Nonlinear mixed model approaches to estimating merchantable bole volume for *Pinus occidentalis Sw.*

Santiago W. Bueno-López* Eddie Bevilacqua[b]

*Corresponding author: [a] Pontificia Universidad Católica Madre y Maestra.
Autopista Duarte Km 1 ½, Santiago de los Caballeros, Republica Dominicana.
swbueno@syr.edu , swbueno@gmail.com
[b] State University of New York, College of Environmental Science and Forestry.
1 Forestry Drive, 320 Bray Hall, Syracuse, N.Y. 13210.

Abstract

The ability to predict cumulative bole volume to any predefined upper stem diameter on a standing tree is essential for estimating current inventory levels and making informed decisions regarding the management of forest resources. Several types of mathematical models have been developed to predict cumulative bole volume, requiring only low cost and high accuracy tree measurements to be obtained. This paper reports on the development and comparison of a variable-exponent taper model and a volume-ratio model for estimating inside-bark cumulative bole volume to three predefined upper stem diameters using stem analysis data *Pinus occidentalis* Sw trees in La Sierra, Dominican Republic. Each sampled tree was measured at multiple points on the bole, making observations spatially correlated within the tree. Inference problems due to autocorrelation were addressed by using a nonlinear mixed-effects model fitted by restricted maximum likelihood (REML)and including two random parameter coefficients at tree level in each of the two models tested. . Using an independent validation data set, the variable-exponent taper model with two random parameters demonstrated better predictive ability as compared to the volume-ratio model when estimating cumulative bole volume to the three predefined upper stem diameters. The taper model allows flexible volume estimation for the population average as well for specific trees.

Key words: nonlinear modeling; mixed-effects models; restricted maximum likelihood; taper model; volume-ratio model; random coefficients; calibration; *Pinus occidentalis* Sw.

1 Introduction

Regression analysis was initiated into forestry research more than 70 years ago. One common use of applying ordinary least squares and non-linear least squares techniques was to fit a regression model to predict volume content in the stem of a standing tree (Gregoire and Schabenberger 1996b). The set of explanatory variables included in a volume model is relatively small and comprises tree dimensions that can be easily and accurately measured in the forest, i.e. stem diameter at breast height (D) and total tree or merchantable height (H).

Due to changes in utilization standards within the forest industry, many different bole-volume equations have been developed. Volume prediction to any merchantability limit has been accomplished in many ways, including: (1) a constrained volume equation for different merchantable limits that differed by a fixed amount (Bailey and Cutter 1970); (2) expressing the merchantable volume for a given set of merchantability limits as a proportion of total tree volume, often called the "volume ratio approach" (Avery and Burkhart 2002); and (3) the development of stem profile equations which predict upper stem diameters at any point along the tree bole and that can be integrated to give volume content within any portion of the tree bole (Kozak 2004). The first approach requires a cumbersome procedure. The last two approaches are more common.

Construction of taper and volume-ratio equations entails the compilation of longitudinal data on each tree, resulting in a lack of independence between observations and biased estimates of parameters if ordinary least squared techniques are used in constructing the models (Calama and Montero 2005). Observations from the same unit of analysis, the trees, are correlated (Gregoire and Schabenberger 1996b, Garber and Maguire 2002, Leites and Robinson 2004).

To properly analyze this type of data, the error structure inherent within the data must be considered in the modeling process. Even though ordinary least squares analysis provides unbiased parameter estimates in the presence of autocorrelation, it does not provide minimum variance estimators (Gregoire and Schabenberger 1996a, Kozak 1997, Garber and Maguire 2002). The violation of the assumption of independence, required by the least squares method, would make inferences and test of significance invalid (West et al. 1984).

In addition to being correlated, the data for the construction of taper and volume-ratio models is unbalanced. Each sample tree is measured by short sections and the desired volume (total or merchantable) is determined by summing up the volumes of all the sections in a tree. Often, the number of observations of cumulative bole volume will vary among trees, as short trees tend to have fewer sections than larger trees. The observed sample will be unbalanced by having unequal numbers of observations per subject (Gregoire and Schabenberger 1996a)

Two general methods have been proposed to fit continuous, unbalanced, multilevel longitudinal data (Garber and Maguire 2002). The first proposed by Gregoire and Schabenberger (1996b) incorporates random subject effects, while the second models the correlation structure directly (Littell et. al. 2006). The former employs nonlinear mixed effects models, inducing the correlations in the marginal distribution of observations from the same tree by random effects that vary across trees, to reduce the impact of autocorrelation. The second approach introduces an assumed correlation structure into a generalized least squares solution (Tassisa and Burkhart 1998).

Mixed-effects models, consisting of both fixed and random-effects parameters have the advantage to allow for modeling of the variance-covariance matrix of correlated data (Yang et. al. 2009). Using random effects to capture tree to tree variability allows for modeling the volume of the individual tree stem (subject specific) as well as the volume of the average tree (population specific) (Schabenberger and Pierce 2002). Jones (1990) pointed out that modeling the variance-covariance structure directly or through random parameters indirectly was often equally effective in accounting for the within-subject correlation.

Pinus occidentalis is an endemic pine species of Hispaniola. In Dominican Republic, is the single most abundant species with the ability to successfully grow and protect the acidic, shallow and infertile sandy soils which characterized the steeped slopes of the mountain ranges. In 1995, the species occupied approximately 302,500 hectares in the DR (Dobler et al. 1995), representing more than the 95% of the species' world-wide distribution. In terms of timber production, *P. occidentalis* is the most important In the DR, comprising approximately 95% of all timber harvested (Banco Central 2004). The most valuable wood product from *P. occidentalis* stands is its timber, which is used for construction and furniture manufacturing.

Other important wood products derived besides saw timber logs are poles and bars used in the tobacco industry.

Despite its economic and ecologic importance, this species has never been the subject of serious growth and yield studies, making it difficult to estimate current inventory levels and to account for the amount of volume harvested. Few volume equations are available for the species at a regional level (Gil and Cuevas 1985, Montalvo et al. 2000). Existing equations have been developed to estimate single tree volume from breast height diameter and total height, but they do not enable product classification and did not consider the inherent difficulties derived from high correlation among observations taken on sample trees, since the parameter estimation method employed was solely based on ordinary least squares.

