









Artificial neural networks, quantile regression, and linear regression for site index prediction in the presence of outliers

Carlos Alberto Araújo Júnior⁽¹⁾ ,
Pábulo Diogo de Souza⁽²⁾ ,
Adriana Leandra de Assis⁽¹⁾ ,
Christian Dias Cabacinha⁽¹⁾ ,
Helio Garcia Leite⁽³⁾ ,
Carlos Pedro Boechat Soares⁽³⁾ ,
Antonilmar Araújo Lopes da Silva⁽⁴⁾  and
Renato Vinícius Oliveira Castro⁽⁵⁾ 

⁽¹⁾Universidade Federal de Minas Gerais, Instituto de Ciências Agrárias, Campus Regional de Montes Claros, Avenida Universitária, nº 1.000, Bairro Universitário, CEP 39404-547 Montes Claros, MG, Brazil. E-mail: araujocaj@gmail.com, alassis@ufmg.br, christian.cabacinha@gmail.com

⁽²⁾Universidade Federal de Santa Maria, Centro de Ciências Rurais, Departamento de Ciências Florestais, Avenida Roraima, nº 1.000, Cidade Universitária, Camobi, CEP 97105-900 Santa Maria, RS, Brazil. E-mail: pabulodiogo@gmail.com

⁽³⁾Universidade Federal de Viçosa, Departamento de Engenharia Florestal, Avenida Purdue, s/nº, Campus Universitário, Edifício Reinaldo de Jesus Araújo, CEP 36570-900 Viçosa, MG, Brazil. E-mail: hgleite@gmail.com, csoares@ufv.br

⁽⁴⁾Celulose Nipo-Brasileira S.A., Rodovia MG 758, Km 3, s/nº, Distrito Perpétuo Socorro, CEP 35196-000 Belo Oriente, MG, Brazil. E-mail: antonilmar@gmail.com

⁽⁵⁾Universidade Federal de São João Del-Rei, Departamento de Ciências Agrárias, Campus Sete Lagoas, Rua Sétimo Moreira Martins, nº 188, Itapoã II, CEP 35702-031 Sete Lagoas, MG, Brazil. E-mail: castrorvo@ymail.com

✉ Corresponding author

Received
February 12, 2018

Accepted
October 30, 2018

How to cite

ARAÚJO JÚNIOR, C.A.; SOUZA, P.D. de; ASSIS, A.L. de; CABACINHA, C.D.; LEITE, H.G.; SOARES, C.P.B.; SILVA, A.A.L. da; CASTRO, R.V.O. Artificial neural networks, quantile regression, and linear regression for site index prediction in the presence of outliers. *Pesquisa Agropecuária Brasileira*, v.54, e00078, 2019. DOI: <https://doi.org/10.1590/S1678-3921.pab2019.v54.00078>.

Abstract – The objective of this work was to compare methods of obtaining the site index for eucalyptus (*Eucalyptus* spp.) stands, as well as to evaluate their impact on the stability of this index in databases with and without outliers. Three methods were tested, using linear regression, quantile regression, and artificial neural network. Twenty-two permanent plots from a continuous forest inventory were used, measured in trees with ages from 23 to 83 months. The outliers were identified using a boxplot graphic. The artificial neural network showed better results than the linear and quantile regressions, both for dominant height and site index estimates. The stability obtained for the site index classification by the artificial neural network was also better than the one obtained by the other methods, regardless of the presence or the absence of outliers in the database. This shows that the artificial neural network is a solid modelling technique in the presence of outliers. When the cause of the presence of outliers in the database is not known, they can be kept in it if techniques as artificial neural networks or quantile regression are used.

Index terms: *Eucalyptus*, artificial intelligence, dominant height, forest inventory, forest modelling, non-sampling errors.

Redes neurais artificiais, regressão quantílica e regressão linear para predição do índice de sítio na presença de “outliers”

Resumo – O objetivo deste trabalho foi comparar métodos para obtenção do índice de sítio para povoamentos de eucalipto (*Eucalyptus* spp.), bem como avaliar seus impactos na estabilidade desse índice em bases de dados com e sem a presença de “outliers”. Foram testados três métodos, com uso de regressão linear, regressão quantílica e rede neural artificial. Foram utilizadas 22 parcelas permanentes de inventário florestal contínuo, medidas em árvores com idade de 23 a 83 meses. Os outliers foram identificados com uso de gráfico de boxplot. A rede neural artificial proporcionou melhores resultados que as regressões linear e quantílica, tanto para as estimativas de altura dominante quanto do índice de sítio. A estabilidade da classificação do índice de sítio obtida pela rede neural artificial também foi melhor que a obtida com os outros métodos, independentemente da presença ou da ausência de outliers na base de dados. Isso indica que a rede neural artificial é uma técnica sólida de modelagem na presença de outliers. Quando a causa da presença de outliers na base de dados não é conhecida, eles podem ser mantidos nela se técnicas como as de redes neurais artificiais ou de regressão quantílica forem utilizadas.

