Height-Diameter Models for *Araucaria angustifolia* (Bertol.) Kuntze in Natural Forests

Emanuel Arnoni Costa[1], André Felipe Hess[2], Danieli Regina Klein[2] & César Augusto Guimarães Finger[1]

[1] Forest Science, Federal University of Santa Maria, Brazil

[2] Forest Science, State University of Santa Catarina, Brazil

Correspondence: Emanuel Arnoni Costa, Forest Science, Federal University of Santa Maria, Brazil. E-mail: emanuelarnonicost@hotmail.com

Abstract

The height-diameter relationship of *Araucaria angustifolia* trees in different sociological positions (dominant, codominant, dominated) was evaluated in a native forest in the south of Brazil, aiming to find accuracy in its estimation and its use as a component of forest description, growth and yield. The total number of trees of the three sociological positions was 657. Part of these trees of each sociological position was used to estimate the parameters of models, and the remaining for model evaluation. Thus, the objective of this work was to find the best height estimate using nonlinear models, linear with dummy variable, principal component with nonlinear regression, and principal component with mixed nonlinear regression. The criteria for accuracy of fit were adjusted coefficient of determination, root mean square error and mean error. The results showed that the fit using principal component with mixed nonlinear regression obtained consistent results and better accuracy. It showed that height growth capacity depends on the sociological position.

Keywords: forest measurement, statistics models, Parana-Pine, regression analysis

1. Introduction

The tree height-diameter relationship is an important component in yield estimation, stand description, and damage appraisals (Zhang, 1997). Total height (h), and diameter at breast height (1.3 m above ground)—including bark—are the two most essential variables in most forest inventories (Huang et al., 2000). However, it is difficult to measure these variables for all trees in a forest, and it is common to adjust hypsometric equations to estimate the height of non-measured trees.

The height-diameter curve is not a strong biological relationship in both native and planted forests, because in many cases, tree trunks present an unusable or deformed portion, which generates high diameter values and low height values, and the opposite is also true (Hess et al., 2014).

Growth and yield models are generally used to predict the temporal development of forest stands. Knowing the diameter at breast height (DBH) and total tree height is fundamental for both developing and applying many growth and yield models. DBH of a tree can be quickly, easily, and accurately measured, but the measurement of total tree height is relatively complex, time consuming, and expensive. Furthermore, tree, stand, and site conditions may prevent accurate height measurements on all trees measured for DBH as it may not be possible to unambiguously observe a given tree, or reach an appropriate vantage point. Therefore, with many permanent and temporary sample plot systems, DBH is conventionally measured for all trees sampled, but height is measured for only a sub-sample of trees selected across the range of diameters observed (Sharma & Parton, 2007).

The hypsometric curve will show a great variability in height for the same diameter when constructing a curve for different sites or ages. The hypsometric relation usually results in estimates with low coefficient of determination and high root mean square error due to the high variability in height for the same diameter class in stands or forests of old age (Machado et al., 1994).

The development of simple and accurate height–diameter models based on easily obtainable tree and stand characteristics is a common precursor to the use of inventory and sample plot data to calculate volume and other stand attributes. However, the relation between the diameter of a tree and its height varies among stands (Calama & Montero, 2004) and depends on the growing environment and stand conditions (Sharma & Zhang, 2004).

Thus, it is a challenging task to choose a mathematical technique of fit and a model that can accurately describe the variability of height-diameter relationships for *Araucaria angustifolia* (Bertol) Kuntze trees in a native forest. An accurate fit of this variable is necessary to estimate forest volumes and describe the forests and their development over time (Huang et al., 2000).

Some functions may produce significant errors by extrapolation when applied beyond the range of the model development data. This is aggravated when the sampled trees are from young or middle-aged stands. Therefore, the models' predictive capabilities (accuracy, precision, time dependence, biological realism, and flexibility) must be carefully evaluated and validated before a long-term forest growth and yield simulator is used (Zhang, 1997).

The objective of this work was to find the best equation to describe the height of individual *Araucaria angustifolia* trees of dominant, codominant, and dominated sociological positions, using different techniques of mathematical adjustment of models, and based on precision statistics, indicate which model describes the height-diameter relationships with more precision.

2. Material and Methods

2.1 Study Area

The study was conducted in a rural property in Lages, in Santa Catarina, Brazil (27°48′S; 50°19′W, and altitude of 987 m) with *A. angustifolia* trees in competition in a natural forest. The local climate is humid subtropical, without dry season, with temperate summer (Cfb), according to the Köppen classification. The local average annual temperature is 15.2 °C, and average annual precipitation is 1685.7 m (Alvares et al., 2013). The predominant soils in the region are Haplic Nitosol (Oxisol), and Humic Cambisol (Inceptisol), developed from basaltic rocks (Embrapa, 1999).

2.2 Data Collection

Data of 286 *A. angustifolia* trees from the 657 trees collected in a natural forest were adapted to hypsometric models, as used by Costa et al. (2014). Statistical techniques that would improve the fit and precision of existing height-diameter models were applied, aiming to validate the old model, as well as a new model using 259 trees that were collected and used for data validation. External data from other sites were used to evaluate the performance of the height-diameter model developed in the validation step using data of 112 trees collected in a natural forest by Klein (2017). This procedure was carried out to ascertain the generalization capacity of the model in terms of height estimation in other places nearby the study area (Figure 1).
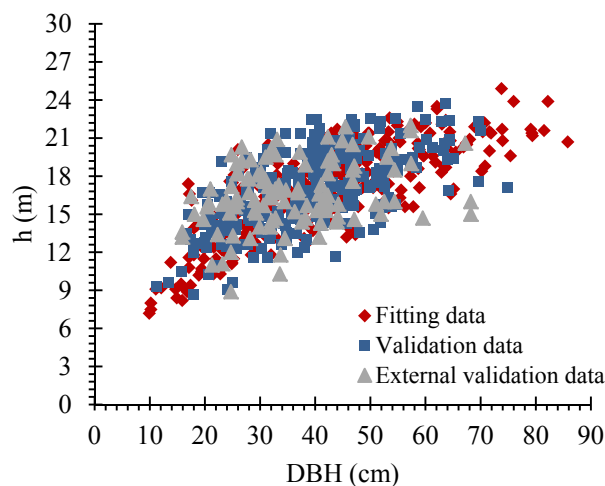