The primary objectives of this study were to: (1) fit both a variable-exponent taper equation (Kozak 1988) and a total volume-ratio model, based on Burkhart (1977) using procedures developed by Gregoire and Schabenberger (1996a), to *P. occidentalis*, Sw. data from La Sierra, Dominican Republic using a mixed-effects modeling approach , (2) compare the accuracy and precision of each model based on evaluation criteria using an independent validation data set, and (3) recommend that best model for estimating merchantable volume for different utilization standards for this species.

2 Material and methods

2.1 Data set

The study area is a region of approximately 1,800 km^2 in the north central portion of Cordillera Central, Dominican Republic. Five natural stands of *P. occidentalis* were selected from three different ecological zones: two in subtropical dry forest; two from the subtropical very humid forest; and one in the subtropical humid forest.

To validate the models, the original data set was randomly divided into two parts. Eighty percent of the observations (n = 149) were used to develop the taper and volume-ratio models, and 20% (n = 39) to validate them. Summary statistics for both fitting and validation data sets for tree diameter at breast height, total tree height, and various merchantable volume measurements are listed in Table 1.

Table 1.Descriptive statistics for the estimation and verification data sets of *P. occidentalis* trees.

Variable	**Mean**	**Std. Dev.**	**Min**	**Max**
---------- Estimation (n = 149) ----------				
Diameter at Breast Height (cm)	29.80	8.31	11.50	53.50
Total Height (m)	21.90	5.51	8.60	35.00
Total volume (m^3)	0.6028	0.4552	0.0346	2.4930
Cumulative Volume (m^3) to 4 cm top diameter	0.5970	0.4554	0.0313	2.4923
Cumulative Volume (m^3) to 8 cm top diameter	0.5876	0.4562	0.0169	2.4800
Cumulative Volume (m^3) to 14 cm top diameter	0.5451	0.4645	0.0006	2.4607
---------- Verification (n = 39) ----------				
Diameter at Breast Height (cm)	30.32	9.13	9.20	53.10
Total Height (m)	20.99	4.33	7.90	29.40
Total volume (m^3)	0.5570	0.4198	0.0346	2.0885
Cumulative Volume (m^3) to 4 cm top diameter	0.5548	0.4201	0.0313	2.0869
Cumulative Volume (m^3) to 8 cm top diameter	0.5454	0.4209	0.0169	2.0824
Cumulative Volume (m^3) to 14 cm top diameter	0.4992	0.4284	0.0006	2.0444

Sample trees identified from the three ecological zones, were felled and measured for diameter outside- and inside-bark at predetermined points along a stem; stump height (0.3 m above ground), breast height (1.3 m above ground), and then in equal intervals of 1.0 m thereafter to a 4-cm upper stem diameter. At each point, outside-bark diameter measurements were taken with diameter tape. Measurements were carried out using diameter tapes and bark gauge. Volume was computed for each one meter long bole section using Smalian´s formula (Gil and Cuevas 1986). The top section was computed using the volume formula of a cone. The sum of the individual section volumes provides an estimate of the total volume of each tree in the sample; identical computations were carried out for inside bark volumes.

2.2 Taper model development

The development of both the taper and volume-ratio models required the use of multiple measurements of diameters up the stem in each individual sampled tree; therefore the nature of the data violated the assumption of independence and absence of autocorrelation (Schabenberger and Pierce 2002). To reduce the impact of autocorrelation, nonlinear mixed effect models were fitted. For the taper model, a nonlinear version of Kozak's (1988) variable-exponent taper equation was fitted to the same data set used for the volume-ratio model. The nonlinear model is expressed as follows:

$$d_{ij} = (a_0 + g_1)D_i^{a_i} \cdot a_2^{D_i} \cdot X_{ij}{}^{u} + \varepsilon_{ij} \qquad (1)$$

$$u = b_1 Q_{ij}{}^2 + b_2 \ln\left(Q_{ij} + 0.001\right) + (b_3 + g_2)\sqrt{Q_{ij}} + b_4 e^{Q_{ij}} + b_5 (D_i/H_i) \qquad (2)$$

where:

d_{ij} = diameter inside bark (cm) at jth measurement for the ith tree

D_i = diameter at 1.30 m outside bark (cm) for the ith tree

h_{ij} = height above the ground (m) at the jth measurement for the ith tree, $0 \le h \le H$

H_i = total tree height (m) for the ith tree

$$X_{ij} = \frac{\left(1 - \sqrt{\frac{h_{ij}}{H_i}}\right)}{\left(1 - \sqrt{p}\right)}$$

Q_{ij} = relative height above the ground (h / H) at the jth measurement for the ith tree

p = location of the inflection point, assumed to be at 22.5% of total height above the ground

$a_0, a_1, a_2, b_1, b_2, b_3, b_4, b_5, g_1, g_2$= parameters to be estimated.

Previous studies (Perez et al. 1990) have found that the location of the inflection point has little effect on the predictive properties of the taper model. We assume, as they did, a constant inflection point p = 0.225.

Using the notation of the Laird-Ware linear mixed model (1982), the nonlinear mixed-effects modeling approach for the taper model equation can be expressed as:

$$d_i = X_i\beta + Z_i b_i + \varepsilon_i \qquad (3)$$

where, d_i is a vector of diameters inside bark observed on a subject tree *i*; β is a vector of fixed-effects parameters common to all trees; b_i is a vector of random effects parameters associated with tree i, assumed to follow a multivariate normal distribution with mean 0 and a variance-covariance matrix as expressed in (5); and X_i and Z_i are design matrices for the fixed and random effects, respectively.