Termos para indexação: *Eucalyptus*, inteligência artificial, altura dominante, inventário florestal, modelagem florestal, erros não amostrais.

Introduction

Understanding growth and yield processes in forests is important to their rational management (Cosenza et al., 2015). Statistical techniques for data modelling can support decision making during forest planning, and information on site productive capacity can help delineating resource allocation strategies for management units, silvicultural treatments, and wood volume prognosis (Campos & Leite, 2017).

The potential for trees to grow at a specific location can be determined by site index estimates, obtained by direct or indirect methods (Binoti et al., 2012). Among direct methods, the guide-curve is the most commonly used, whose application includes the adjustment of regression models that relate dominant height and forest age data (Scolforo, 2006). When this procedure is used for the classification of site productive capacity, its evaluation generally includes the stability analysis (Machado et al., 2011), which deals with the estimation of the number of plots (or samples) that remain in the same site index class over time.

Databases used for dominant height modelling in forest stands are obtained from permanent or temporary plots, during a forest inventory, or from stem analyses (Scolforo, 1997). These measurements, although simple, must be done carefully in order to avoid non-sampling errors (Soares et al., 2011), which can impact the analysis when one or more values are out of the general trend of the data and are considered as outliers. Because of their marked difference, according to Santos et al. (2015), the number of outliers should be as low as possible. There are several ways to identify them, among which stands out the boxplot graphic (Schwertman et al., 2004).

It should be noted that sometimes outliers are not measurement errors, but values with a different distribution. In these cases, a deep evaluation about the cause of their discrepant behaviour is necessary before excluding them from database. In other situations, the set of discrepant data can result in new strata for modelling. When the analysis of their origin is not possible, due to the amount of data or to the impossibility of assessing sampling procedures, it is important to keep the outliers in the database. Therefore, mathematical techniques suitable for other types of distribution are required to incorporate these values into the database, aiming at a robust modelling

process, without the possibility of phasing out real information, although discrepant.

The use of the quantile regression (QR) and artificial neural network (ANN) is common for this kind of robust modelling. The first type of analysis was proposed by Koenker & Basset (1978) and is rather robust when there are outliers in the database, since they have no effect on the distribution of the conditional median and can be used to model any specified quantile of a dataset (Abellanas et al., 2016). The second is part of the artificial intelligence field. It was firstly described by McCulloch & Pitts (1943) and has been considered as an excellent alternative to the traditional regression models. Both the QR and ANN have already been used with success to fit dominant height data, with focus on site index estimates (Cosenza et al., 2015; Araújo Júnior et al., 2016); however, these studies did not consider the presence of outliers in the database.

The objective of this work was to compare methods of obtaining the site index for eucalyptus stands, as well as to evaluate their impact on the stability of this index in databases with and without outliers.

Materials and Methods

The study was developed using data from 22 rectangular, permanent plots (approximately 330 m²) of a continuous forest inventory carried out in forest stands in the Vale do Rio Doce region, in the east of the state of Minas Gerais, Brazil. Measurements were done in trees with ages from 20 to 83 months, with six observations per plot. Dominant height was obtained for each measurement, considering the average height of the 100 largest trees per hectare, and diameter at breast height was determined as in Assman (1970). Descriptive statistics are shown in Table 1.

Observations were grouped according to age classes, with a 12-month range, in order to identify possible discrepant data. A boxplot was built for each age class, and data out of the confidence interval were considered as outliers, as proposed by Schwertman et al. (2004). Therefore, two databases were created: one containing outliers and another without them.

Schumacher's model was adjusted for each database, as in Demolinari et al. (2007), and is described as: $\ln H = \beta_0 + \beta_1 A^1 + \varepsilon$, where H is the dominant height (meters); A is the age of the trees at measurement (months); β_0 and β_1 are parameters; Ln is the natural

logarithm; and ε is the random error, with $\varepsilon \sim \text{NID}(0, \sigma^2)$. For Schumacher's model, the adjustment was done with the ordinary least square method and with the minimization of the absolute error for the QR (Araújo Júnior et al., 2016), using the following equations:

$$\min \sum_{i=1}^n \rho_{\tau} |r_i|; \text{ and } r_i = y_i - \sum_{j=1}^k x_{ij} \beta_j;$$

where n is the number of observations; ρ_{τ} is the weight for residual i , determined as $2q$ (quantile) if $r_i > 0$ or $2(1 - q)$ if otherwise; r_i is the residual for the i^{th} observation; y_i is the i^{th} observed value; k is the number of parameters to be estimated; x_{ij} is the value of the j^{th} independent variable for the i^{th} observed data; and β_j is the j^{th} parameter to be estimated.

For the QR method, a percentile of 50% (median) was considered. In this case, the adjustment was made using the quantreg statistical package (Koenker, 2013), developed for the R software (R Core Team, 2014).

