Breda Binoti, Daniel Henrique; Marques da Silva Binoti, Mayra Luiza; Garcia Leite, Helio;
Vivas Andrade, Alessandro; Saraiva Nogueira, Gilciano; Leles Romarco, Marcio; Grijó
Pitangui, Cristiano
SUPPORT VECTOR MACHINE TO ESTIMATE VOLUME OF EUCALYPT TREES

SUPPORT VECTOR MACHINE TO ESTIMATE VOLUME OF EUCALYPT TREES[1]

Daniel Henrique Breda Binoti[2]*, Mayra Luiza Marques da Silva Binoti[3], Helio Garcia Leite4, Alessandro Vivas Andrade[5], Gilciano Saraiva Nogueira[6], Marcio Leles Romarco[6] and Cristiano Grijó Pitangui[7]

[2] DAP Engenharia Florestal,Viçosa, MG - Brasil. E-mail: <danielhbbinoti@gmail.com>.
[3] Universidade Federal do Espírito Santo, Centro de Ciências Agrárias (CCA), Departamento de Ciências Florestais e da Madeira, Jerônimo Monteiro, ES - Brasil. E-mail: <MayraBinoti@gmail.com>.
[4] Universidade Federal de Viçosa, Centro de Ciências Agrárias, Departamento de Engenharia Florestal, Viçosa, MG - Brasil. E-mail: <hgleite@gmail.com>.
[5] Universidade Federal dos Vales do Jequitinhonha e Mucuri, Campus JK, Departamento de Sistema de Informação, Diamantina, MG - Brasil. E-mail: <alessandro.vivas@gmail.com>.
[6] Universidade Federal dos Vales do Jequitinhonha e Mucuri, Departamento de Engenharia Florestal, Diamantina, MG - Brasil, E-mail:<gnogueira@gmail.com> and <marcioromarco@gmail.com>.
[7] Universidade Federal de São João Del-Rei, Departamento de Tecnologia e Engenharia Civil, Computação e Humanidades, Ouro Branco, MG - Brasil. E-mail: <pitangui.cristiano@gmail.com>.
*Corresponding author.

ABSTRACT – This study aimed to test the application of the technique of support vector machines (SVM) to estimate the volume of eucalyptus trees. The data used in this study were from of 2307 trees of clonal hybrids (*Eucalyptus grandis* x *Eucalyptus urophylla*) located in southern Bahia. In the definition of stratification traditionally used 53 stratums were defined (defined by the stratification project and clone). He set the model of Schumacher and Hall for each stratum. The SVM were constructed to correlate the volume of trees on the basis of other independent variables which may be numeric as dbh and height and categorical as genetic material and design. The estimates were analyzed using statistical and graphical analysis of residues. The analysis consisted of the graphical inspection statistical dispersion of errors (residuals) in relation to the percentage of the values observed, and the analysis of the histogram of residues. The statistics used were the correlation between the observed and estimated volumes. The model of Schumacher and Hall showed the correlation between observed and predicted values of 0,993, and the SVM set of correlated 0,994. The SVM technology showed good adaptation to the problem, and this can use to predict the volumetric production of planted forests.

Keywords: Volume equations; Machines Learning; Schumacher and Hall.

MÁQUINAS DE VETORES DE SUPORTE PARA ESTIMAR VOLUME DE ÁRVORES DE EUCALIPTO

*RESUMO – O presente trabalho objetivou-se testar a aplicação da técnica de Máquinas vetores de suporte (SVM) para a estimação do volume de árvores de eucalipto. Os dados utilizados neste estudo foram provenientes de cubagens de 2.307 árvores de povoamentos clonais híbridos (**Eucalyptus grandis** x **Eucalyptus urophylla**) localizados no sul da Bahia. Na definição de estratificação de cubagens tradicionalmente usada foram definidos 53 estratos (definidos pela estratificação projeto e clone). Ajustou-se o modelo de Schumacher e Hall para cada estrato. As SVM foram construídas para correlacionar o volume das árvores em função das demais variáveis independentes que podem ser numéricas como dap e altura, e categórica como material genético e projeto. As estimativas foram analisadas empregando estatísticas e análise gráfica de resíduos. A análise gráfica consistiu na inspeção estatística da dispersão dos erros (resíduos) percentuais em relação aos valores observados, bem como na análise do histograma de resíduos. As estatísticas empregadas foram a correlação entre os volumes estimados e observados. O modelo de Schumacher e Hall apresentou a correlação entre valores observados e estimados de 0,993, sendo que a SVM ajustada apresentou correlação de 0,994. A tecnologia das SVM apresentou boa adequação ao problema, sendo esta possível de utilização para previsão volumétrica da produção de florestas plantadas.*

Palavras-chave: Volume equations; Machines Learning; Schumacher and Hall.