Model (1) represents the fixed effects mean structure given by $X_i\beta$ in the Laird-Ware linear mixed model. The b_i's are called random effects or random coefficients depending on the nature of the Z_i matrix. The extent to which trees vary about the population average response is expressed by the variability of the b_i's. It is assumed that the b_i's and the ε_i's are normally distributed with:

$$E\begin{bmatrix} b_i \\ \varepsilon_i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (4)$$

$$Var\begin{bmatrix} b_i \\ \varepsilon_i \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix} \quad (5)$$

Variations in tree taper can be expressed using matrix notation as:

$$V = ZGZ' + R \quad (6)$$

V can be modeled by setting up the random effects design matrix *Z* or by specifying covariance structures for *G* and *R* (Littell et al. 2006). Studies have shown that within-tree autocorrelation can be successfully removed or reduced by modeling random effects alone. However, Garber and Maguire (2003) and later Trincado and Burkhart (2006) demonstrated that the results are not always satisfactory. In this study, autocorrelation was addressed by varying the location of two random effects within the terms of the variable exponent coefficient $(b_1, b_2, b_3, b_4, b_5)$, and choosing the one providing the best fit according to Akaike's information criterion (AIC) and Schwarz's Bayesian information criterion (BIC) values. This permitted for the effects of correlation among observations of longitudinal data to be accounted for (Lindstrom and Bates 1990).

2.3 Volume-ratio model development

To develop the volume-ratio model, we followed the combined modeling approaches first reported by Burkhart (1977) and later improved by Gregoire and Schanbenberger (1996a) to explicitly consider the correlations among the measurements on a single tree bole. We employed nonlinear mixed effect models to simultaneously fit in one single model both a total volume and volume proportion component to estimate merchantable volume of *P. occidentalis* trees. With this approach, the joint estimation of the parameters would be more efficient in providing estimations of bole volume to specified upper-bole diameters. We used random effects to capture tree-to-tree variability in size and shape to model the volume of the individual tree bole as well as the volume of the average tree.

The mixed effects model for estimating the cumulative bole volume was specified as,

$$V_{id_j} = [\beta_0 + (\beta_1 + u_{1i}) D_i^2 H_i] exp\left[-(\beta_2 + u_{2i}) * \frac{t_{ij}}{1000} * exp(\beta_3 t_{ij})\right] + \varepsilon_{ij} \quad (7)$$

where,

V_{id_j} = cumulative volume of the ith tree up to an upper diameter d_j

D_i = diameter (cm) at 1.30 m outside bark for the ith tree

H_i = total height (m) for the ith tree

d_j = upper diameter at jth measurement

$$t_{ij} = \frac{d_j}{D_i}$$

$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, u_{1i}, u_{2i}$ = fixed and random parameters to be estimated

ε_{ij} = within-cluster error term for the jth measurement of the ith tree

In model [6] above, $u_{1i} \sim N(0, \sigma_i^2)$, $u_{2i} \sim N(0, \sigma_i^2)$, $cov\ (u_{1i}, u_{2i}) = \sigma_{12}$; $\beta_k, k = 0, \ldots., 4$, are fixed parameters; $\varepsilon_{ij} \sim N(0, \sigma)$; and $E[\varepsilon_{ij}\varepsilon_{ik}] = 0$ for all distinct pairs. The u_{1i} model random slopes in the total volume equation and the u_{2i} model the rate of change and point of inflection in the ratio term (Schanbenberger and Pierce 2002).

The inclusion of the random effects in the total volume component is due to the variability in size which is reflected in total volume variations and the random effects in the ratio term are included to reflect the variability in shape of the stem profiles (Gregoire and Schanbenberger 1996b). To chose how to compute the total bole volume content component on the equation above, i.e., $\beta_0 + (\beta_1 + u_{1i}) D_i^2 H_i$, several mathematical models were preliminarily examined. The combined variable equation, as it is known in quantitative silviculture, provided a straightforward model capable of relating total volume to easily measurable tree explanatory variables; although the constant variance assumption was expected to be violated due to the fact that a direct measure of stem content was used as the dependent variable in a regression model (stem volume variability of larger trees is usually greater than the variability of smaller trees).

The volume proportion component is always positive and tends to one as d_j tends to 0. This portion of the model (7) is constrained so that volume content up to any specified upper stem diameter is greater or equal to 0, and volume content when the upper diameter is equal to 0 is equivalent to total volume.

The conditional distribution $f_{y/b}(y/b)$ is stated as $V_{idj}/\sim N\left(V_{i0} R_{idj}, \sigma^2\right)$ where $V_{i0} = \beta_0 + (\beta_1 + u_{1i}) D_i^2 H_i$ and $R_{idj} = \left[exp\left[-(\beta_2 + u_{2i}) * \frac{t_{ij}}{1000} * exp\left(\beta_3 t_{ij}\right)\right]\right]$.

The distribution of the u_i is as follows,

$$u_i \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, D = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}\right); i = 1, \ldots., n; Cov\left[u_i, u_j\right] = 0$$

2.4 Parameter estimation

PROC NLMIXED in SAS (SAS Institute Inc. 2006) was used to estimate the fixed-effects and variance–covariance parameters associated with the random-effects for the taper and volume-ratio models (Littell et al. 2006).

Fitting the models by nonlinear least squares procedures requires initial estimates of parameters to be input at the start of the iteration procedure. In the case of the taper model, we used the values obtained from the linearized version of the taper equation which has a mathematical value described by the formula:

$$\ln(d) = \ln(a_0) + a_1\ln(D) + \ln(a_2)D + b_1\ln(X)Z^2 + b_2\ln(X) * \ln(Z + 0.001)$$
$$+ b_3\ln(X)\sqrt{Z} + b_4\ln(X)e^Z + b_5\ln(X)(D/H) + \varepsilon \quad (8)$$

where *ln* is natural logarithm and all other variables as previously defined. To locate the position of the random parameters (g_1 and g_2) within the non-linear taper model, every parameter (a_0, a_1, a_2, b_1, b_2, b_3, b_4, and b_5) was tested individually as being random. The final model was selected based on the achievement of best fit as indicated by the lowest AIC, BIC and -2LLvalues. For the volume-ratio model, starting values for the parameters (b_0, b_1, b_2 and b_3) were chosen as the converged iterates from the RMLE fit base on linearization. The positions of the random effects (u_1 and u_2) within the model were chosen following the procedure used with the taper model.