The two databases were, then, used for training the ANN, considering a multilayer perceptron structure, with age as an input variable and dominant height as an output variable. ANN training was carried out using the resilient propagation algorithm, 5 neurons in the hidden layer and 3,000 epochs of training; these parameters were defined according to Binoti et al. (2015). The NeuroForest software was used to obtain ANN parameters (Binoti et al., 2014).

The average percentage relative error (E_{APR}), the mean absolute error (MAE), the root mean square error (RMSE), and the correlation coefficient between observed and estimated values ($r_{y\hat{y}}$) were used to evaluate the quality of the adjustments. In addition, histograms of the E_{APR} for each method were built. The estimated parameters were obtained with the following equations:

$$E_{\text{APR}} = \frac{100}{n} \sum_{i=1}^n \left(\frac{\hat{y}_i - y_i}{y_i} \right), \text{ MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|,$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}, \text{ and } r_{y\hat{y}} = \frac{\text{cov}(y, \hat{y})}{\sqrt{\text{var}(y) \text{var}(\hat{y})}}$$

where n is the total number of observations, \hat{y}_i is the estimated value for observation i , y_i is the correspondent observed value, cov is the covariance, and var is the variance. The estimated parameters for the QR and linear regression (LR) were evaluated by the t-test, at 5% probability.

The guide-curve method, applying a reference age of 72 months to Schumacher's model, was used to determine the site index using the estimates obtained from the LR and QR models (Campos & Leite, 2017), through the equation: $\text{LnS} = \text{LnH} - \beta_1(1/A - 1/I_A)$, where S is the site index value (meters), H is the dominant height (meters), A is the age at measurement time (months), and I_A is the index age (months).

A new training – considering a data structure similar to the one used for Schumacher's model – was performed to obtain the site index by a neural network. Therefore, the ANN was trained considering a multilayer perceptron structure. Dominant height at the current age, current age, and future age were considered as input variables, and dominant height at a future age, as an output variable. The trained ANN was applied using the index age of 72 months as the future age in order to obtain the site index for each measurement.

The site index estimates obtained from the LR, QR, and ANN were classified using the equation:

$$\text{Class} = \text{int} \left(\frac{S}{R} \right) R + \frac{R}{2}$$

where int is a function that returns the integer value of a real number, S is the site index estimated for each plot in each measurement time, and R is the range of class (meters) considered. This equation returns the central value of the class as a function of the class range – in this study, an interval equal to 5.0 m was considered.

The results of the classification of productive capacity using the site index were subjected to the stability analysis, as suggested by Scolforo (2006) and Chaves et al. (2016). In this case, five intervals of last measurements were considered: interval 1, measurements from 1 to 6; interval 2, from 2 to 6;

Table 1. Descriptive data statistics for eucalyptus (*Eucalyptus* spp.) stands.

Age class (months)	Basal area (m ² ha ⁻¹)		Dominant height (m)		Total volume (m ³ ha ⁻¹)	
	Average	Standard deviation	Average	Standard deviation	Average	Standard deviation
18	6.09	1.40	10.43	1.40	25.44	8.48
30	12.05	2.44	17.92	3.23	90.46	32.45
42	17.74	2.59	22.80	2.60	166.86	34.71
54	20.23	3.06	25.44	2.94	209.67	42.30
66	22.43	3.73	27.43	3.32	257.45	60.62
78	23.54	3.49	28.30	3.92	284.44	66.10

interval 3, from 3 to 6; interval 4, from 4 to 6; and interval 5, from 5 to 6. The amount of plots that remained in a same class, in all measurements, was counted for each range. Stability was obtained from the percentage of stable plots.

Results and Discussion

Using a boxplot, six outliers, all located below data distribution, were detected for each dominant height measurement period (Figure 1). When the presence of outliers is as strong as the one observed here, considering them in modelling generally enhances database consistency. It should be highlighted that,