Figure 1. Height-diameter curve for *A. angustifolia* trees in a natural forest

The diameter at breast height (DBH) of each tree was measured with a diametrical tape, the height (h) and the height to crown base (HCB) were measured with the Vertex IV hypsometer and classified according to their sociological position in the forest. Different social positions and number of trees were used for fitting height-diameter models (SP_1: dominant, N = 186; SP_2: codominant, N = 50; SP_3: dominated, N = 50), validation (SP_1: dominant, N = 125; SP_2: codominant, N = 64; SP_3: dominated, N = 70), and validation using external data

(SP$_1$: dominant, N = 53; SP$_2$: codominant, N = 41; SP$_3$: dominated, N = 18). The data showed large variation in diameters and heights among the sociological positions (Table 1).

Table 1. Summary statistics for model fitting and validation data

Variable	Social position	Fitting data*		Validation data		External validation data[#]	
		Mean±SD	Range	Mean±SD	Range	Mean±SD	Range
DBH (cm)	Dominant	47.8±15.3	14.7-85.8	46.4±11.2	13.4-74.8	42.2±11.1	23.2-68.2
h (m)		18.1±3.3	8.4-24.9	18.4±2.7	9.6-23.7	17.2±2.7	8.9-22.0
HCB (m)		12.8±4.1	3.0-19.1	13.2±3.0	4.7-18.8	9.3±3.0	3.2-13.1
HCB/DBH		27.6±8.8	9.7-58.9	29.4±7.7	12.4-57.2	23.5±10.1	8.0-49.6
DBH (cm)	Codominant	33.8±12.6	9.9-58.9	36.7±9.2	15.7-64.3	33.4±8.7	18.1-57.4
h (m)		15.5±3.7	7.2-21.2	16.7±3.2	9.6-22.6	16.7±2.4	10.3-21.9
HCB (m)		11.3±4.1	3.6-17.8	12.4±3.5	3.6-18.5	10.3±2.4	4.6-14.4
HCB/DBH		35.3±11.9	18.2-66.5	34.9±11.1	15.6-74.5	33.0±11.7	11.2-56.9
DBH (cm)	Dominated	29.1±10.2	12.1-49.5	27.5±7.3	11.2-43.4	24.4±5.3	15.9-32.1
h (m)		14.9±2.6	8.9-19.6	14.9±2.3	8.7-20.1	15.0±2.0	11.0-18.2
HCB (m)		11.2±2.9	4.4-15.9	11.7±2.1	6.2-16.4	9.8±2.9	3.6-16.8
HCB/DBH		41.5±15.0	18.6-87.1	44.9±11.4	24.4-68.8	41.6±14.4	16.9-69.1
DBH (cm)	All	42.2±16.2	9.9-85.8	38.9±12.6	11.2-74.8	36.1±11.4	15.9-68.2
h (m)		17.1±3.6	7.2-24.9	17.1±3.1	8.7-23.7	16.6±2.6	8.9-22.0
HCB (m)		12.2±4.0	3.0-19.1	12.6±3.0	3.6-18.8	9.8±2.8	3.2-16.8
HCB/DBH		31.3±11.9	9.7-87.1	34.9±11.6	12.4-74.5	29.9±13.2	8.0-69.1

Note. SD = standard deviation; h = height; HCB = height to crown base; DBH = diameter at breast height; HCB/DBH = HCB to DBH ratio. *Data from Costa et al. (2014). [#]Data from Klein (2017).

2.3 Models for Height-Diameter Relationship

2.3.1 Nonlinear [NL]

The Michailoff nonlinear model [NL] (Equation 1) was chosen because of its well explained data trend of height-diameter in *A. angustifolia* trees (Costa et al., 2016). The model was adjusted according to the groups of sociological position of trees (SP$_1$: dominant, SP$_2$: codominant, and SP$_3$: dominated) (Costa et al., 2014).

$$h = 1.3 + \beta_0 \cdot \exp(-\beta_1/DBH) \tag{1}$$

Where, h is total tree height (m), DBH is tree diameter at breast height (cm); β_0 and β_1 are the estimated parameters.

2.3.2 Linear With Dummy Variable [LDV]

The linear model with dummy variable [LDV] (Equation 2) applied in the present study was used by Costa et al. (2014) in a natural forest to stratify the sociological position of trees.

$$\ln(h - 1.3) = 3.2334 - 18.3308 \cdot (1/DBH) - 0.1190 \cdot D_1 - 0.2755 \cdot D_2$$
$$+ 4.4283 \cdot [(1/DBH) \cdot D_1] + 8.5324 \cdot [(1/DBH) \cdot D_2] \tag{2}$$

Where, h is total tree height (m), DBH is tree diameter at breast height (cm), D_1 and D_2 are dummy variables to stratify trees by the codominant and dominated social positions, respectively. The equation has R^2adj. 69.86% and root mean square error (RMSE) 1.9567.

2.3.3 Principal Component With Nonlinear Regression [PC-NLR]

The principal component with nonlinear regression [PC-NLR] (Equation 5) was developed in three stages. The first stage was the optimization of the value of coefficient η_0 (Equation 3), that typifies the level (β_0) of Equation 1, fitted according to the sociological position.

$$\eta_0 = [(h - 1.3)/\exp(-\beta_1/DBH)] \tag{3}$$

Where, h is total tree height (m), DBH is tree diameter at breast height (cm), β_1 is the slope coefficients defined by sociological position (see Table 3, [NL] SP$_1$ = 17.7633; SP$_2$ = 11.6296; and SP$_3$ = 7.9829), and η_0 is the value obtained for each tree.