2.5 Comparative analysis of two volume approaches.

Once the taper and volume ratio models were obtained, the fixed-effects parameters from the best linear unbiased predictions were used to predict a mean merchantable volume inside-bark up to three pre-established upper stem diameters (14, 8 and 4 cm), which defined merchantability standards in the study region. Interpolation was performed to estimate these diameters. The random effects parameters were set to their expected value of 0.

The better model was selected based on its ability to accurately predict merchantable volume to the selected upper stem diameters using an independent verification data set. In the case of the volume-ratio model, best linear unbiased predicted merchantable volumes were directly obtained from SAS output for the selected upper stem diameters. For the taper model, the fixed-effect parameters from the best linear unbiased prediction were incorporated using a procedure developed by (Huang 1994, Huang et al. 1994) to obtain the desired merchantable volumes.

To calculate tree merchantable volume (m^3) to specified upper stem diameters with the taper model, an iteration procedure, using estimated coefficients for *P. occidentalis* to define Z, the initial relative height above the ground. The iteration process was repeated until the desired precision of having the absolute value of g0-g1 < 0.00000001 was obtained. A 4.0 cm top diameter inside bark was assumed. To compute merchantable height and merchantable length, a stump height of 0.30 m was assumed. Then merchantable length was divided into 10 sections of equal length and the height above ground from the middle and the top of each section was computed. Prediction of diameter inside bark at the middle and top of each section was accomplished using the taper equation. The final product of the program allowed computing the merchantable volume of the tree to 4.0 cm top dib and calculating merchantable volume in terms of cubic meter using Newton's formula.

Using the observed and calibrated volumes, goodness-of-fit statistics were used to assess the predictive capability of the selected models (Cochran 1963). The following criteria were considered for selection were mean squared error (MSE), coefficient of determination (R^2), average bias (B), sum of squared relative residuals (SSRR), root mean square error (RMSE), and mean average deviation (MAD). They were respectively calculated by:

$$MSE = \frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{n-m} \quad (9)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2} \quad (10)$$

$$B = \frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)}{n} \quad (11)$$

$$B\% = \left[\frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)}{n}\right] * 100 \quad (12)$$

$$RMSE = \sqrt{\sum \frac{(Y_i - \hat{Y}_i)^2}{n}} \quad (13)$$

$$RMSE\% = \left[\sqrt{\sum \frac{(Y_i - \hat{Y}_i)^2}{n}}\right] * 100 \quad (14)$$

$$MAD = \sum \frac{|\hat{e}|}{n} \quad (15)$$

where Y_i and $\widehat{Y}_i$ are the observed and estimated volumes for the i^th observation (i = 1, 2,…, n), with n being the total number of trees for volume comparisons; $\bar{Y}$ is the mean of observed volumes; B is mean prediction error; $B\%$ is percent prediction error; R^2 is the coefficient of determination; MSE is mean square error of the predictions; RMSE is root mean square error of the predictions; RMSE% is the percent root mean square error of the predictions; and MAD is the mean absolute error of prediction. Both the mean and percent prediction errors give an average measure of the prediction bias, in absolute and relative terms respectively. MSE combines the mean bias and the variation of the biases, which gives a better measure of model accuracy and was used as a primary criterion for model evaluation.

In the DR, three main products are derived from *P. occidentalis* boles: "sawlogs" with a small end diameter of 14 cm, "poles" with a small end diameter of 8 cm and "bars" with a small end diameter of 4 cm. The volumes inside-bark to each of those end points was determined and used to develop the models in the study.

3 Results

The fixed effects parameters of both the taper and the volume-ratio model were fitted to the validation data set, to serve as reference for comparing the fitting of both models but including two random parameters. The results on Table 2 show the improvement obtained on the AIC, BIC and -2 log likelihood statistic on both models after including in each two random parameters. The models with two random effects have the smallest values for the goodness-of-fit statistics. The fixed effect models fit only a population-averaged curve and do not take into account any clustering (Calama and Montero 2005). The fixed effect models consider all observations as independent and the residuals are measured against the population average. On the contrary, the residuals of the model including random parameters take into consideration tree-specific predictions.

Table 2. Akaike's Information Criterion (AIC), Baye's Information Criterion (BIC) and minus twice log likelihoods (-2 LL) values for the taper and volume ratio fixed effect models and corresponding models including two random parameters each.

Model	Random Effects	AIC	BIC	-2 LL
Taper	None	14608	14689	14582
	g_1 and g_2	12197	12225	12179
Volume-ratio	None	-5931	-5875	-5949
	u1 and u2	-14638	-14617	-14652

3.1 The Taper model.

The coefficients b_1 and b_5 of the taper equation were not statistically significant and remove from the original model thus, the final taper model was,

$$\boldsymbol{d_{ij}} = (\mathbf{a_0} + \boldsymbol{u_1})\mathbf{D_i}^{\mathbf{a_1}}\mathbf{a_2^{D_i}}\mathbf{X_i}^{\mathbf{b_2 ln(Z_i+0.001)+(b_3+\boldsymbol{u_2})\sqrt{Z_i}+b_4 e^{Z_i}}} + \boldsymbol{\varepsilon_{ij}} \quad (16)$$

After testing every parameter of the non-linear taper model individually as being random, the best fit (lowest AIC, BIC and -2LL values) was achieved when a_2 and b_3 were set as random. Fixed parameter estimates for model (16), along with their corresponding standard error and p-values and the estimates for the residual variance and the variances of the two random effects are displayed in Table 3. Between trees variance of stem form is estimated to be 1.44 cm^2 while within tree variance as reflected by the random effects variance components var(g1) and var(g2) is in the order of 0.01 and 0.004 cm^2.