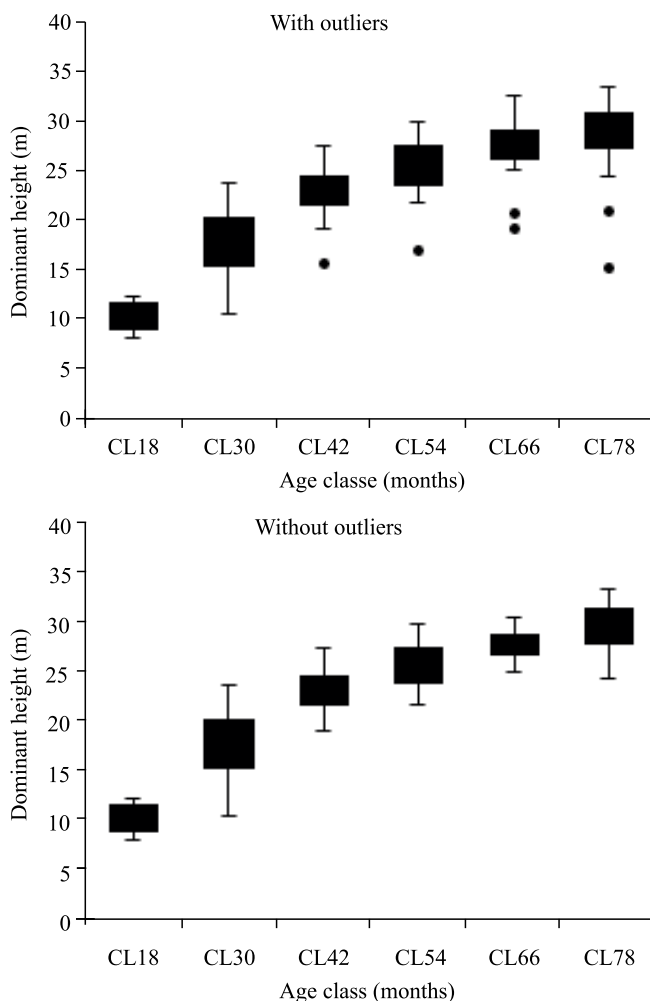


Figure 1. Boxplot graphics for the variables dominant height and age class of eucalyptus (*Eucalyptus* spp.) stands from databases with and without outliers. Age class was measured with a 12-month range (CL18 to CL78).

on the one hand, their use may cause a loss in the explanatory power of regression models (Schwertman et al., 2004), whereas, on the other hand, it can completely change the analysis if the causes for their presence cannot be determined.

The boxplot analysis did not evidence discrepant observations after the exclusion of the outliers (Figure 1). The comparison of databases with and without outliers indicated little discrepancy for the averages and medians, in each class of dominant height (Table 2).

Considering Schumacher's model for the two databases, the adjusted equations generated parameters statistically different from zero ($p < 0.05$), both for the LR and QR (Table 3). It is possible to note that the values obtained for the parameters using the different

Table 2. Averages and medians according to the dominant height classes of eucalyptus (*Eucalyptus* spp.) stands, measured with a 12-month range (CL18 to CL78).

Dominant height class (m)	With outliers		Without outliers ⁽¹⁾	
	Average	Median	Average	Median
CL18	10.43	10.54	10.43 (0.00%)	10.54 (0.00%)
CL30	17.92	18.27	17.92 (0.00%)	18.27 (0.00%)
CL42	22.80	22.72	23.14 (1.49%)	22.87 (0.66%)
CL54	25.44	25.76	25.84 (1.57%)	26.17 (1.59%)
CL66	27.43	27.87	27.80 (1.35%)	27.87 (0.00%)
CL78	28.30	28.77	29.15 (3.00%)	28.81 (0.14%)

⁽¹⁾Values within parenthesis define the variation percentage between databases with and without outliers.

Table 3. Estimated parameters for Schumacher's model considering different adjustment methods.

Regression type	Parameter	Value	Standard error	p-value
Database with outliers				
Linear	β_0	3.7430	0.0282	< 0.05
	β_1	-29.7093	1.0902	< 0.05
Quantile	β_0	3.7747	0.0312	< 0.05
	β_1	-30.2833	1.1967	< 0.05
Database without outliers				
Linear	β_0	3.7886	0.0224	< 0.05
	β_1	-30.9034	0.8490	< 0.05
Quantile	β_0	3.7807	0.0304	< 0.05
	β_1	-30.4042	1.1436	< 0.05

methods differed in the database with outliers, but were similar in the one without them. This result is due to the fact that the LR is less tolerant to the presence of noise than the QR (Araújo Júnior et al., 2016), mainly when the discrepant values promote a change in average values, but interfere little in the median, as was the case for some dominant height classes (Table 2).

Graphically, it is possible to observe a discrepancy between the curves for the values estimated by the LR and QR only when there are outliers in the database (Figure 2). This is in alignment with the results from

