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

Santiago Wigberto Bueno-López [(1)], Eddie Bevilacqua [(2)]

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 the collection of low cost data but high accuracy tree measurements. This paper reports on the 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* 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.

Keywords: Nonlinear Mixed-effects Models, Restricted Maximum Likelihood, Taper Model, Volume-ratio Model, Random Coefficients, Calibration, *Pinus occidentalis*

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

(1) Vicerrectoría de Investigaciones e Innovación, Pontificia Universidad Católica Madre y Maestra, Autopista Duarte km 1.500, Santiago de los Caballeros (República Dominicana); (2) State University of New York, College of Environmental Science and Forestry, 1 Forestry Drive, 13210 Syracuse (NY - USA)

@ Santiago W Bueno-López (swbueno@gmail.com)

(Gregoire & 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 & Clutter 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 & Burkhart 2002); and (3) stem profile equations which predict upper stem diameters at any point along the tree bole developed by Kozak (2004). They can be integrated to give volume content within any portion of the tree bole. 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 longi-

tudinal 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 & Montero 2006). Observations from the same unit of analysis, the trees, are correlated (Gregoire & Schabenberger 1996b, Garber & Maguire 2003, Leites & Robinson 2004).

The fixed effects model that includes explanatory variables accounts for a part of the total variation in tree volume content. There is a residual variability component, representing differences in attributes acting at stand or tree level, that are difficult to observe, encompassing heredity and environmental factors, or even measurement errors, introducing a stochastic level of variability in tree volume content. Calama & Montero (2006) stated: "The use of mixed models allows us to identify the different sources of stochastic variability that are not explained by the fixed part of the model by dividing the residual variance into different components".

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 & Schabenberger 1996a). The violation of the assumption of independence required by the linear regression model would make inferences and test of significance invalid (West et al. 1984). The properties of the least square estimators would be affected when observations are not independent, although the approach itself does not make assumption of independent observations.

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 & Schabenberger 1996a).

Two general methods have been proposed to fit continuous, unbalanced, multilevel longitudinal data. The first proposed by Gregoire & 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 im-

pact of autocorrelation. The second approach introduces an assumed correlation structure into a generalized least squares solution (Tasissa & 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 & 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 Sw. is an endemic pine species of Hispaniola Island (Caribbean Sea). In Dominican Republic (DR), it 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 (Bueno-López & Bevilacqua 2010). 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 de la República Dominicana 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 (Bueno-López & Bevilacqua 2011). Few volume equations are available for the species at a regional level (Montalvo et al. 2001). 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 and a total volume-ratio model to *P. occidentalis* 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 valida-

tion data set; and (3) recommend that best model for estimating merchantable volume for different utilization standards for this species. Parameter estimates along with their respective significance levels and goodness-of-fit statistics of the models tested will be reported.

Material and methods

Data set

The study area is a region of approximately 1800 km² 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 Tab. 1.

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 (Husch et al. 2002). 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.

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 & Pierce 2002). To reduce the impact of autocorrelation, nonlinear mixed effect models were fitted. For the taper model, a nonlinear version of Kozak (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 (eqn. 1, eqn. 2):

$$d_{ij} = (a_0 + g_1) D_i^{a_1} \cdot a_2^{D_i} \cdot X_{ij}^{u} + \varepsilon_{ij}$$

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

where d_{ij} represents the j-th diameter measurement inside bark (in cm - $j = 1, 2, 3, \ldots$) at the i-th tree ($i = 1, 2, \ldots$); D_i is the corresponding diameter at 1.30 m outside bark (cm) for the i-th tree; $a_0, a_1, a_2, b_1, b_2, b_3, b_4, b_5$ are the fixed-effects parameters; g_1, g_2 are the random-effect parameters assumed to be independent and identically distributed random variables (iid): $N(0, \sigma_g^2)$; ε_{ij} are the residual errors assumed to be iid $N(0, \sigma_\varepsilon^2)$; and independent of g_1, g_2; H_i is the corresponding total tree height (m) for the i-th tree; Q_{ij} is the relative height above the ground (h/H) at the j-th measurement for the i-th tree. In X_{ij} ahead (eqn. 3), h_{ij} is the height above the ground (m) at the j-th measurement for the i-th tree ($0 \le h \le H$), and p is the location of the inflection point, assumed to be at 22.5% of total height above the ground (eqn. 3):

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

Previous studies (Perez et al. 1990) have found that the location of the inflection point