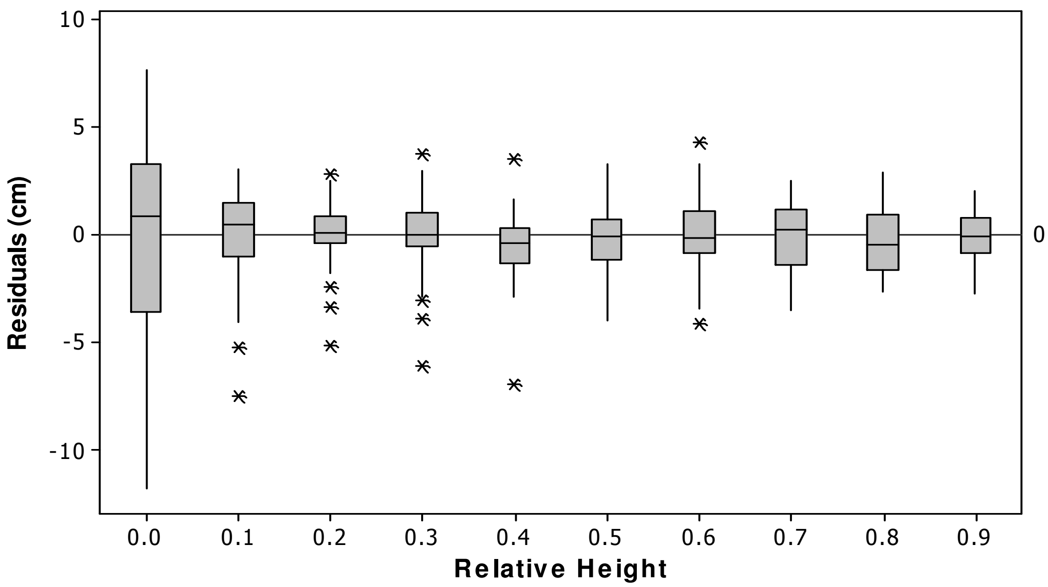
Table 3. Parameter estimates, corresponding standard error and p-values for the taper model .

Parameter	Estimate	Standard Error	p-value
a0	1.05800	0.20090	<.0001
a1	0.89480	0.08280	<.0001
a2	1.00140	0.00304	<.0001
b1		Non-significant	
b2	0.02473	0.00683	0.0004
b3	-1.09360	0.09958	<.0001
b4	0.70910	0.05280	<.0001
b5		Non-significant	
Residual variance	1.44310	0.03584	<.0001
Var(g1)	0.01722	0.00237	<.0001
Var(g2)	0.00412	0.00165	0.0138

Residuals from the model fitting portion plotted against the predicted diameters inside bark values showed no conspicuous pattern in their distribution. The inclusions of the random effects in the model, which account for a large portion of the variability, have effectively removed the autocorrelation among the observations. The predicted diameters (inside bark) by the taper model were compare to the observed values on the basis of bias and precision (RMSE %). To evaluate these upper stem diameter predictions we measured the bias for the mean response using stem diameter measurements from the verification data set at 10 different relative heights (Fig. 1). It does not show a pattern of increasing residual variance with greater relative height, but it shows predictions being less accurate for the lower part of the stem. The relative heights were created by dividing heights by total height. At 0.0 (stump) and 0.1 relative height, the taper model under predicts the diameter slightly, but remains relatively unbiased throughout the upper parts of the bole.

In terms of precision, as measured by the root-mean-square error as a percentage of the mean observed diameter, the corresponding percentages from relative height 0.0 to 0.9 were 10.7, 6.0, 4.8, 5.5, 6.6, 6.5, 8.3, 11.2, 16.4 and 20.6 respectively. These results show low precision at the bottom and top of the stem, being worse at the tip.

Figure 1.Residual values (diameter, cm) from the fitting of the taper model to the validation data set at different relative heights.



3.2 The volume-ratio model

For the volume-ratio model (6), its fixed parameter estimates, corresponding standard error and p-values and the estimates for the residual variance and the variances of the two random effects are displayed in Table 4.

Table 4. Parameter estimates, corresponding standard error and p-values for the volume-ratio model (see equation [7] in text).

Parameter	Estimate	Standard Error	Pr > |t|
b0	0.01814	0.00374	<.0001
b1	0.02974	0.00047	<.0001
b2	10.28940	0.43790	<.0001
b3	5.93180	0.03826	<.0001
Residual variance	0.00087	0.00002	<.0001
Var(u1)	0.00001	0.00000	<.0001
Var(u2)	12.93320	1.86190	<.0001

In terms of precision, as estimated by RMSE%, the performance of the volume-ratio model is poorest at the lower portions of the stem and improves towards the top of the tree.

RMSE% ranges from a high of 48% at 0.1 relative heights down to 10% at the top of the tree. This could be a result of the great variability in bole volume content occurring in the lower portion of the trees due to differences in tree form. The fact that we are interested in predicting volume to upper diameters and RMSE% is below 17% above 0.5 relative heights makes this lack of fit negligible. The predicted cumulative volumes (inside bark) up the stem with the volume-ratio model were compare to the observed values up to 10 different relative heights. Bias, as percentage of the averaged observed cumulative volume, was computed at each relative height. It can be considered negligible. The greatest lack of accuracy occurred at 0.1 relative height where the bias was 2.89% (Fig. 2).

Figure 2.Bias performance of the volume-ratio model as percentage in the validation data set at different relative heights.