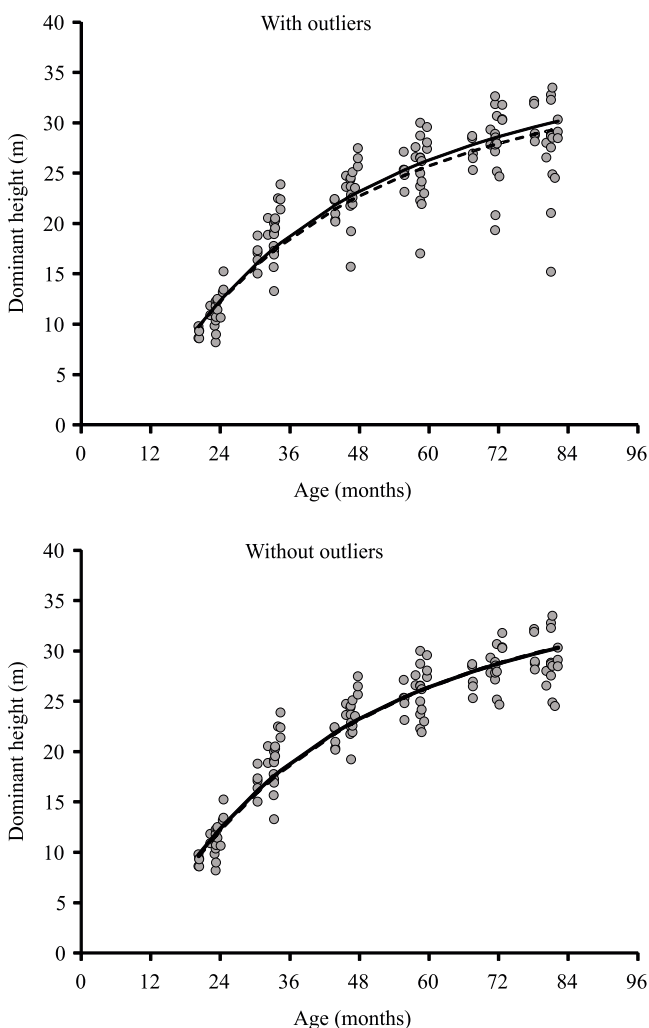


Figure 2. Schumacher's model adjusted considering the quantile (continuous line) and linear (dotted line) regressions for the variables dominant height and age class of eucalyptus (*Eucalyptus* spp.) stands from databases with and without outliers.

Araújo Júnior et al. (2016), who reported a dislocation of the curve in the direction of the outliers. These variations can promote imprecise values, mainly when there are no guarantees that the outliers really must be part of the sampling.

The statistics for regression and for the ANN (Table 4) reveal better MAE, RSME, and $r_{y\hat{y}}$ values for the latter, which showed greater robustness in estimating dominant height values. The QR had a slightly superior adjustment than the LR, which was also found by Araújo Júnior et al. (2016).

When the database had no outliers, the statistics were improved, compared with the previous analysis (Table 4). However, the order of the quality of the adjustment remained the same, i.e., the ANN results were superior to the ones obtained by the QR and LR. Indeed, recent literature shows accuracy gains due to the adoption of neural networks instead of classic regression models (Binoti et al., 2015; Miguel et al., 2015).

The histograms of the estimated values from the database with outliers showed errors above 50% (Figure 3). Even when techniques considered robust to outliers, such as the QR and ANN, were used, errors still occurred at high levels. This can be explained by the behaviour of the techniques that consider outliers – or give them a great importance – when modelling a denser set of data.

Considering the second ANN, trained to obtain the values of dominant height at a future age, errors between -20 and +30% were observed (Figure 4), as

Table 4. Dominant height statistics for different methods of estimation, according to eucalyptus (*Eucalyptus* spp.) plant age.

Statistic ⁽¹⁾	Linear regression	Quantile regression	Artificial neural network
Database with outliers			
E_{APR}	1.0259	2.8944	1.5875
MAE	2.0960	2.0931	1.8407
RQEM	2.9424	2.9850	2.7128
$r_{y\hat{y}}$	0.8984	0.8981	0.9133
Database without outliers			
E_{APR}	0.5613	0.9764	0.9737
MAE	1.7256	1.7128	1.5267
RMSE	2.1710	2.1557	1.9739
$r_{y\hat{y}}$	0.9468	0.9471	0.9553

⁽¹⁾ E_{APR} , average percentage relative error; MAE, mean absolute error; RMSE, root mean square error; and $r_{y\hat{y}}$, correlation coefficient between observed and estimated values.

well as E_{APR} , MAE, and RSME values near zero for databases with and without outliers (Table 5). The difference between the estimates of the two ANNs can be attributed to the projection of the dominant height in the second one, where the neural network output (dominant height at a future age) was dependent on the dominant height at the current age. This might

provide a better learning to the ANN, compared with the patterns of inputs and outputs of the system. These results agree with those found by Cosenza et al. (2015), who pointed out that the classification of productive capacity using neural networks provided consistent results, superior to those observed with the application of a support machine vector.