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

Datasets	Variables	Mean	St. Dev.	Min	Max
Estimation (n=149)	Diameter at breast height (cm)	29.8	8.31	11.5	53.5
	Total height (m)	21.9	5.51	8.6	35
	Total volume (m³)	0.6	0.45	0.03	2.49
	Cumulative volume (m³) to 4 cm top diameter	0.59	0.45	0.03	2.49
	Cumulative volume (m³) to 8 cm top diameter	0.58	0.45	0.01	2.48
	Cumulative volume (m³) to 14 cm top diameter	0.54	0.46	0	2.46
Verification (n=39)	Diameter at breast height (cm)	30.32	9.13	9.2	53.1
	Total height (m)	20.99	4.33	7.9	29.4
	Total volume (m³)	0.55	0.42	0.03	2.08
	Cumulative volume (m³) to 4 cm top diameter	0.55	0.42	0.03	2.08
	Cumulative volume (m³) to 8 cm top diameter	0.54	0.42	0.02	2.08
	Cumulative volume (m³) to 14 cm top diameter	0.49	0.43	0.01	2.04

has little effect on the predictive properties of the taper model. We assume, as they did, a constant inflection point $p = 0.225$.

Using a general notation, the nonlinear mixed-effects modeling approach for the taper model equation can be expressed as (eqn. 4):

$$d_{ij} = f(x_{ij}, \theta, b_i) + e_{ij}$$

where x_{ij} is a vector of regressor variables and θ is a vector of fixed effects (Schabenberger & Pierce 2002).

Random effects are denoted by the vector b_i, modelers of heterogeneity of observations between trees. The b_i are zero mean random variables with variance covariance matrix as expressed in eqn. 6; d_{ij} are diameters inside bark observed on a subject tree i, and e_{ij} are within three errors with mean 0 and possibly correlated. The extent to which trees vary around the population average response is expressed by the variability of the b_i's. It is assumed that b_i's and the e_{ij}'s are normally distributed with (eqn. 5, eqn. 6):

$$E\begin{bmatrix} b_i \\ \varepsilon_i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$Var\begin{bmatrix} b_i \\ \varepsilon_i \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}$$

Variations in tree taper can be expressed using matrix notation as follows (eqn. 7):

$$V = ZGZ' + R$$

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 & Maguire (2003) and later Trincado & 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 & Bates 1990).

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 & Schabenberger (1996a) to explicitly consider the correlations among the measurements on a single tree bole. We employed nonlinear mixed ef-

fect 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 follows (eqn. 8):

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

where V_{id_j} is the cumulative volume of the i-th tree up to an upper diameter d_j; D_i is the diameter (cm) at 1.30 m outside bark for the i-th tree; H_i is the total height (m) for the i-th tree; d_j is the upper diameter at j-th measurement ($t_{ij} = d_j / D_i$); $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, u_{1i}, u_{2i}$ are fixed and random parameters to be estimated; ε_{ij} is the within-cluster error term for the j-th measurement of the i-th tree.

In the model represented by eqn. 7 above, $u_{1i} \sim N(0, \sigma_1^2)$, $u_{2i} \sim N(0, \sigma_1^2)$, $cov(u_{1i}, u_{2i}) = \sigma_{12}$; β_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} models random slopes in the total volume equation and the u_{2i} models the rate of change and point of inflection in the ratio term (Schabenberger & 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 & Schabenberger 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 1 as d_j tends to 0. This portion of the model in eqn. 8 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 $f_{(y/\beta)}(y/\beta)$ is stated as $V_{idj} / \sim N(V_{i0} R_{idj}, \sigma^2)$ where (eqn. 9):

$$V_{i0} = \beta_0 + (\beta_1 + u_{1i})D_i^2 H_i$$

and (eqn. 10):

$$R_{idj} = \exp[-(\beta_2 + u_{2i}) \cdot \frac{t_{ij}}{1000} \cdot \exp(\beta_3 t_{ij})]$$

The distribution of the u_i is as follows (eqn. 11):

$$u_i \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, D = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}\right)$$

where $i=1, \ldots, n$ and $cov[u_i, u_j] = 0$.

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 following formula (eqn. 12):

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

where ln is the natural logarithm and all other variables are 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 -2LL values. 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.

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),

Tab. 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
	u_{1i} and u_{2i}	-14638	-14617	-14652

which defined merchantability standards in the study region. Linear interpolation (Husch et al. 2002) 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 et al. 1994) to obtain the desired merchantable volumes.

To calculate tree merchantable volume (m³) to specified upper stem diameters with the taper model, an iteration procedure was applied 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 $h_0 - h_1 < 0.00000001$ was obtained, where h_0 represent an initial value for relative height of diameters up the stem and h_1 represents relative height of the defined diameter of the presumed utilization standard. 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 m³ using Newton's formula (Husch et al. 2002).