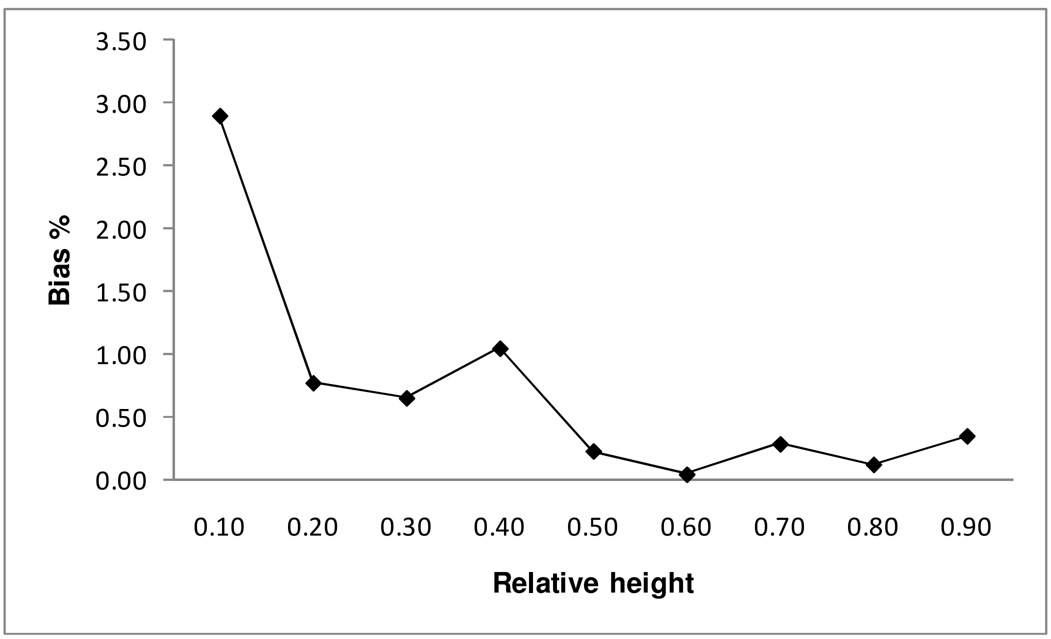
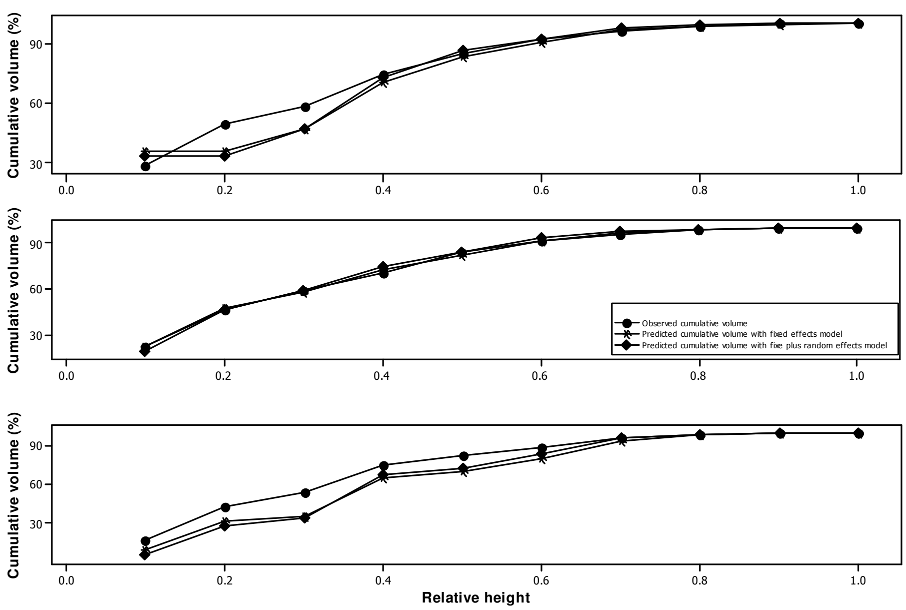


Fig. 3 shows the observed cumulative volume as percentage and superimposed to the corresponding predicted cumulative volume percentages by the fixed effect parameters model and the model including random parameters at different relative heights on three contrasting trees (small, medium and large) of the verification data set. The fixed effect model approximates better the cumulative observed volume between 0.1 and 0.3 relative heights in all three trees. From 0.35 to 0.7 the fixed plus random effects model approximates better the observed cumulative volume percentage and from 0.7 to 0.1 relative heights, at the tip, both models predict similar cumulative volume values. It is evident from Figure 4, that the volume-ratio model including random parameters fit the verification data better.

Cumulative bole volume predictions by the taper and volume-ratio models on verification data set were estimated to three predefined upper stem diameters normally used by forest industries in the study region, employing both, the taper and the volume-ratio models. For “saw logs” that upper stem diameter is 14 cm; for “posts” it is 8 cm and; for “bars”, 4 cm. The predictions were carried out considering both, the fixed and the random part of the models. The predicted cumulative volumes were then compared to the corresponding observed cumulative volumes by means of residual analysis. The results are shown on Table 5.

Figure 3. Observed cumulative volume as a percentage superimposed to the corresponding averaged predicted cumulative volumes by the fixed effect parameters (dotted line) model and the model including random parameters (red line) at different relative heights on the 39 trees of the verification data set.



When comparing the volume-ratio model to the taper model for their ability to predict merchantable volume to varying merchantability limits using independent data, the taper model proved slightly superior to the volume-ratio model according to most goodness-of-fit statistics. Only for the ‘bar” product (4 cm upper diameter) was the volume-ratio model slightly better than the taper model in terms of bias and MAD.

Table 5. Goodness-of-fit statistics from the residual analysis performed on the validation data set for cumulative bole volume predictions to three predefined upper stem diameters of 14 cm, 8 cm and 4 cm by the taper and volume-ratio models.

Goodness of fit statistic	Upper Stem Diameter 14 cm		8 cm		4 cm	
	Taper	Volume Ratio	Taper	Volume Ratio	Taper	Volume Ratio
MSE	0.014	0.022	0.014	0.021	0.014	0.020
RMSE	0.090	0.115	0.090	0.110	0.090	0.110
RMSE%	15.650	19.970	14.340	17.690	14.276	17.408
Bias	0.021	-0.042	0.032	-0.037	0.036	-0.034
Bias%	3.690	7.240	5.130	5.873	5.648	5.297
MAD	0.049	0.054	0.051	0.050	0.052	0.050
R^2	0.930	0.917	0.928	0.920	0.926	0.920

4 Discussion

The selection of the random parameters in the taper models was chosen based on the values of AIC, BIC and -2LL according to the model validation results. The lowest values for the fit statistics above were obtained when the first random coefficient g_1 was added to the intercept term $a_0 D_i^{a_1} a_2^{D_i}$, and the second random coefficient g_2 was added to the fix parameter β_3 accompanying the $\sqrt{Z}$ term in the variable-exponent taper model. Both random parameters were statistically significant, demonstrating significant variability in fixed parameters between trees. The correlation coefficient of g_1 and g_2 was low (0.3) indicating that variation has been accounted by current covariates. Yang et al. (2009) also incorporated one random parameter into the exponent (b_0 parameter) and another for a_1 to obtain the best results for several candidate variable-exponent taper equations for lodgepole pine in Alberta, Canada.