1. INTRODUCTION

The use of volumetric equations is one of the main tools for the quantification of volume production of forest stands, and the basis for decision making in the management of forest. Conventionally the volume of cubed trees closely is correlated with easily measurable variables of the population as total height and diameter of 1.30 m in height (dbh) (SILVA et al., 2009).

Several models are used to estimate the volume of trees, among them stands out the model proposed by Schumacher and Hall (1933) is one of the most widespread in the forest area (LEITE; ANDRADE, 2002; CAMPOS; LEITE, 2009). However, the introduction of artificial intelligence techniques has aroused great interest in the sector by generating volumetric estimates with greater accuracy and efficiency, among them stand out the Artificial Neural Networks (ANN). ANN are parallel systems composed of simple processing units, called artificial neurons connected to each other in a specific way (SILVA et al., 2009; BINOTI, 2010). Artificial neurons are simplified mathematical models of biological neurons (HAYKIN, 2001).

ANN are powerful tools used to approach functions is to generate a neural network that approximates the unknown function f(x), which describes the relationship of the input-output pairs {(x1, y1), (x2, y2) , ..., (xn, yn)} from a set of n training patterns. However, some problems related to training and sizing of ANN have led to its misuse resulting in pretentious estimates of its potential.

The training an ANN consists of the design of the topology as well as the determination of the values of the weights between connections. This is equivalent to determining the existing mathematical relationships between variables and the determination of the parameters of a statistical model (VALENÇA, 2009). Another extremely important feature when training an ANN is to determine the stopping criteria. The determination of these criteria defines the quality of ANN during use.

In view of this, came the support vector machine (SVM), these networks are a type of ANN fed forward (HAYKIN, 2001). These networks have a high ability to generalize because their training search beyond the error minimization training to minimize the complexity of the obtained ANN (VALENÇA, 2009). This study aimed to evaluate the use of SVM technique to estimate the volume of eucalyptus trees.

2. MATERIAL AND METHODS

2.1 Data

The data used in this study were 2,307 trees stands clonal hybrids (*Eucalyptus grandis* x *Eucalyptus urophylla*) located in southern Bahia. The data showed DBH ranging from 4.5 to 28.3 cm, total height (Ht) ranging from 6.6 to 33.8 m, age ranging from 2.7 to 11 years. The data come from 21 projects and 15 clones. In cubagens stratification definition used traditionally defined 53 strata (defined by layering design and clone). The sample trees were cubed in 1 m sections, the total volume in shell (V) obtained by applying the Smalian formula.

2.2 Model Volumetric

It was performed to obtain estimates of the volume using the method conventionally used in forestry. For each layer adjusted to Schumacher model and Hall, as described below:

$$LnV = \beta_0 + \beta_1 * LnDAP + \beta_2 * LnHt + \varepsilon$$

where, V = volume, m3; DAP = diameter at 1.3 m height, cm; Ht = total tree height, m; Ln = natural logarithm; = βi parameters; and ε = random error, ε ~ NID (0, σ²).

2.3 Support Vector Machines

The construction of vectors is based on statistical machine learning process described in detail by Haykin (2001) and Steinwart and Christmann (2008). Briefly, the use of SVM to estimate the volume of trees consists in determining the volume dependence in relation to the remaining independent variables that can be numerical and height as dap, or genetic material as categorical and stand. In this work it was used as continuous independent variables dap (cm), height (m), age (years) and categorical variables genetic material (clone name) and project (project name).

It was used as error function type II function also known as nu-SVM which has the following function.

$$\frac{1}{2} w^T w - c \left(v\varepsilon + \frac{1}{N} \sum_{i-1}^{N} (\zeta_i + \zeta_i^*) \right)$$

Subject to the restrictions

$$(w^t \ \phi(x_i) + b) - y_i \leq \varepsilon + \zeta_i$$

$$y_i - (w^t \ \phi(x_i) + b_i) \leq \varepsilon + \zeta_i^*$$

$$\zeta_i^*, \zeta_i > 0, i = 1, \dots, N, \ \varepsilon \geq 0$$

wherein C is the capacity constant, w is the vector of coefficients, and b is a constant parameter are the index i represents the training cases, N is the total number of cases, ö is the kernel used. The kernel function used was the type RBF (radial basis function). The construction of the SVM has been made in Excel environment with the aid of techniques for *Visual Basic for Applications.*

2.4 Evaluation of Estimates

The volumes by using equations and SVM were observed compared with the corresponding volumes. Estimates were analyzed using statistical and graphical analysis of waste. The graphical analysis consisted of the statistical dispersion of inspection errors (waste) percentage relative to values observed, as well as waste histogram analysis. The statistics used were correlation between the estimated and observed volumes.