The second stage was the determination of the linear regression with principal component (Equation 4, see Table 3, PC-NLR-coefficients of regression: $\alpha_0 = 23.1129$; $\alpha_1 = 1.9362$; $\alpha_2 = -0.6472$; $\alpha_3 = -3.7318$). The use of linear regression with principal component can be beneficial where modeling is used with quantitative and qualitative variables simultaneously. In addition, a multicollinearity problem can be solved using the principal component regression (Montgomery, 2006). The modeling of regression coefficients in h/d models (β_0-level and β_1-slope) were tested at plot level using multiple linear regression as a function of stand variables (Adame et al., 2008), being the complete model later adjusted. Similarly, in the present study the value obtained for each tree (η_0, according to the first stage in Equation 3) was used by modeling the principal component regression in function of variables at tree level. We evaluated the influence of variables: diameter at breast height (DBH-quantitative), height to crown base (HCB-quantitative) and social position (SP-qualitative) with regards to the level (η_0) of model.

$$\eta_0 = \alpha_0 + \alpha_1 . K_1 + \alpha_2 \cdot K_2 + \alpha_3 \cdot K_3 \tag{4}$$

For the third stage, we used the result of Equation 4 (η_0) obtained from trees in (Equation 5), which was symbolized by (δ_0). Thus, Equation 5 was fitted and new regression parameters were defined (*e.g.* see Table 3, PC-NLR-estimated coefficients: $\xi_0 = 1.1512$ and $\xi_1 = 9.4520$). This was designated principal component with nonlinear regression [PC-NLR]. Therefore, a unique model with biological realism of estimated coefficients of regression was obtained, since the linearization of a nonlinear model generates changes in its mathematical structure and adds bias.

$$h = 1.3 + (\delta_0/\xi_0) \cdot \exp[-(\xi_1/DBH)] \tag{5}$$

Where, K_1, K_2, K_3 are eigenvectors; *e.g.* see Table 3 in PC-NLR, $K_1 = 0.6700 \cdot [(DBH_i - Mean_{DBH})/SD_{DBH}] + 0.5642 \cdot [(HCB_i - Mean_{HCB})/SD_{HCB}] - 0.4827 \cdot [(SP_i - Mean_{SP})/SD_{SP}]$, inserting the values in the expression: $K_1 = 0.6700 \cdot [(DBH_i - 42.5815)/16.1674] + 0.5642 \cdot [(HCB_i - 12.2000)/3.9961] - 0.4827 \cdot [(SP_i - 1.5245)/0.7753]$, K_2 and K_3 followed the same procedure; SD = standard deviation; α_0, α_1, α_2, α_3 are the estimated coefficients of linear regression with principal component, $\alpha_0 = 23.1129$; $\alpha_1 = 1.9362$; $\alpha_2 = -0.6472$; $\alpha_3 = -3.7318$; h is total tree height (m); DBH is tree diameter at breast height (cm); HCB is the height to crown base (m); SP is sociological position of the tree in the forest; δ_0 is the result of the regression with principal component; and ξ_0 and ξ_1 are the estimated coefficients.

2.3.4 Principal Component With Mixed Nonlinear Regression [PC-MNLR]

In this model, we used the result of Equation 4 (η_0) obtained from trees, referring to item 2.3.3 in stages 1 and 2, according to (Equation 6), which was symbolized by (δ_0). Thus, Equation 6 was fitted and new regression parameters were defined (*e.g.* see Table 3, PC-MNLR-coefficients estimated: $\theta_0 = 1.0062$ and $\theta_1 = 12.3262$; where $SP_1 - u_{k1} = 5.1354$; $SP_2 - u_{k2} = -0.2625$; $SP_3 - u_{k3} = -4.8947$). This was designated principal component with mixed nonlinear regression [PC-MNLR]. The PC-MNLR was used attributing random coefficients to stratify the trees according to their sociological position because of its easy calibration. The introduction of random coefficients in slope model (θ_1) was tested. However, in the Plenterwald system, the height curve does not change because the social position of trees of certain diameter classes remains the same (Prodan, 1944). The height-diameter trend in *A. angustifolia* trees was investigated using the nonlinear model with inclusion of quantitative variables: diameter at breast height (DBH), height to crown base (HCB), and a qualitative variable represented by the sociological position of trees in the model. Since the height curve is very similar to the growth development curve, and the height curves of normal-management classes are similar to those of uneven-aged forests (Loetsch et al., 1973).

$$h = 1.3 + [(\delta_0/\theta_0)] \cdot \exp[-(\theta_1 + u_{ki})/DBH] \tag{6}$$

Where h is total tree height (m), DBH is tree diameter at breast height (cm), δ_0 is the value of regression with principal component, θ_0 and θ_1 are the estimated coefficients of fixed effects, and u_{ki} is the estimated coefficient of random effects that specify the sociological position.

2.4 Criteria for Height-Diameter Model Evaluation

The adjusted coefficient of determination (R^2adj.), root mean square error (RMSE), and mean error (e_{mean}) were evaluated (Table 2). The graphical residual distribution (%) as a function of DBH was assessed. The accuracy of the estimation models for the data reserved for validation was based on the t-test ($\alpha = 0.05$). All statistics were processed with the Statistical Analysis System (SAS-9.2) (SAS Institute 2011).

Table 2. Criteria used to evaluate the height-diameter estimation models

Criterion	Expression
R^2adj. =	$1 - \left[\dfrac{\sum_{i=1}^{n}(y - \hat{y})^2}{\sum_{i=1}^{n}(y - \bar{y})^2}\right] \cdot \left[\dfrac{n-1}{n-p}\right]$
RMSE =	$\sqrt{\left[\dfrac{\sum_{i=1}^{n}(y - \hat{y})^2}{n-p}\right]}$
e_{mean} =	$\dfrac{y - \hat{y}}{n}$

Note. y_i = observed value for the ith observation; $\hat{y}_i$ = predicted value for the ith observation; $\bar{y}_i$ = mean of the y_i; n = number of observations in the dataset; p = number of estimated parameters.