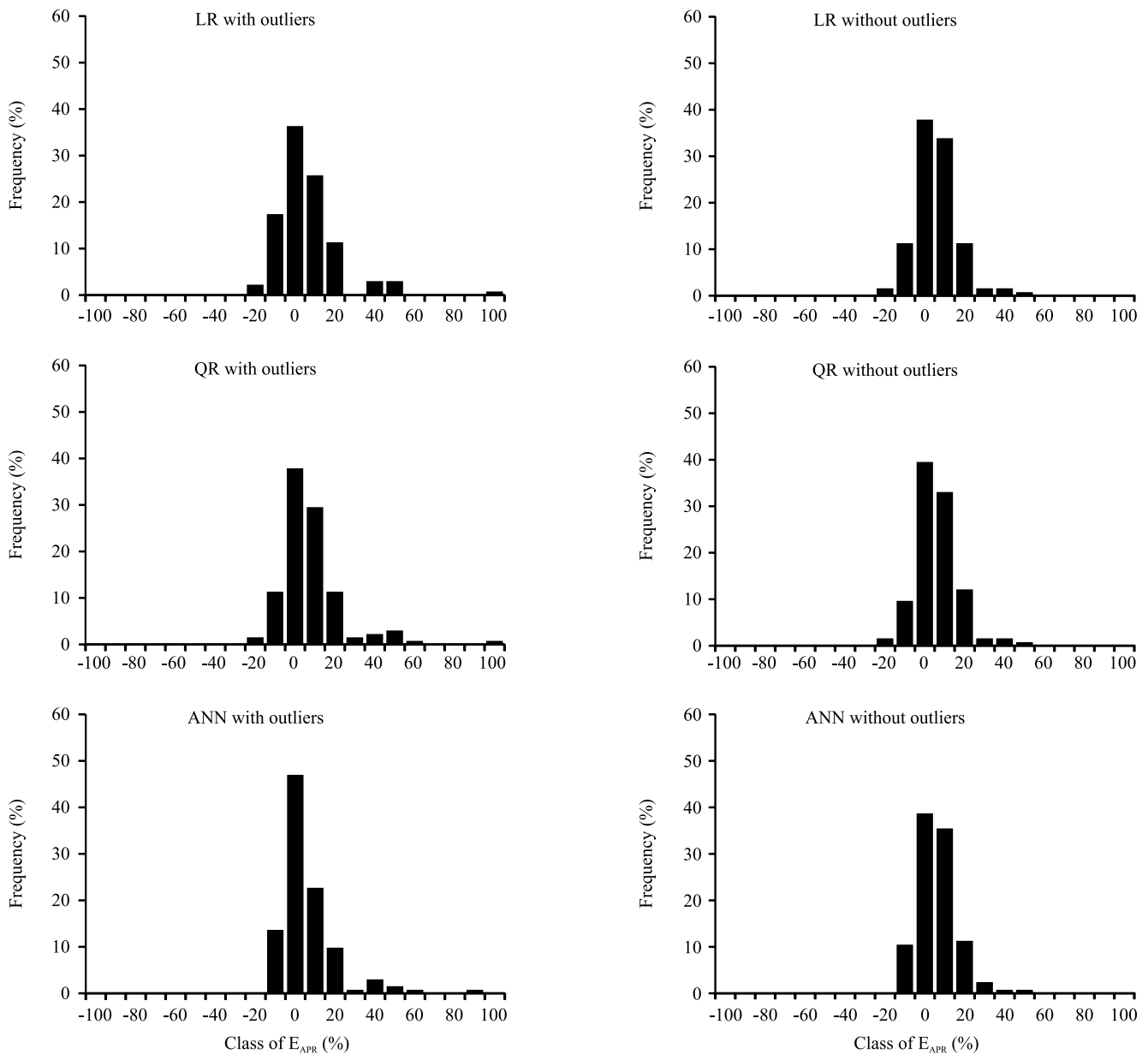


Figure 3. Histograms of relative percentage error (E_{APR}) for dominant height estimates according to eucalyptus (*Eucalyptus* spp.) plant age, considering different methods and the presence or absence of outliers in the database. LR, linear regression; QR, quantile regression; and ANN, artificial neural network.

The analysis of site classification stability showed that the ANN provided good results, both for databases with and without outliers, especially when six or five measurements were considered (Figure 5). This reinforces the quality of the estimates obtained with this technique, mainly for the dominant height ranges, for which a lower stability is expected. The LR provided

good estimates when less than four measurements were considered. The worse results were observed for the QR (six measurements) when the outliers were excluded; however, when they were considered, this regression provided good results. When low amounts of measurements are available, the percentage of stable plots increases for all methods, which does not depend on the presence or absence of outliers. This pattern is commonly reported in similar studies (Chaves et al., 2016) and occurs because stability is greater in more advanced ages, represented here by the two last measurements.

Although better results were expected for the QR with the presence of outliers in the database, the opposite occurred when five and four measurements