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

wherein the variance s² and cov is the covariance.

3. RESULTS

They were adjusted 53 models of Schumacher and Hall and SVM to describe the job data. Both tested models showed results without bias and bias-free. The model Schumacher and Hall showed the correlation between observed and estimated values of 0.993, and the correlated set of SVM 0.994. The presented set SVM 63 vectors, and these 9-limiting.

The graphical analysis of adjustments Schumacher and Hall model and SVM is shown in Figure 1. It can be seen by residues histogram a higher dispersion for Schumacher and Hall volume model.

4. DISCUSSION

This study was motivated mainly by the growing interest of forest science in the use of artificial intelligence techniques. The application of ANN in the forest area has aroused great interest in the forest area, mainly
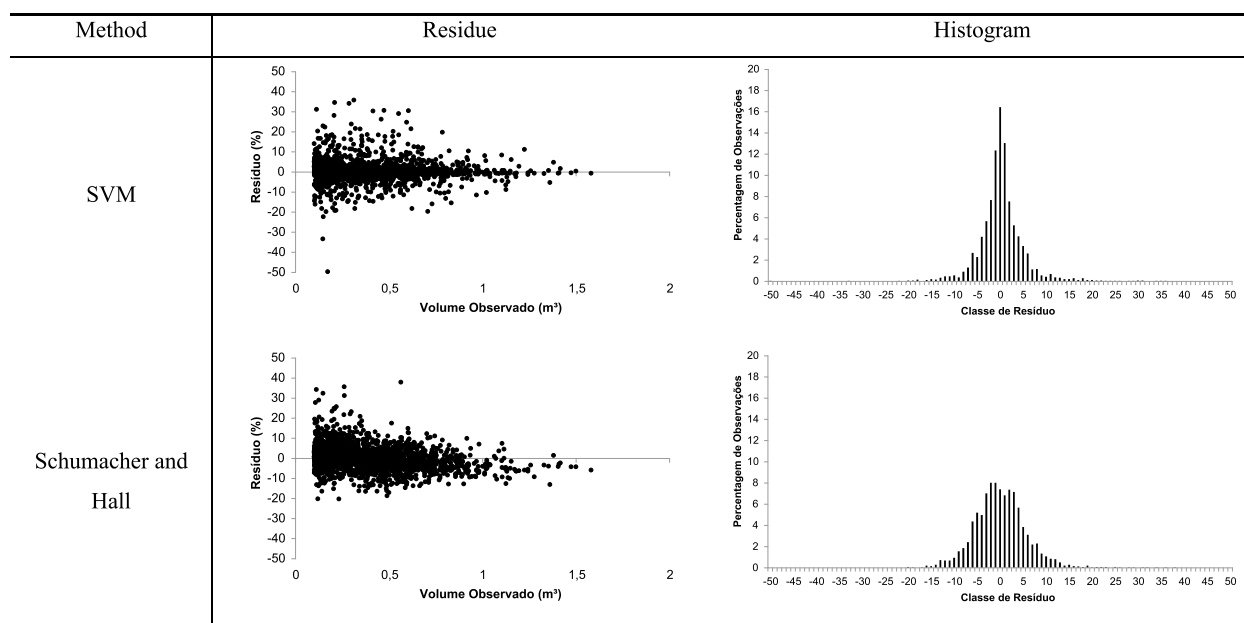

Figure 1 – Graphical analysis of the model fit the Schumacher and Hall and SVM.
Figura 1 – *Análise gráfica do ajuste do modelo de Schumacher e Hall e da MVS.*

due to the work of Silva et al., (2009) Leite et al., (2010) and Binoti (2010). However, the use of ANN has as main concern the definition of the number of neurons in the great hidden layers as well as in the definition of stopping criteria. The concern with these factors is directly related to the generalizability of ANN generated, ie a very large amount of neurons in the hidden layer can cause the network only memorize the training data, over their generalizability, which causes a biased result during implementation. The same problem is seen with overtraining (HAYKIN, 2001; VALENÇA, 2009).

The use of SVM has the main objective to obtain ANN with high generalizability, since, during the training of these ANN is sought not only minimize the error training, but also the complexity of this (VALENÇA, 2009). Various applications using the SVM to approximate functions are found (DRUCKER et al., 1997; CHERKASSKY, MULIER, 1998; SCHÖLKOPF et al, 1998; SMOLA; SCHÖLKOPF, 1998; MATTERA; HAYKIN, 1999; MULLER et al., 1999; VAPNIK, 1998, 1999; SCHÖLKOPF et al., 1999; KWOK, 2001; SCHÖLKOPF; SMOLA, 2002).