3. Results

The h and HCB variability among the sociological positions and their means are shown in Table 1. This variability indicates the importance of separating the data for fitting to the hypsometric function (Costa et al., 2014), since the mean for the set of trees is close to the mean of dominant trees. However, for the other sociological positions, the variable will be overestimated when using the mean of the set of trees. Differences in h and HCB between sociological positions showed a standard deviation ±3 meters, denoting differences in height growth, amount of light entering the canopy, density and competition of trees in the forest.

The models were adjusted for each sociological position (NL-Equation 1), the NL model tended to explain less the variability in the codominant and dominated sociological positions, and had greater RMSE (less precision), while the adjustment explained 71.0% of the variance of the dominant position. These results denote the difference in height growth associated with competition for light and ability to compete, thus indicating a great difference in the height-diameter ratio. The height growth of trees in the dominant position stabilized and reached the same level in the canopy (similar heights), presenting better fit and precision in the estimation of the variable. The adjustment of the NL model was satisfactory regarding the accuracy with the validation data (Table 3) according to sociological position, showing that estimates did not differ from those observed by the t-test (SP_1: p = 0.378635[ns]; SP_2: p = 0.368907[ns] and SP_3: p = 0.754448[ns]).

The model with Dummy variable (LDV-Equation 2) explained 70.0% of the variance in data used for fitting and 42.0% in validation data (*e.g.* see in Table 3), but its p-value was significant by the t-test (SP_1, SP_2 and SP_3: p < 0.011326), showing that the predicted values differed from those observed, confirming the results in Figures 2 and 3, what evidences the low performance of the model. The use of a linear model with categorical dummy variables to stratify the sociological positions (Costa et al., 2014) showed the variance inflation factor (FIV) higher than 5, indicating that the model presents collinearity for the estimators (Montgomery et al., 2006). It is worth noting that the multicollinearity problem can be solved by other statistical methods such as the use of principal component analysis, nonlinear model, ridge regression, among others (Hoerl & Kennard, 1970; Montgomery, 2006).

The effect of social position on the intercession and slope of the regression curves using covariance analysis, based on the Dummy variable (LDV) in Table 3, and according to the study by Costa et al. (2014), showed significant values for the variables D1 (codominant) (p = 0.0256) and D2 (dominated) (p < 0.0001), indicating differences between levels of codominant and dominated sociological positions in relation to dominant trees. The adjustment of the model with Dummy variables must have a restricted use due to the occurrence of collinearity. In this sense, observing the mean error for the individual models, it is preferable to use the NL model to estimate height instead of models with Dummy variable (LDV).

The use of principal component analysis in models of regression can be beneficial where modeling is used with quantitative and qualitative variables simultaneously. In regression models with principal components, they are explanatory variables that are not correlated, a factor that eliminates the problem of multicollinearity (Draper & Smith, 1981). A high correlation of DBH and HCB was evidenced by the correlation matrix, indicating that the larger the diameter, the greater the height to crown base (Table 3). The negative SP value indicates that it decreases with increasing height (β_0-level in equation 4), *i.e.*, dominant (values of 1 for SP_1), codominant (values of 2 for SP_2) and dominated (values of 3 for SP_3) trees. DBH and HCB are the variables with greater weight, being more influential for the first principal component (PC1). In the same relation, for the second principal component (PC2), is SP. With PC1 (62.93%) and PC2 (27.92%), it is possible to explain 90.85% of the total

variance of data. The fitting by linear regression with principal components in Equation 4 [see Table 3, PC-NLR coefficients of regression: (α_0 = 23.1129; α_1 = 1.9362; α_2 = -0.6472; α_3 = -3.7318)] showed values of R²adj. = 0.7488 and RMSE = 1.9392.

The inclusion of the quantitative variable HCB and the qualitative variable SP in the principal component with nonlinear regression model (PC-NLR-Equation 5) showed a similar fitting performance to that of the Dummy variable model (LDV) in Table 3. The PC-NLR explained 68.0% of variance of data used for fitting and 52.0% of validation data (*e.g.* see Table 3),but its p-value was significant by the t-test (p < 0.000042), evidencing low accuracy in estimates, with the predicted values different from observed.

The best estimates were obtained with the adjustment of the principal component with mixed nonlinear regression (PC-MNLR-Equation 6), which explained 86.0% of variance using data of model adjustment and 73.0% with validation data (Table 3). According to the t-test, estimates did not differ from those observed in the validation data (p = 0.658524^{ns}). The graphical analysis of residual distribution in the PC-MNLR model was better than that of LDV, these results can be observed in the data used for the fitting (Figure 2) and in the data used for the validation (Figure 3). The PC-MNLR model had lower RMSE and homogeneous residual distribution for all sociological positions, concentrating its distribution around 20.0%, and reducing discrepant points.