Using the observed and calibrated volumes, goodness-of-fit statistics were used to assess the predictive capability of the selected models (Cochran 1977). The following criteria were considered for selection: mean squared error (*MSE*), coefficient of determination (*R²*), average bias (*B*), sum of squared relative residuals (*SSRR*), root mean square error (*RMSE*), and mean average deviation (*MAD*). They were respectively calculated by (eqn. 13-19):

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

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

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

$$B\% = \frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)}{(n - m)} \cdot 100$$

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

$$RMSE\% = 100 \cdot \sqrt{\sum \frac{(Y_i - \hat{Y}_i)^2}{n - m}}$$

$$MAD = \frac{\sum |\hat{e}|}{n}$$

where Y_i and $\hat{Y}_i$ are the observed and estimated volumes for the *i*-th observation ($i = 1, 2, \ldots, n$), 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; *R²* 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.

Results

The fixed effects parameters of both the taper and the volume-ratio model were used on the validation data set to generate estimates, and to serve as reference for comparing the fitting of both models, including two random parameters. The results in Tab. 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 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.

The Taper model

The coefficients b_1 and b_5 of the taper equation were not statistically significant at $\alpha = 0.05$ and removed from the original model thus; the final taper model was (eqn. 20):
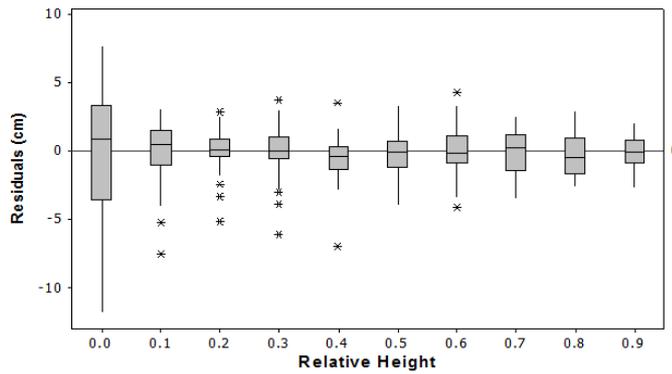
$$d_{ij} = (a_0 + u_{1i}) D_i^{a_1} \cdot a_2^{D_i} \cdot \\ \cdot X_i^{b_2 \ln(Z_i + 0.001) + (b_3 + u_{3i})\sqrt{Z_i} + b_4 e^{Z_i}} + \varepsilon_{ij}$$

After testing every parameter of the nonlinear 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 in eqn. 20, along with their corresponding standard error and p-values, the estimates for the residual variance and the variances of the two random effects are displayed in Tab. 3. Between trees variance of stem form computed with the taper model

Tab. 3 - Parameter estimates, corresponding standard error and *P*-values for the taper model. (**) = *P* < 0.01; (****) = *P* < 0.0001.

Parameter	Estimate	Standard error	*P*-value
a_0	1.058	0.201	****
a_1	0.894	0.082	****
a_2	1.001	0.003	****
b_1		Non-significant	
b_2	0.024	0.006	****
b_3	-1.093	0.099	****
b_4	0.709	0.052	****
b_5		Non-significant	
Residual variance	1.443	0.035	****
Var(g_1)	0.017	0.002	****
Var(g_2)	0.004	0.001	**

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

Tab. 4 - Parameter estimates, corresponding standard error and p-values for the volume-ratio model (see eqn. 8 in the text). (****): $P < 0.0001$.
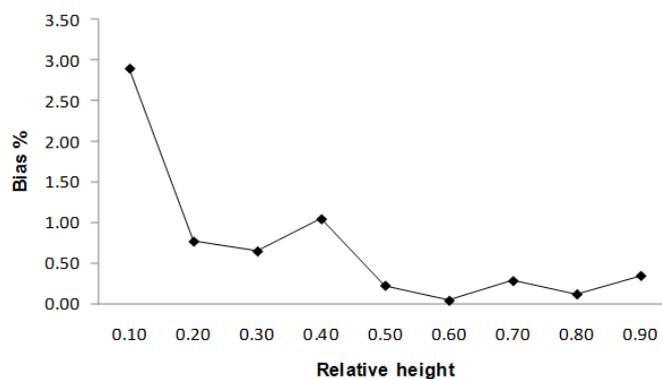
Parameter	Estimate	Standard error	Pr > |t|
b_0	0.018	0.004	****
b_1	0.029	0.001	****
b_2	10.289	0.437	****
b_3	5.931	0.038	****
Residual variance	0.001	0.000	****
Var(u_1)	0.000	0.000	****
Var(u_2)	12.933	1.861	****

(eqn. 20), is estimated to be 1.44 cm², while within tree variance as reflected by the random effects variance components var(g_1) and var(g_2) is in the order of 0.01 and 0.004 cm².