Several authors have reported the need to incorporate surrogates of stem form into taper models (Burkhart and Walton 1985, Kozak 1988) as a means to improve the predictive ability of those types of models. Garber and Maguire (2002) and Yang et al. (2009) found functions using D/H ratio, a proxy for tree form, to be significant, the former in their taper models for *Abies grandis*, *P. ponderosa* and *P. contorta* in Oregon; and the later for lodgepole pine in Alberta. Our results for *P. occidentalis* using D/H were not significant in our model.. The effects from this tree form alternative variable may have been captured by the random component on the exponent part of the model.

The location term (i.e., Z) was not significant on its own, although transformed Z functions were statistically significant. The fact that all these terms are functions of the same variable Z, may result in the presence of multicollinearity among the independent variables. Although as Kozak (1997) demonstrated, in our case the predictions were not significantly affected. The inclusion of random effects may have also helped in alleviating autocorrelation problems.

Our taper model had somewhat large prediction bias in the lower portion of the bole (21% at a relative height Z = 0.1). Higher up the bole, the bias percentage decreases substantially, to values between 5 and 9%. These results agree with the findings of Garber and Maguire (2003) but are contrary to what was reported by Calama and Montero (2006). The later paper found their linear mixed taper model to be was less accurate in predicting tree upper stem diameters. However, they developed a linear, rather than nonlinear, mixed model.

Previous studies (VanderSchaaf and Burkhart 2007, Yang et al. 2009) found that including random effects in the modeling process reduces the majority of the autocorrelation among observations. Others report a reduction, but not elimination, of autocorrelation by including random effect parameters (Trincado and Burkhart 2006). Our results for *P. occidentalis*, similar to what Yang et al. (2009) found in lodgepole pine, did not find obvious trends in the residuals of the variable-exponent equation (Fig. 1), an indication that the inclusion of random effects in the model effectively accounted for the autocorrelation on the observations.

The random effects in the volume-ratio model also take into account inter-tree variation through the marginal covariance structure. As with the taper model, the random parameters allow individualization of the model fit to each subject tree, accounting for within-tree covariances. The random parameters in the volume-ratio model explain variability in size (total volume) and shape of the volume profile. In this model,the random u_{1i} term models random slopes in the total volume component of the equation and u_{2i} models the rate of change of the tree profile.

The improvement of the model fit by the inclusion of the random effect parameters can be verified by the value of the model fit statistics (i.e., AIC, BIC and -2LL) when the random effects are included or not. The model including the two random parameters has the smallest values on the fitting statistics. Fig. 3 also clearly shows that the model including the random parameters fits more closely the observed cumulative volume at most relative heights of the tree profiles for the three selected trees. By not including the random effect parameters it is assumed that all the cumulative volume observations are independent. In the random effects models, the residuals are measured against the tree-specific predictions whereas in the fixed effects model, the residuals are measured against the mean tree predictions.

To evaluate between variable-exponent taper equation and the volume-ratio model in estimating cumulative bole volume to three predefined upper stem diameters for *P. occidentalis*, Sw. trees, seven statistics were employed. The taper model was superior for merchantable upper stem diameters in terms of RMSE. In terms of percent bias, the taper model was superior by 3.55% and 0.74% upto 14 and 8 upper stem diameters, respectively, while the volume-ratio was marginally better by 0.35% at a 4 cm upper stem diameter.

Although the taper model was superior in almost all the statistics in terms of accuracy and precision for estimating merchantable volume for different utilization standards, the differences in performance of prediction by the two models were very similar. However, if merchantability standards change, the taper model would be more flexible to adapt to the new standards.

Although the difference in goodness-of-fit values is minimal for these comparisons, the use of these two different methods to predict merchantable volume in *P. occidentalis* produced results inconsistent with other pine species. In predicting merchantable volume, other studies (Gaztelurrutia and Montero 2001) have shown the total volume-ratio approach with advantages over taper models because in their development, merchantable cubic-meter volumes and heights are calculated directly, making them easier to apply. Moreover, taper equations are fitted with the purpose of diameter prediction, contrary to the volume-ratio model, which is used to optimize the prediction of volume.

The improvement of the volume-ratio model to predict cumulative bole volume to any merchantability limit and the taper model to predict bole diameter at a stipulated height based on measurements of overall tree size is significant when random effect parameters are used in conjunction with a nonlinear mixed procedure. The random effects allow to individualize the fit of the model to each subject tree and to explain the inter-tree variation by means of the marginal covariance structure. Garber and Maguire (2003) found that the inclusion of random parameters only was not enough to address autocorrelation in finding variable-exponent taper models for *A. grandis*, *P. contorta* and *P. ponderosa*. They had to directly model the autocorrelation by adding a variance-covariance matrix in addition to the random parameters. The main disadvantage with these models is the complexity of the fitting process, which requires iterative methods, although this drawback is easily overcome with the speed of computers. Another disadvantage is that the fitting depends heavily on the choice of initial parameters to start the iterations.

From a financial perspective, the most important part of a tree bole is within the first half of the height. In the Dominican Republic, the most valuable product is the log with a top diameter of 14 cm. Knowing that the bias and precision of the nonlinear mixed taper model were superior to the volume-ratio model for that portion of the bole, model (16) should be used to estimate inside-bark volume content of *P. occidentalis.*

Acknowledgements

This investigation was carried out thanks to important contributions of FONDOCYT, a research grant from the government of the Dominican Republic administered by the Ministerio de Educación Superior de Ciencia y Tecnología. Great minds should be recognized for their contributions to the successful finalization of this investigation. I would like to express my gratitude to the members of my doctoral committee, Doctor Ralph D. Nyland, Doctor Lianjun

Zhang and Doctor Edwin H. White. Working with this dream team is one of those rare opportunities that happen once in a lifetime.

References

Avery, T.E. & Burkhart, H.E. 2002. Forest Measurements. Fifth Edition. McGraw Hill, New York. 456 p.