The results showed that the LDV model is biased and the PC-MNLR model is the most suitable for estimating a height of araucaria trees for this forest typology. The estimate values using the equation of PC-MNLR with external data show no differences from those values observed and predicted by the t-test (p = 0.95886) and evidences high accuracy of equation in the prediction of total height of araucaria trees (Figure 4). Therefore, the inclusion of the quantitative and qualitative variables promoted the generation of uncorrelated coefficients (multicollinearity) using the principal component, the PC-MNLR model has good generalization capacity with better accuracy to describe data. When the independent variables necessary to use the PC-MNLR are not measured, it is recommend using the individual NL models. These do not show significant differences by the t-test between observed and estimated data for the positions: SP1, SP2 and SP3.
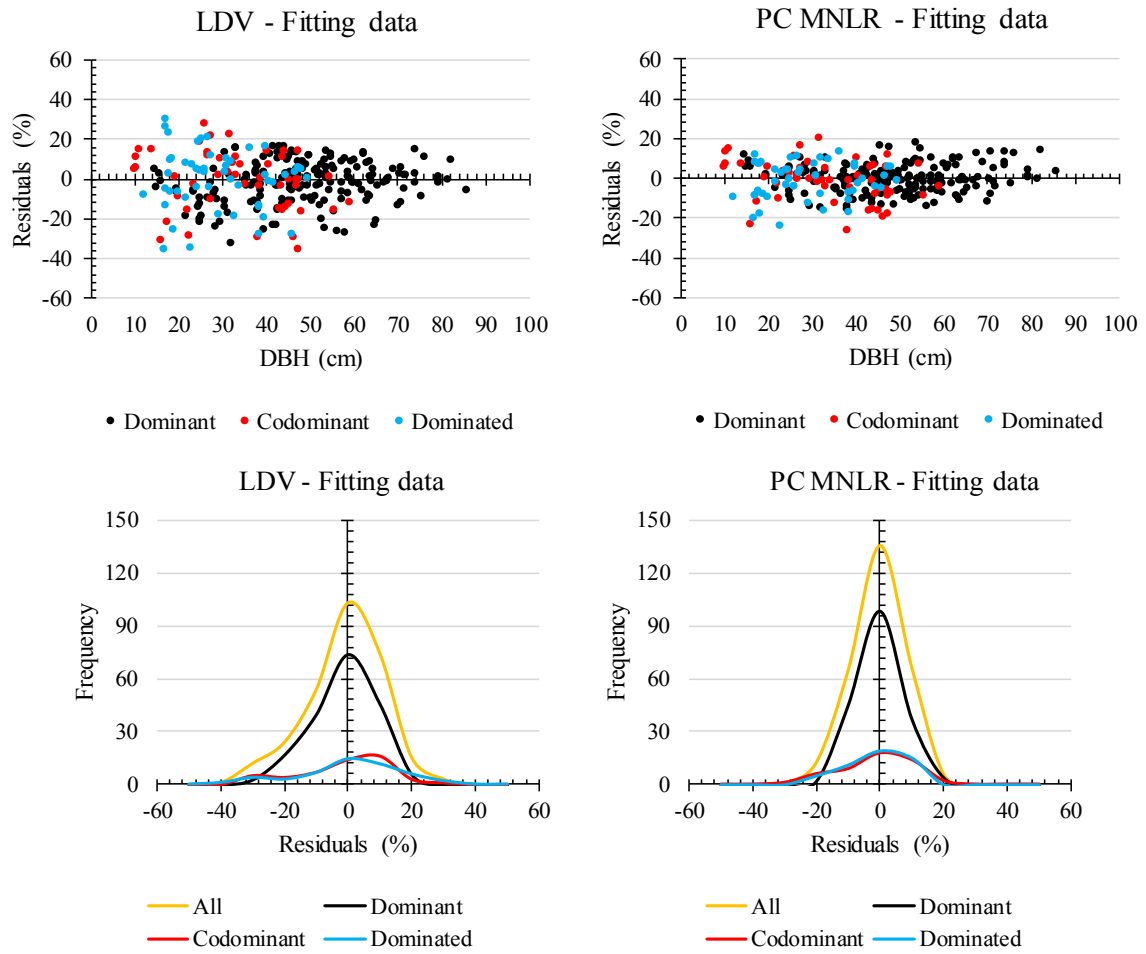
Figure 2. Residual distribution of the best-fit equations for the total height data of *Araucaria angustifolia* trees

Table 3. Performance of different models and fitting techniques to estimate the height-diameter relationship of *Araucaria angustifolia* trees

Type	Social position	CL[b]	Coefficients		Statistics fitting			Statistics validation			
			β_0	β_1	R^2adj.	RMSE	e_{mean}	R^2adj.	RMSE	e_{mean}	t test (p-value)
Nonlinear[a] [NL]	SP$_1$		25.3155	17.7633	0.7121	1.7950	-0.1986	0.3630	2.1753	0.1707	0.378635 [ns]
	SP$_2$		21.1504	11.6296	0.6187	2.2837	-0.0112	0.1963	2.8747	0.3205	0.368907 [ns]
	SP$_3$		18.4356	7.9829	0.3136	2.1516	-0.0019	0.2115	2.0565	0.0766	0.754448 [ns]
Linear with Dummy variable[a] [LDV]	SP$_1$, SP$_2$, SP$_3$		3.2334	-18.3308	0.6986	1.9567	0.0642	0.4244	2.3667	0.3669	0.011326 [*]
		D1	-0.1190	4.4283							
		D2	-0.2755	8.5324							

Principal component with nonlinear regression [PC-NLR]

Statistics fitting: R^2adj. 0.6821, RMSE 2.0093, e_{mean} 0.1140
Statistics validation: R^2adj. 0.5214, RMSE 2.1581, e_{mean} 0.5358, t test (p-value) 0.000042[*]

Simple statistics

	DBH	HCB	SP
Mean	42.5815	12.2000	1.5245
SD	16.1674	3.9961	0.7753

Correlation matrix

	DBH	HCB	SP
DBH	1.0000	0.6273	-0.4990
HCB	0.6273	1.0000	-0.1670
SP	-0.4990	-0.1670	1.0000

Eigenvectors

	K_1	K_2	K_3
DBH	0.6700	0.0510	0.7407
HCB	0.5642	0.6136	-0.5525
SP	-0.4827	0.7880	0.3823

Coefficients of regression

α_0	α_1	α_2	α_3
23.1129	1.9362	-0.6472	-3.7318
(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
ξ_0	ξ_1		
1.1512	9.4520		
(<0.0001)	(<0.0001)		

Principal component with mixed nonlinear regression [PC-MNLR]

Statistics fitting: R^2adj. 0.8586, RMSE 1.3403, e_{mean} -0.0036
Statistics validation: R^2adj. 0.7316, RMSE 1.6162, e_{mean} -0.0439, t test (p-value) 0.658524 [ns]

