

MAPEAMENTO DA RELAÇÃO ALTURA-DIÂMETRO UTILIZANDO *SUPPORT VECTOR MACHINE***HIGHT-DIAMETER RELATION MAPPING USING *SUPPORT VECTOR MACHINE***

Sthefany Mendes Zuba¹ Emanuely Canabrava Magalhães Gabriela Leticia Ramos Carvalho
Thais Sales Gonçalves Adriana Leandra de Assis Carlos Alberto Araújo Júnior

RESUMO

A determinação da altura é uma operação onerosa e sujeita a erros, mais recentemente, com a evolução das tecnologias de sensoriamento remoto, principalmente do LiDAR, a obtenção da variável altura é feita direta e imediatamente. Dessa forma, buscou-se nesse trabalho avaliar as estimativas de diâmetro à altura de 1,30 m do solo a partir da ferramenta de análise Support Vector Machine (SMV), para isso foram utilizados dois critérios para a divisão do banco de dados para treinamento e validação, sendo o Treinamento 1 feito utilizando dados dos anos de 2012, 2013 e 2014 para treinamento e do ano de 2015 para validação, e o Treinamento 2 considerando uma divisão aleatória dos dados. Utilizou-se o software Statistica para treinamento das SMV. Após o processamento foram calculadas as estatísticas Bias, RQME, correlação e erro percentual médio; gerados gráficos de dispersão, considerando os valores estimados e observados; e histograma de resíduos. Foi possível concluir que a SMV podem ser utilizadas para estimativa de diâmetros à altura de 1,30 m. e que para estimativas mais precisas é necessário que os dados de treinamento sejam representativos da população em estudo.

Palavras-chave: LiDAR; mensuração florestal; inteligência artificial

ABSTRACT

Height determination is an onerous and error-prone operation, more recently with the evolution of remote sensing technologies, especially LiDAR, obtaining the height variable is done directly and immediately. In this work, we sought to evaluate the diameter estimates at 1.30 m height from the Support Vector Machine (SMV) analysis tool. Two criteria were used to divide the database for training and validation, with Training 1 being done using data from the years 2012, 2013 and 2014 for training and the year 2015 for validation, and Training 2 considering a random division of data. Statistica software for SMV training was used. After the processing, the Bias, RQME, correlation and mean percent error statistics were calculated; dispersion charts were generated, considering the estimated and observed values; and histogram of residues. It was possible to conclude that the SMV can be used to estimate diameters at a height of 1.30 m. and that for more accurate estimates it is necessary that the training data be representative of the study population.

Keywords: Lidar; forest measurement; artificial intelligence

INTRODUÇÃO

Duas das variáveis medidas durante a realização de inventários florestais são o diâmetro à altura do peito (Dap), medido a 1,3 metros do solo, e a altura total das árvores (Ht). O diâmetro é facilmente medido para todas as árvores, já a determinação da altura, de acordo com Couto e Bastos (1986), é uma operação onerosa e sujeita a erros. Apesar disso, a quantificação das alturas das árvores é fundamental, principalmente para a estimativa do volume de madeira dos povoamentos e para fazer classificações destes quanto as suas produtividades (THIERSCH *et al.*, 2013).

Para contornar os inconvenientes das medições das alturas, tem-se como prática obter a altura de algumas das árvores e os diâmetros de todos os indivíduos existentes na parcela amostral. Com isso, e através da relação entre os valores de Dap e Ht, pode-se estabelecer equações matemáticas, conhecidas como equações hipsométricas, que possibilitam estimar as alturas das demais árvores, reduzindo-se, portanto, o tempo e os custos de mensuração (THIERSCH *et al.*, 2013).