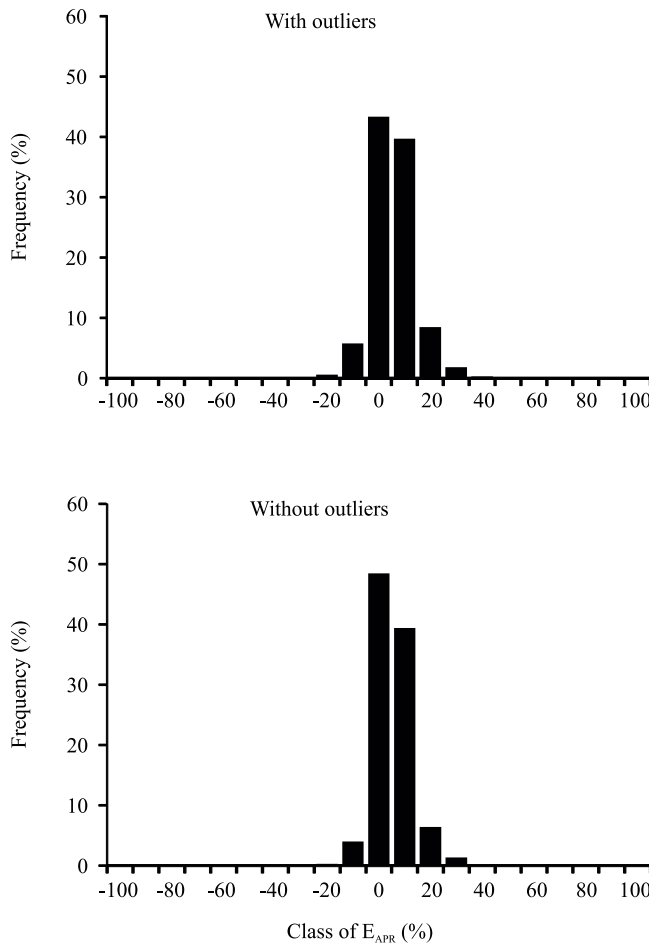


Figure 4. Histograms of relative percentage error (E_{APR}) for the artificial neural network estimates, used to classify the productive capacity of the eucalyptus (*Eucalyptus* spp.) site, with and without outliers in the database.

Table 5. Statistics of neural network trained to obtain the site index values for eucalyptus (*Eucalyptus* spp.) stands⁽¹⁾.

Outliers	E_{APR}	MAE	RMSE	$r_{y\hat{y}}$
With	0.4678	1.4336	1.9099	0.8942
Without	0.3131	1.2623	1.7327	0.8840

⁽¹⁾ E_{APR} , average percentage relative error; MAE, mean absolute error; RMSE, root mean square error; and $r_{y\hat{y}}$, correlation coefficient between observed and estimated values.

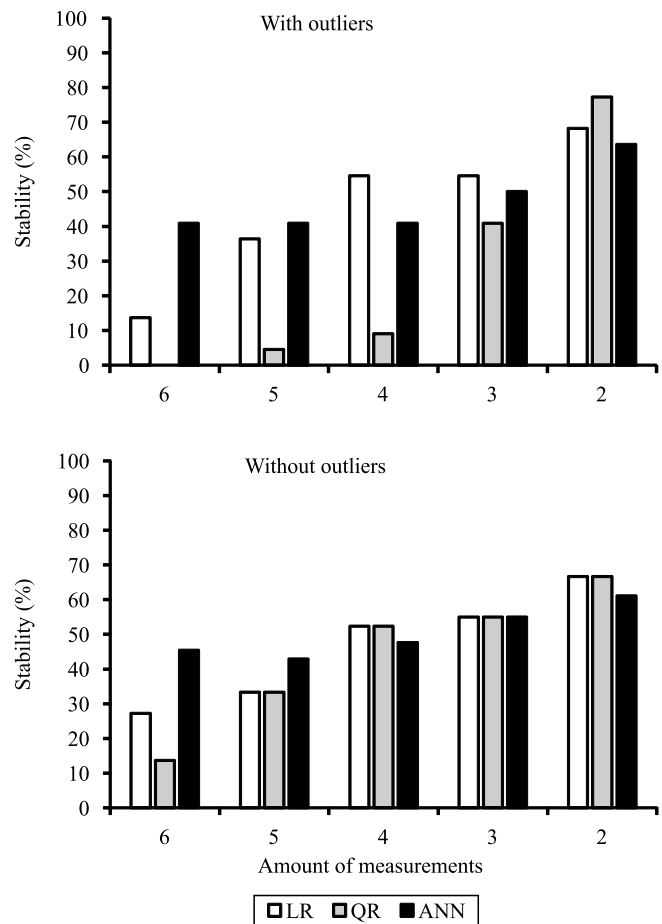


Figure 5. Percentage of stable plots according to the considered amount of recent measurements of eucalyptus (*Eucalyptus* spp.) stands, in databases with and without outliers. LR, linear regression; QR, quantile regression; and ANN, artificial neural network.

were considered. These discrepancies were caused by one plot, in each case. In the first one, the site index estimates for the QR were 30.27 m for class 32.5 m, and, for the LR, they were 29.95 m for class 27.5 m; in the second, for the QR, they were 30.10 m for class 32.5 m and, for the LR, 29.96 m for class 27.5 m.

Conclusions

1. The artificial neural network (ANN) is a robust technique to cope with the presence of outliers in databases, and can be used for the classification of the productive capacity of even-aged eucalyptus (*Eucalyptus* spp.) stands.

2. A better stability in the classification of forest sites can be obtained using the ANN, both with the presence or absence of outliers.