	Parameters	Estimate	Pr > |t|
Fixed effects	θ_0	1.0062	(0.0002)
	θ_1	12.3262	(0.0369)
Variance components	var (μ_1)	16.9001	(0.3485)
	var (ε)	1.7710	(0.0070)
Random effects	SP$_1$ – u$_{k1}$	5.1354	(0.1650)
	SP$_2$ – u$_{k2}$	-0.2625	(0.9226)
	SP$_3$ – u$_{k3}$	-4.8947	(0.1776)

Note. [a] Coefficients estimated and defined as described by Costa et al. (2014); [b] definition of *Dummy* variable, D$_1$ = codominant trees; D$_2$ = dominated trees. [ns] non-significant (α = 5%); SP = sociological position of tree in the forest; SD = standard deviation.
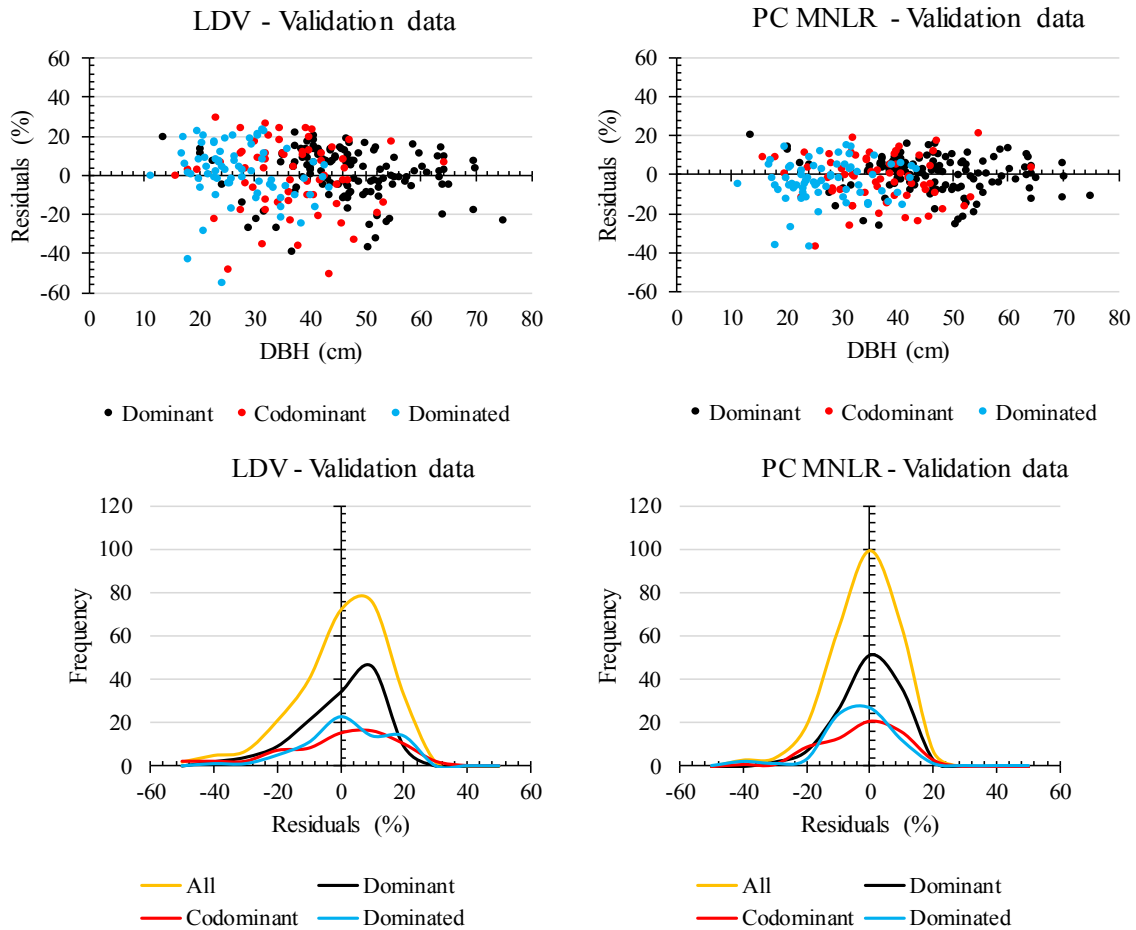
Figure 3. Residual distribution of the best-fit equations using the validation data for the variable of total height of *Araucaria angustifolia* trees

Example:

DBH	HCB	SP	K_1	K_2	K_3	u_{ki}	h
37.2	11.3	1	0.8293	-0.3911	-0.5918	5.1354	18.2
33.1	10.4	2	-0.2487	0.4641	-0.1593	-0.2625	17.1
29.6	11.5	3	-0.8875	1.7615	-0.0837	-4.8947	17.2

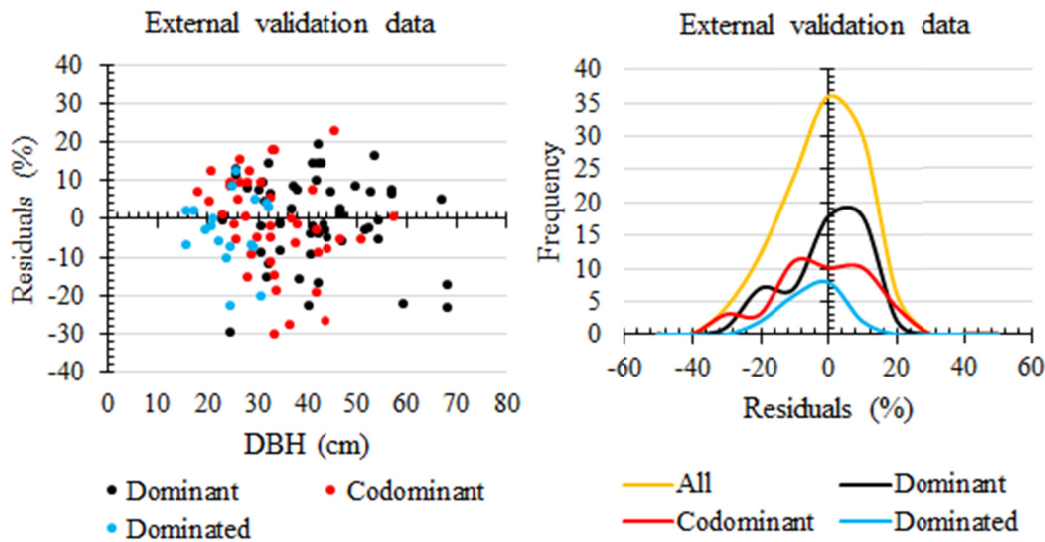| Statistics using external validation data	
R^2adj.	0.4587
RMSE	1.9086
e_{mean}	-0.0091
t test (p-value)	0.95886^{ns}