Mais recentemente, com a evolução das tecnologias de sensoriamento remoto, principalmente do LiDAR (*Ligth Detection and Ranging*), a obtenção da variável altura tem sido feita direta e imediatamente (DUBAYAH *et al.*, 2000). O LiDAR é um sensor que emite repetidamente pulsos infravermelhos em direção à

¹Estudante de graduação em Engenharia Florestal, Universidade Federal de Minas Gerais, Avenida Universitária, 1000, Universitário, CEP 39404-547, Montes Claros, MG, Brasil. E-mail: sthefanyzuba@gmail.com

superfície imageada. Um temporizador mede o tempo de viagem do pulso da aeronave para a superfície da Terra e da sua volta para o sensor. O tempo de ida e volta é então diretamente relacionado à distância, possibilitando, assim, a obtenção da altura total das árvores (WAGNER *et al.*, 2004). Entretanto, algumas características da floresta, como o diâmetro, não podem ser determinadas diretamente pelo LiDAR aerotransportado, sendo necessário que isso seja feito com utilização de modelos de regressão ou de ferramentas de inteligência computacional (DUBAYAH *et al.*, 2000).

Nesse sentido, é interessante que pesquisas sejam desenvolvidas com o intuito de avaliar diferentes técnicas, tal como o Support Vector Machine (SVM) para estimar o diâmetro das árvores a partir de dados de altura total. As SVM, desenvolvidas inicialmente por Vladimir Vapnik (VAPNIK e LERNER 1963; VAPNIK e CHERVONENKIS, 1964), tem como objetivo maximizar a capacidade de generalização, baseando nos princípios da Teoria de Aprendizado Estatístico (TAE). Esta teoria estabelece as condições matemáticas essenciais que permitem uma boa generalização, ou seja, um bom desempenho tanto para os conjuntos de treinamento quanto para os de teste (SMOLA e SCHOLKOPF, 2004).

Dessa forma, este trabalho teve como objetivo avaliar as estimativas de diâmetro à altura de 1,30 m do solo utilizando Support Vector Machine e tendo como principal variável independente o valor da altura total das árvores.

MATERIAIS E MÉTODOS

Os dados utilizados neste estudo são oriundos de inventários florestais contínuos realizados em plantios florestais na região do norte do estado de Minas Gerais durante o período de 2013 a 2016. Foram consideradas 604 parcelas amostrais totalizando 12.131 pares de dados de Dap e altura total (Ht) das árvores medidas em diferentes condições de sítio, espaçamento, material genético e rotação. As idades das árvores variaram entre 2 e 9 anos.

Para a divisão do banco de dados para treinamento e validação foram utilizados dois critérios, sendo o Modelo 1 aquele que utilizou dados dos anos de 2013, 2014 e 2015 para treinamento e do ano de 2016 para validação, e o Modelo 2 o que considerou uma divisão aleatória dos dados.

O processamento dos dados foi realizado no programa Statistica versão 13.0. O kernel escolhido foi a Função de Base Radial (RBF) (MONTAÑO, 2016), utilizando o padrão do próprio sistema para definir o parâmetro Gama e o otimizador *v-fold-cross-validation* para encontrar o fator Capacidade e o epsilon (COSENZA *et al.*, 2015), foi utilizado a regressão do tipo 1. Para estimativa do Dap foi utilizado com variáveis independentes a área útil por planta (AP), a idade (I) e a altura total (Ht).

Gráficos de dispersão considerando os valores estimados e observados foram construídos para avaliação dos resultados, bem como histogramas de resíduos. Os valores de erro para cada observação (e_i) foram obtidos utilizando-se a equação 1. Também foram calculadas as estatísticas de bias (equação 2), raiz quadrada do erro médio quadrático (equação 3), correlação entre valores estimados e observados (equação 4) e erro médio percentual (equação 5). Sendo:

$$e_i = \hat{y}_i - y_i \quad (\text{equação 1})$$

$$\text{bias} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i) \quad (\text{equação 2})$$

$$\text{RQME (\%)} = \frac{100}{\bar{y}} \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (\text{equação 3})$$

$$r = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (\text{equação 4})$$

$$\text{EMP} = \frac{1}{n} \sum_{i=1}^n \left[100 \cdot \frac{(y_i - \hat{y}_i)}{y_i} \right] \quad (\text{equação 5})$$

Em que: \hat{y}_i é o valor estimado e y_i é o valor observado para o i -ésimo dado, n é o número total de observações, \bar{y} é a média dos valores observados e $\bar{\hat{y}}$ é a média dos valores estimados.

RESULTADOS E DISCUSSÃO

Os dois modelos apresentam valores de bias, RQME (%) e EMP baixos e elevada correlação (Tabela 1), o que demonstra que o treinamento conseguiu mapear os padrões existentes em ambos os conjuntos de dados.