References

- ABELLANAS, B.; ABELLANAS, M.; POMMERENING, A.; LODARES, D.; CUADROS, S. A forest simulation approach using weighted Voronoi diagrams. An application to Mediterranean fir *Abies pinsapo* Boiss stands. **Forest Systems**, v.25, p.1-16, 2016. DOI: <https://doi.org/10.5424/fs/2016252-08021>.
- ARAÚJO JÚNIOR, C.A.; SOARES, C.P.B.; LEITE, H.G. Curvas de índices de local em povoamentos de eucalipto obtidas por regressão quantílica. **Pesquisa Agropecuária Brasileira**, v.51, p.720-727, 2016. DOI: <https://doi.org/10.1590/S0100-204X2016000600003>.
- ASSMAN, E. **The principles of forest yield study**: studies in the organic production, structure, increment and yield of forest stands. Oxford: Pergamon Press, 1970. 506p.
- BINOTI, D.H.B.; BINOTI, M.L.M. da S.; LEITE, H.G. Aplicação da função hiperbólica na construção de curvas de índice de local. **Revista Árvore**, v.36, p.741-746, 2012. DOI: <https://doi.org/10.1590/S0100-67622012000400015>.
- BINOTI, D.H.B.; BINOTI, M.L.M. da S.; LEITE, H.G. Configuração de redes neurais artificiais para estimação do volume de árvores. **Ciência da Madeira**, v.5, p.58-67, 2014. DOI: <https://doi.org/10.12953/2177-6830.v05n01a06>.
- BINOTI, M.L.M. da S.; LEITE, H.G.; BINOTI, D.H.B.; GLERIANI, J.M. Prognose em nível de povoamento de clones de eucalipto empregando redes neurais artificiais. **Cerne**, v.21, p.97-105, 2015. DOI: <https://doi.org/10.1590/01047760201521011153>.
- CAMPOS, J.C.C.; LEITE, H.G. **Mensuração florestal**: perguntas e respostas. 5.ed. atual. e ampl. Viçosa: Ed. da UFV, 2017. 636p.
- CHAVES, A.G.S.; DRESCHER, R.; CALDEIRA, S.F.; MARTINEZ, D.T.; VENDRUSCOLO, D.G.S. Capacidade produtiva de *Tectona grandis* L.f no Sudoeste de Mato Grosso. **Scientia Forestalis**, v.44, p.415-424, 2016. DOI: <https://doi.org/10.18671/scifor.v44n110.14>.
- COSENZA, D.N.; LEITE, H.G.; MARCATTI, G.E.; BINOTI, D.H.B.; ALCÂNTARA, A.E.M. de; RODE, R. Classificação da capacidade produtiva de sítios florestais utilizando máquina de vetor de suporte e rede neural artificial. **Scientia Forestalis**, v.43, p.955-963, 2015. DOI: <https://doi.org/10.18671/scifor.v43n108.19>.
- DEMOLINARI, R.A.; SOARES, C.P.B.; LEITE, H.G.; SOUZA, A.L. de. Crescimento de plantios clonais de eucalipto não desbastados na região de Monte Dourado (PA). **Revista Árvore**, v.31, p.503-512, 2007. DOI: <https://doi.org/10.1590/S0100-67622007000300016>.
- KOENKER, R. **quantreg**: quantile regression. R package version 5.05. 2013. Available at: <http://CRAN.R-project.org/package=quantreg>. Accessed on: Jan. 15 2018.
- KOENKER, R.; BASSETT JR., G. Regression quantiles. **Econometrica**, v.46, p.33-50, 1978.
- MACHADO, S. do A.; SOUZA, R.F. de; JASKIU, E.; CAVALHEIRO, R. Construction of site curves for native *Mimosa scabrella* stands in the metropolitan region of Curitiba. **Cerne**, v.17, p.489-497, 2011. DOI: <https://doi.org/10.1590/S0104-77602011000400007>.
- MCCULLOCH, W.S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics**, v.5, p.115-133, 1943. DOI: <https://doi.org/10.1007/BF02478259>.
- MIGUEL, E.P.; REZENDE, A.V.; LEAL, F.A.; MATRICARDI, E.A.T.; VALE, A.T. do; PEREIRA, R.S. Redes neurais artificiais para a modelagem do volume de madeira e biomassa do cerradão com dados de satélite. **Pesquisa Agropecuária Brasileira**, v.50, p.829-839, 2015. DOI: <https://doi.org/10.1590/S0100-204X2015000900012>.
- R CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2014.
- SANTOS, A.J.A.; ANJOS, O.; PEREIRA, H. Estimation of *Acacia melanoxylon* unbleached Kraft pulp brightness by NIR spectroscopy. **Forest Systems**, v.24, eRC03, 2015. DOI: <https://doi.org/10.5424/fs/2015242-07580>.
- SCHWERTMAN, N.C.; OWENS, M.A.; ADNAN, R. A simple more general boxplot method for identifying outliers. **Computational Statistics & Data Analysis**, v.47, p.165-174, 2004. DOI: <https://doi.org/10.1016/j.csda.2003.10.012>.
- SCOLFORO, J.R.S. **Biometria florestal**: modelos de crescimento e produção florestal. Lavras: UFLA, 2006. 393p.
- SCOLFORO, J.R.S. **Biometria florestal**: módulo 3: métodos para classificação de sítios florestais. Lavras: UFLA, 1997. 151p.
- SOARES, C.P.B.; PAULA NETO, F. de; SOUZA, A.L. de. **Dendrometria e inventário florestal**. 2.ed. Viçosa: Ed. da UFV, 2011. 272p.