Figure 4. Statistics, example and residual distribution of the best-fit equations using external validation data to describe total height of *Araucaria angustifolia* trees with the PC-MNLR model

A practical example will be demonstrated using the fitted PC-MNLR (Equation 6), whose values are: $\delta_0 = 23.1129 + 1.9362 \cdot K_1 - 0.6472 \cdot K_2 - 3.7318 \cdot K_3$, $\theta_0 = 1.0062$ and $\theta_1 = 12.3262$; the rewriting of the model resulted in h = 1.3 + [(23.1129 + 1.9362·K_1 − 0.6472·K_2 − 3.7318·K_3)/1.0062]·exp[-(12.3262 + u_{ki})/DBH]. The calculation is: $K_1 = 0.6700 \cdot [(DBH_i − 36.0866)/11.4459] + 0.5642.[(HCB_i − 9.7509)/2.7923] − 0.4827 \cdot [(SP_i − 1.6875)/0.7356]$, with eigenvectors obtained in Table 3 (0.6700, 0.5642, -0.4827; 0.0510, 0.6136, 0.7880; 0.7407, -0.5525, 0.3823) and values ($Mean_{DBH} = 36.0866$; $SD_{DBH} = 11.4459$; $Mean_{HCB} = 9.7509$; $SD_{HCB} = 2.7923$; $Mean_{SP} = 1.6875$; $SD_{SP} = 0.7356$) obtained for external validation data (Table 1). The calculation of K_2 and K_3 followed the same procedure. Thus, random effects according to the social position are given by SP_1 (Dominant) − $u_{k1} = 5.1354$; SP_2 (Codominant) − $u_{k2} = -0.2625$; SP_3 (Dominated) − $u_{k3} = -4.8947$. With the values obtained from the dominant tree (SP = 1) in Figure 4 and using Equation 7, it is possible to estimate total tree height in 18.2 meters. The same procedure must be performed to estimate other heights of trees from different social positions.

$$h = 1.3 + [(23.1129 + 1.9362 \cdot (0.8293) − 0.6472 \cdot (-0.3911)$$
$$− 3.7318 \cdot (-0.5918))/1.0062] \cdot \exp[-(12.3262 + 5.1354)/37.2] \quad (h \approx 18.2 \text{ m}) \quad (7)$$

Where h is total tree height (m), DBH is tree diameter at breast height (cm), δ_0 is the result of the regression with principal component, θ_0 and θ_1 are the estimated coefficients of fixed effects, and u_{ki} are the estimated coefficients of random effects that specify each sociological position (SP).

4. Discussion

This study compared different fitting techniques of regression models aiming to improve the height-diameter relationship estimation for different sociological positions of *A. angustifolia* in a natural forest. The PC-MNLR (Equation 6) predicted the height of a tree in the forest requiring the measurement of DBH, HCB and their classification by SP with high accuracy of estimates (Table 3).

The variable HCB represents crown recession (Russell et al., 2014), which is an easier measurement variable depending on the density of the forest. Thus, this model has improved the height-diameter relationship to define the vertical and horizontal structure of the forest, which shows a different development according to age, density, sociological position, competition and site.

Several factors affect the tree height-diameter relationship, especially competition, age, and site (Adame et al., 2008). High-accuracy models for tree height estimates include variables that characterize the forest stand such as age, mean square diameter, basal area, density, and dominant height.

At early ages, codominant and dominated trees of *A. angustifolia* invest more in height growth due to the lack of light in the crown and, consequently, they present smaller increment in diameter. As they age, height growth stabilizes, causing a decrease in their vigor, which is more evident in codominant and dominated individuals (Costa et al., 2014).

The PC-MNLR change in height for a given diameter level is not the same for all sociological positions and height to crown base. Based on this evaluation, the height-diameter relationship is dependent on tree sociological positions in the forest. Therefore, estimating the height of trees using a simple height-diameter model for data set will lead to errors.

The determination of fixed and random effects in a model is a flexible decision, subject to debate, and all parameters in the model must be first considered mixed when the convergence is possible (Sharma and Parton, 2007). According to Calama and Montero (2004), predicting random parameters for a plot using complementary observations of height increases the predictive ability of the model for all trees. Fang and Bailey (2001) suggest that the parameters with high variability and less overlapping in confidence intervals, obtained by fitting each individual plot separately, must be considered mixed if the convergence is not achieved when considering all the parameters as mixed.

In mature forests, competition among trees creates different extracts, sociological positions, and variation in diameter and height, allowing a more stable hypsometric curve over time, by associating different diameters for the same height, since the height-diameter relationship can be affected by external factors such as density and sociological position (Machado et al., 2008).

The low biological relationship of height and diameter is evident due to height variability depending on the sociological position, which generates incoherent fitting of equations depending on forest typology, site conditions, and forest density; thus, testing different models and regression techniques is necessary to increase the accuracy in estimating this variable.

PC-MNLR is flexible and can be used in other sites, and locations with the same forest typology and species. This was observed in the residual distribution of the model, since they are more centralized near the zero (*e.g.* see Figures 2, 3 and 4). It is important when we use the total tree height variable for production planning and forest management, as it assigns greater precision to volume estimates.