Entretanto, ao considerar a avaliação da capacidade de generalização do aprendizado da SMV o Modelo 2 apresentou valores melhores que aqueles encontrados para o Modelo 1.

Tais resultados podem estar relacionados com o fato de que para o Modelo 2, os dados de validação estão melhor distribuídos entre as diferentes classes diamétricas (Figura 1), o que permite que uma melhor estimativa dos valores (PEREIRA e CENTENO, 2017). Apesar de apresentar baixa correlação apresentada para o Modelo 1, as demais estatísticas ficaram próximas aos valores obtidos para o treinamento e para o Modelo 2, o que sugere que a SMV apresenta um elevado potencial para estimativa de valores de diâmetro à altura de 1,30 m do solo.

TABELA 1: Estatísticas calculadas para os dois tipos de treinamento. O treinamento 1 os dados foram divididos em 2012-2014 para treinamento e o ano de 2015 para validação. O treinamento 2 houve a divisão aleatória dos dados.

TABLE 1: Calculated statistics for the two types of training. Training 1 The data were divided into 2012-2014 for training and the year 2015 for validation. Training 2 randomly divided the data.

Etapa	Modelo 1				Modelo 2			
	Bias	RQME	Corr.	EMP	Bias	RQME	Corr.	EMP
Treinamento	-0,20	11,83	0,92	0,88	-0,05	11,64	0,92	1,43
Validação	0,12	11,41	0,62	2,49	0,17	12,22	0,90	2,89

Em que: RQME = raiz quadrada do erro médio quadrático; Corr.= correlação entre valores estimados e observados; EMP = erro médio percentual.

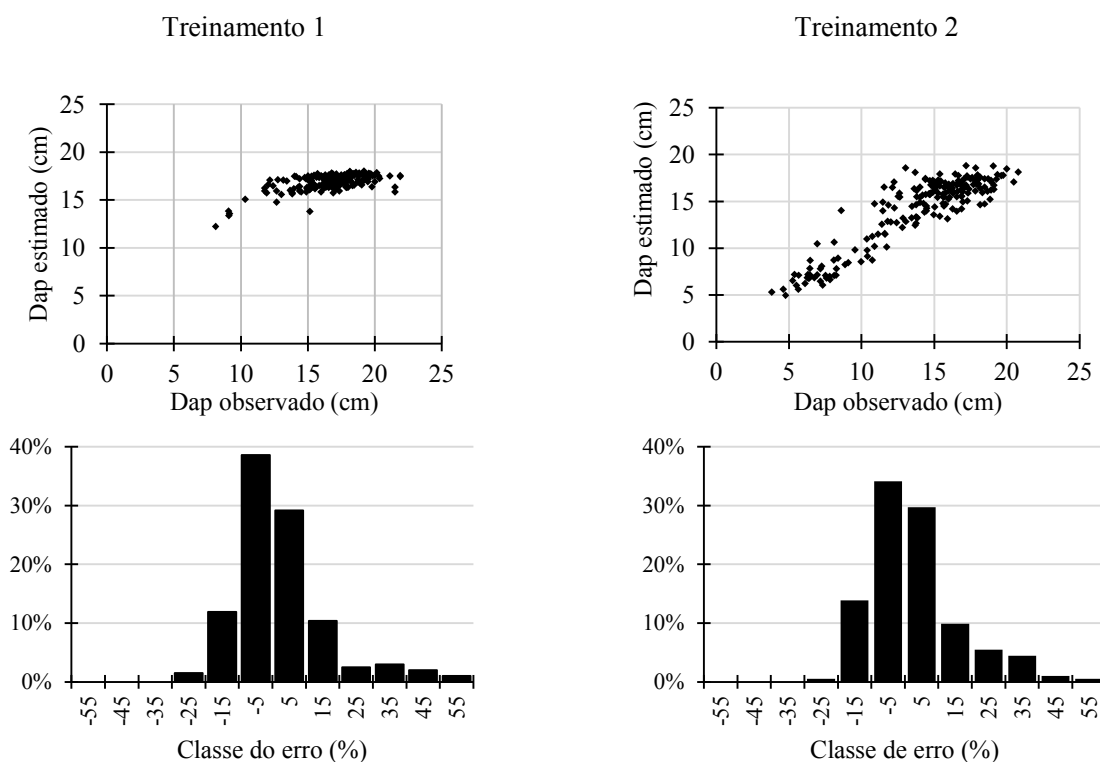


FIGURA 1: Gráficos de resíduos das estimativas de diâmetro utilizando Support vector machine.