Bailey, R.L. & Clutter, J.L. 1970. Volume tables for old-field Loblolly pine plantations in the Georgia Piedmont. Georgia Forest Research Council, Report No. 22—Series 2, 4. 71 p.

Banco Central de la República Dominicana. 2004. Annual Report. Gobierno Dominicano. 2056 p.

Burkhart, H.E. 1977. Cubic foot volume of loblolly pine to any merchantable top limit. Southern Journal of Applied Forestry 1(2): 7-9.

Burkhart, H.E. & Walton, S.B. 1985. Notes: Incorporating crown ratio into taper equations for loblolly pine trees. Forest Science 31(2): 478-484.

Calama, R. & Montero, G. 2005. Multilevel linear mixed model for tree diameter increment in stone pine (Pinus pinea): a calibrating approach. Silva Fennica 39(1): 37–54.

_____ & Montero, G. 2006. Stand and tree-level variability on stem form and tree volume in Pinus pinea L.: A multilevel random components approach. Investigación Agraria: Sistemas Recursos Forestales 15(1): 24-41.

Cochran, W.G. 1977. *Sampling Techniques.* Third Edition. John Wiley and Sons. New York. Chichester. Brisbane. Toronto. 428 p.

Dobler, G., Peralta, L.E., Debord, L.T. & Torres, J.G. 1995. Investigación y Manejo de Especies Maderables de Uso Comun en La Sierra: Guía Técnica. Plan Sierra, Inc., San Jose de las Matas, Dominican Republic. 269 p.

Garber, S.M. & Maguire, D.A. 2003. Modeling stem taper of three central Oregon species using nonlinear mixed effects models and autoregressive error structures. For. Ecol. Manage. 179(1–3):507–522.

Gaztelurrutia, M. & Montero, G. 2001. Modelo de simulación de claras en masas de *Pinus silvestris* L. Monografias Investigación Agraria N. 3. Instituto nacional de investigación y tecnología agraria y alimentaria. Madrid. España. 114 p.

Gil, T. & Cuevas, B. 1986. Tabla de volumen local y general sobre Pinus occidentalis para el Proyecto La Celestina, San José de las Matas. ISA-PUCAMAYMA, Santiago de los Caballeros. 58 p.

Gregoire, T.G. & Schabenberger, O. 1996a. Nonlinear mixed-effects modeling of cumulative bole volume with spatially correlated within-tree data. Journal of Agricultural, Biological, and Environmental Statistics, 1 (1): 107-119.

_____ & Schabenberger, O. 1996b. A nonlinear mixed-effects model to predict cumulative bole volume of standing trees. Journal of Applied Statistics 23, 257-271.

Huang, S. 1994. Ecologically based individual tree volume estimation for major Alberta tree species. Alberta Sustainable Resource Development. Public Land and Forest Division. 41 p.

_____ , Price, D., Morgan, D. & Peck, K. 1994. Kozak's Variable-exponent taper equation regionalized for White spruce in Alberta. Western Journal of Applied Forestry 15(2): 75-85.

Husch, B., Beers, T.W. & Kershaw Jr., J.A. 2002. Forest mensuration, 4rd edn. Wiley, New York. 456p.

Jones, R.H. 1990. Serial correlation or random subject effects? Commun Stat Simul Comput 19:1105–1123.

Kozak, A. 1997. Effects of multicollinearity and autocorrelation on the variable-exponent taper function. Canadian Journal of Forest Research. 27: 619-629.

_____ 1988. A variable-exponent taper equation. Canadian Journal of Forest Research. 18: 1363-1368.

_____ 1998. Effects of upper stem measurements on the predictive ability of a variable-form taper equation. Canadian Journal of Forest Research. 28(7): 1078-1084.

_____ 2004. My last words on taper equations. Forest Chronicle. 80:507–515.

Laird, N.M. & Ware, J.H. 1982. Random-effects models for longitudinal data. Biometrics 38:963–974.

Leites, L.P. & Robinson, A.P. 2004. Improving taper equations of loblolly pine with crown dimensions in a mixed-effects modeling framework. Forest Science 50:204–212.

Lindstrom, M.J. & Bates, D.M. 1990. Nonlinear mixed-effects models for repeated measures data. Biometrics 46:673–687.

Littell, R.C., Henry, P.R. & Ammerman, C.B. 2006. Statistical analysis of repeated measures data using SAS procedures. Journal of Animal Science. 76:1216-1231.

Montalvo, J.M., Bueno, S.W. & Torres, J.G. 2001.Ecuacion para el cálculo de volumen del árbol para Pinus occidentalis, Sw.; en el Plan Sierra, Republica Dominicana. Revista Chapingo. Serie ciencias forestales y del ambiente, Universidad Autonoma Chapingo, Mexico. Vol 7, No. 001. Pp 49-53.

Perez, D.N., Burkhart, H.E. & Stiff, C.T. 1990. A variable-form taper function for Pinus oocarpa Schiede in Central Honduras. Forest Science 36: 186-191.

SAS Institute Inc. 2006. SAS System for Regression. Third Edition. SAS Institute Inc., Cary, N.C.

Schabenberger, O. & Pierce, F.J. 2002. Contemporary statistical models for the plant and soil sciences. Boca Raton, FL: CRC Press. 738p.

Tasissa, G. & Burkhart, H.E. 1998. Modelling thinning effects on ring specific gravity of loblolly pine (Pinus taeda L.). Forest Science 44(2), 212-223.

Trincado, G. & Burkhart, H.E. 2006. A generalized approach for modeling and localizing stem profile curves. Forest Science 52:670–682.

VanderSchaaf, C.L. & Burkhart, H.E. 2007. Comparison of methods to estimate Reineke's maximum size-density relationship species boundary line slope. Forest Science. 53:435–442.

West, P.W., Ratkowsky, D.A. & Davis, A.W. 1984. Problems of hypothesis testing of regressions with multiple measurements from individual sampling units. Forest Ecology and Management. 7:207–224.

Yang, Y., Huang, S., Trincado, G. & Meng, S.X. 2009. Nonlinear mixed-effects modeling of variable-exponent taper equations for lodgepole pine in Alberta, Canada. European Journal of Forest Research. 128:415–429.