In general, the model shows biological coherence and might help reduce prediction bias, particularly at the extremes of any dataset (Rijal et al., 2012). The PC-MNLR model outperformed LDV in all sociological positions, especially in the codominant and dominated classes, since the greatest variation for a given height was due to the variability among sociological positions.

5. Conclusion

The data showed the relation between dendrometrics and categorical variables. Categorical variables were used in the model to describe the social position of trees and their variation in height depending on their increase in diameter.

The proposed PC-MNLR model showed better accuracy for height estimates according to the sociological position of the trees. The use of random effects in the sociological position had easy application of the model in the field. The model showed good generalization in predicting tree heights using external data, confirming its potential for this purpose.

There is a need for silvicultural interventions in trees that grow at low diameter rates, and density regulation of trees in forests with selection of promising trees of better social position and with well-formed and vigorous crowns.

References

Adame, P., Del Río, M., & Cañellas, I. (2008). A mixed nonlinear height-diameter model for pyrenean oak (*Quercus pyrenaica* Willd.). *Forest Ecology and Management, 256*, 88-98. https://doi.org/10.1016/j.foreco. 2008.04.006

Alvares, C. A., Stape, J. L., Sentelhas, P. C., Gonçalves, J. L. M., & Sparovek, G. (2013). Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift, 22*, 711-728. https://doi.org/10.1127/0941-2948/ 2013/0507

Calama, R., & Montero, G. (2004). Interregional nonlinear height-diameter model with random coefficients for stone pine in Spain. *Canadian Journal of Forest Research, 34*, 150-163. https://doi.org/10.1139/x03-199

Costa, E. A., Finger, C. A. G., & Cunha, T. A. (2014). Influência da posição sociológica na relação hipsométrica de *Araucaria angustifolia*. *Revista Brasileira de Ciências Agrárias, 9*, 110-116. https://doi.org/10.5039/ agraria.v9i1a3201

Costa, E. A., Schroder, T., & Finger, C. A. G. (2016). Height-Diameter Relationships for *Araucaria angustifolia* (Bertol.) Kuntze in Southern Brazil. *Revista Cerne, 22*, 493-500. https://doi.org/10.1590/010477602016 22042182

Draper, N., & Smith, H. (1981). *Applied regression analysis* (2nd ed.). New York: John Wiley & Sons.

Embrapa. (1999). *Sistema brasileiro de classificação de solos*. Brasília: Embrapa Produção de Informação. Rio de Janeiro: Embrapa Solos.

Fang, Z. X. & Bailey, R. L. (2001). Nonlinear mixed effects modeling for slash pine dominant height growth following intensive silvicultural treatments. *Forest Science, 47*, 287-300.

Hess, A. F., Braz, E. M., Thaines, F., & Matto, P. P. (2014). Adjustment of the hypsometric relationship for species of Amazon Forest. *Ambiência Guarapuava, 10*, 21-29. https://doi.org/10.5935/ambiencia.2014. 01.02

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for no orthogonal problems. *Technometrics, 12*, 55-68.

Huang, S., Price, D., & Titus, S. J. (2000). Development of ecoregion-based height–diameter models for white spruce in boreal forests. *Forest Ecology and Management, 129*, 125-141. https://doi.org/10.1016/S0378-1127(99)00151-6

Klein, D. R. (2017). *Morfometria e crescimento diamétrico de Araucaria angustifolia (Bertol.) Kuntze no Planalto Sul Catarinense* (Dissertação, Universidade do Estado de Santa Catarina, Lages).

Loetsch, F., Zohrer, F., & Haller, K. (1973). *Forest inventory* (V. 2). Frankfurt: BLV.

Machado, S. A., Bailey, R. L., Basso, S. F., & Bevilacqua Júnior, V. G. (1994). Análise do comportamento da relação hipsométrica com respeito a idade para plantações de *Pinus elliotti* no Estado do Paraná. *Revista Cerne, 1*, 5-12.

Machado, S. A., Nascimento, R. G. M., Augustynczik, A. L. D., Silva, L. C. R., Figura, M. A., … Téo, S. J. (2008). Comportamento da relação hipsométrica de *Araucaria angustifolia* no capão da Engenharia Florestal da UFPR. *Pesquisa Florestal Brasileira, 56*, 5-15. Retrieved from http://pfb.cnpf.embrapa.br/pfb/ index.php/pfb/article/view/58

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2006). *Introduction to linear regression analysis* (4th ed.). New York: John Wiley & Sons.

Prodan, M. (1944). *Zuwachs-und Ertragsuntersuchungen in Plenterwald* (PhD Thesis, University of Freiburg, Freiburg).

Rijal, B., Weiskittel, A. R., & Kershaw, J. A. (2012). Development of regional height to diameter equations for 15 tree species in the North American Acadian Region. *Forestry, 85*, 379-390. https://doi.org/10.1093/ forestry/cps036

Russel, M. B., & Weiskittel, A. R. (2014). Comparing strategies for modeling individual-tree height and heigth-to-crown base increment in mixed-species Acadian forests of northeastern North America. *Europan Journal of Forest Research, 133*, 1121-1135. https://doi.org/10.1007/s10342-014-0827-1

SAS Institute. (2011). *The SAS system for Windows* (Release 9.2). SAS Inst., Cary, NC.

Sharma, M., & Parton, J. (2007). Height-diameter equations for boreal tree species in Ontario using a mixed-effects modeling approach. *Forest Ecology and Management, 249*, 187-198. https://doi.org/10.1016/j.foreco.2007.05.006

Sharma, M., & Zhang, S. Y. (2004). Height-diameter models using stand characteristics for *Pinus banksiana* and *Picea mariana*. *Scand. J. For. Res., 19*, 442-451. https://doi.org/10.1080/02827580410030163

Zhang, L. (1997). Cross-validation of nonlinear growth functions for modeling tree height–diameter relationships. *Annals of Botany, 79*, 251-257. https://doi.org/10.1006/anbo.1996.0334

Copyrights