FIGURE 1: Waste graphs of diameter estimates using Support vector machine.

A utilização de técnicas de inteligência artificial tem sido amplamente difundidas no meio florestal, entretanto é necessário que estudos criteriosos sejam realizados de modo a garantir aos manejadores o conhecimento suficiente para que os procedimentos de mensuração florestal sejam adaptados. Nesse sentido, deve-se procurar ferramentas que possam minimizar a quantidade de dados de campo obtidos de maneira manual. Como demonstrado, a SMV apresenta potencial para auxiliar nessa tarefa, no entanto ficou evidente que ainda não é possível utilizar apenas dados históricos para realizar estimativas para um inventário atual, principalmente em termos da obtenção da variável Dap como saída do modelo.

Recomenda-se, portanto, que novos estudos sejam realizados de maneira a obter as variáveis do povoamento que possam auxiliar nas estimativas considerando apenas os dados já coletados em inventários florestais realizados anteriormente.

CONCLUSÕES

Support Vector Machines podem ser utilizadas para estimativa de diâmetros à altura de 1,30 m. Para estimativas mais precisas é necessário que os dados de treinamento sejam representativos da população em estudo.

AGRADECIMENTOS

Os autores agradecem à Universidade Federal de Minas Gerais e à Fundação de Amparo à Pesquisa do Estado de Minas Gerais pelo apoio técnico e financeiro.

REFERÊNCIAS

- COSENZA, D. N., *et al.* Classificação da capacidade produtiva de sítios florestais utilizando máquina de vetor de suporte e rede neural artificial. **Scientia Forestalis**, v. 43, n. 108, p. 955-963, 2015.
- COUTO, H. T. Z.; BASTOS, N. L. M. Modelos de equações de volume e relações hipsométricas para plantações de *Eucalyptus* no Estado de São Paulo. **IPEF**, n. 37, p. 33-44, 1987.
- DUBAYAH, R., *et al.* Land surface characterization using LiDAR remote sensing. In: HILL, M. J.; ASPINALL, R. J. (Ed.). **Spatial Information for Land Use Management**. Amsterdam. Gordon & Breach Science Publishers, 2000, p. 25-38.
- MONTAÑO, R. A. N. R. **Aplicação de técnicas de aprendizado de máquina na mensuração florestal**. 2016. 102 f. Tese (Doutorado em Informática) – Universidade Federal do Paraná. Curitiba, Paraná, 2016.
- PEREIRA, G. H. de A.; CENTENO, J. A. S. Avaliação do tamanho de amostras de treinamento para redes neurais artificiais na classificação supervisionada de imagens utilizando dados espectrais e laser scanner. **Boletim de Ciências Geodésicas**, v. 23, n. 2, p. 268-283, 2017.
- SMOLA, A. J.; SCHOLKOPF, B. A tutorial on support vector regression. *Statistics and Computing*. **Kluwer Academic Publishers**, v.14, n.1, p.199–222, 2004.
- THIERSCH, C. R. *et al.* Estimativa da relação hipsométrica em clones de *Eucalyptus sp.* com o modelo de curtis ajustado por métodos bayesianos empíricos. **Revista Árvore**, v. 37, n. 1, p. 01-08, 2013.
- VAPNIK, V. N.; LERNER, A. Pattern recognition using generalized portrait method. **Automation and Remote Control**, v. 24, n. 1, p. 774-780, 1963.
- VAPNIK, V.N., CHERVONENKIS, A.Y. On the uniform convergence of relative frequencies of events to their probabilities. **Theory of Probability and Its Applications**, v. 16, n. 1, p. 264-280, 1971.
- WAGNER, W. *et al.* From single-pulse to full-waveform airborne laser scanners: potential and practical challenges. **Geo-Imagery Bridging Continents**, v. 35, n. 1, p. 201-206, 2004.