution through Evolutionary Complexification. **Journal of Artificial Intelligence Research**, [s.1.], v. 21, p. 63-100, 2004.
- STANLEY, K. O.; MIIKKULAINEN, R. Evolving Neural Networks through Augmenting Topologies. **Evolutionary Computation**, [s.1.], v. 10, n. 2, p.99-127, 2002.