The SVM technique presented a considerable gain in accuracy when compared to the model of Schumacher and Hall. This difference is evident when analyzing the waste histogram (Figure 1). It can be seen that for the estimations originated by SVM 90% of the residues are in range between classes of about 5%, while for the models Schumacher and Hall this amplitude reaches the plus or minus 10 classes %.

The application of the SVM technique has the major advantage of using only a tool to estimate volume of the entire population independent of the number of layers that may exist. The use of traditional volumetric equations despite having greater simplicity requires a longer time to adjust and obtain the equations, as well as their storage and application. The case study demonstrates that, for the problem in question were generated 53 volumetric equations to be examined individually, in contrast to SVM generated only a mathematical model.

One of the main characteristics for SVM technique is that they are able to cope with imperfect or outlines data. The formulation of SVM encompass encompasses the principle of minimizing the structural risk - structural risk minimization (SRM), and the principle of minimizing the empirical risk - empirical risk minimization (ERM).

SRM involves minimizing the generalization error, whereas RME involves minimizing the error to the training data, this fact guarantees a better estimate for data not observed (HAYKIN, 2001).

This work presents itself as the introductory SVM technique for obtaining volumetric estimates, and technical parameterization work should be done as well, its application to other problems of the forest area.

5. CONCLUSIONS

The technology SVM showed good adaptation to the problem, which is possible to use for volume production forecast of planted forests.

6. REFERENCES

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

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

CHERKASSKY, V.; MULIER, F. **Learning from data:** concepts, theory, and methods. New York: Wiley, 1998.

DRUCKER, H.; BURGES, C.; KAUFMAN, L.; SMOLA, A.; VAPNIK, V. **Support vector regression machines**. In: MOSER, M.; JORDAN, J.; PETSCHE, T. (Ed.) **Neural information processing systems**. Cambridge: MIT Press, 1997. v.9. p.155-161.

HAYKIN, S. **Redes neurais:** princípios e prática. Porto Alegre: 2001. 900p.

KWOK, J.T. **Linear dependency between and the input noise in 1-support vector regression**. In: DORFFNER, G.; BISHOF, H.; HORNIK, K. (Ed.) **ICANN 2001**. p.405-410.

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

LEITE, H.G.; SILVA, M.L.S.; BINOTI, D.H.B.; FARDIN, L.; TAKIZAWA, F.H. Estimation of inside-bark diameter and heartwood diameter for *Tectona grandis* Linn. trees using artificial neural networks. **European Journal Forest Research**, v.130, p.263-269, 2011.

MATTERA, D.; HAYKIN, S. Support vector machines for dynamic reconstruction of a chaotic system. In: LKOPF, SCHO, B.;¨BURGES, J.; SMOLA, A. (Ed.) **Advances in kernel methods**: Support vector machine. Cambridge: MIT Press, 1999.

MULLER, K.; SMOLA, A.; RATSCH, G.; SCHÖLKOPF, B.; KOHLMORGEN, J.; VAPNIK, V. Using support vector machines for time series prediction. In: SCHO¨LKOPF, B.; BURGES, J.; SMOLA, A. (Ed.) **Advances in kernel methods**: Support vector machine. Cambridge, MA: MIT Press, 1999.

SCHÖLKOPF, B.; BARTLETT, P.; SMOLA, A.; WILLIAMSON, R. Support vector regression with automatic accuracy control. In: NIKLASSON, L.; BODE,´M.; ZIEMKE, N, T. (Ed.) **Proceedings of ICANN'98.** Perspectives in neural computing. Berlin: Springer, 1998. p.111–116.

SCHÖLKOPF, B.; BURGES, J.; SMOLA, A. **Advances in kernel methods**: Support vector machine. Cambridge, MA: MIT Press, 1999.

SCHÖLKOPF, B.; SMOLA, A. **Learning with kernels:** Support vector machines, regularization, and beyond. Cambridge, MA: MIT Press, 2002.

SILVA, M.L.M.; 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.

SMOLA, A.; SCHÖLKOPF, B. A tutorial on support vector regression. **NeuroCOLT Technical Report NC-TR-98-030.** London: Holloway College, University of London, 1998.

STEINWART, I.; CHRISTMANN, A. **Support vector machines**. New York: Springer, 2008.

VALENÇA, M.J.S. **Fundamentos das redes neurais:** exemplos em Java. Recife: Livro Rápido, 2009. 382p

VAPNIK, V. **Statistical learning theory** . New York: Wiley, 1998.

VAPNIK, V. **The nature of statistical learning theory**. Berlin: Springer, 1999.