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). No patterns of increasing residual variance with greater relative height were detected, but predictions resulted 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 underpredicts 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.

The volume-ratio model

For the volume-ratio model (eqn. 8), fixed parameter estimates, corresponding standard error and p-values, the estimates for the residual variance and the variances of the two random effects are displayed in Tab. 4.

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, and may be considered negligible. The greatest lack of accuracy occurred at 0.1 relative height where the bias was 2.89 % (Fig. 2).

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 were studied. The fixed effect model approximated 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 percen-

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

Tab. 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 statistics	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	0.02
RMSE	0.09	0.115	0.09	0.11	0.09	0.11
RMSE %	15.65	19.97	14.34	17.69	14.276	17.408
Bias	0.021	-0.042	0.032	-0.037	0.036	-0.34
Bias %	3.69	7.24	5.13	5.873	5.648	5.297
MAD	0.049	0.054	0.051	0.05	0.052	0.05
R²	0.93	0.917	0.928	0.92	0.926	0.92

tage and from 0.7 to 0.1 relative heights, at the tip, both models predict similar cumulative volume values.

Cumulative bole volume predictions by the taper and volume-ratio models on verification data set were estimated to three pre-defined 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 in Tab. 5.

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) the volume-ratio model was slightly better than the taper model in terms of bias and MAD.

Discussion

The selection of the random parameters in the taper models was 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 $a_0 D_i{}^{a1} a_2{}^{Di}$, and the second random coefficient g_2 was added to the fix parameter β_3 accompanying the 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 & Walton 1985, Kozak 1988) as a means to improve their predictive ability. Garber & Maguire (2003) and Yang et al. (2009) found significant functions using D/H ratio (a proxy for tree form) in their taper models for *Abies grandis*, *P. ponderosa* and *P. contorta* in Oregon (the former), and for lodgepole pine in Alberta (the latter). Our results for *P. occidentalis* shows that D/H ratio was 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. 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 & Maguire (2003), but are contrary to what reported by Calama & Montero (2006). The latter paper found their linear mixed taper model to be less accurate in predicting tree upper stem diameters. However, they developed a linear, rather than nonlinear, mixed model.

Previous studies (Van der Schaaf & 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 the complete elimination, of autocorrelation by including random effect parameters (Trincado & Burkhart 2006). Our results for *P. occidentalis*, analogously to what Yang et al. (2009) found in lodgepole pine, did not reveal 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 the 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. By excluding 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 predic-

tions, whereas in the fixed effects model, the residuals are measured against the mean tree predictions.

In the present work, we fitted both a variable-exponent taper equation and a total volume-ratio model to *P. occidentalis* data from La Sierra in the Dominican Republic, using a mixed-effects modeling approach, with the goal of comparing the accuracy and precision of each model based on evaluation criteria using an independent validation data set, and recommend that best model for estimating merchantable volume for different utilization standards for this species in the region. To reach a decision in selecting the best function, caution must be taken. The taper model is superior for predicting merchantable upper stem diameters in terms of the RMSE evaluation. In terms of percent bias, the taper model is superior by 3.55 % and 0.74 % up to 14 and 8 upper stem diameters, respectively; however, the volume-ratio is marginally better by 0.35 % at a 4 cm upper stem diameter.

Unbiased and accurate individual tree volume prediction is useful in determining stand volume in management inventories. Based on this standpoint, the prediction performance by the two models are very similar. However, whenever merchantability standards should change, the taper model would be more flexible to adapt to the new standards and to predict stand volume estimates.

The main concern of these models to predict merchantable volume in *P. occidentalis* is that they are contrasting with the results found for other pine species when comparing taper and volume ratio to estimate bole volume. In predicting merchantable volume, other studies (Gaztelurrutia & Montero 2001) have shown that the total volume-ratio approach have advantages over taper models, because in their development, merchantable cubic-meter volumes and heights are calculated directly. In contrast, taper models are developed using inside diameters which are indirectly related to volume content. 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.

Nonetheless, it is important to point out that, despite the inconsistent results obtained, the methodology proposed in the present work could be useful. The improvement of the predictions by both models (eqn. 8 and eqn. 20) 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 & Maguire (2003) found that the inclusion of random parameters only was not enough to address autocorrelation in finding variable-ex-

ponent taper models for *Abies grandis* (Douglas ex D. Don) Lindl., *Pinus contorta* Dougl., *and Pinus ponderosa* Lawson & C. Lawson. They had to directly model the autocorrelation by adding a variance-covariance matrix in addition to the random parameters.

Conclusions

We fitted a variable-exponent taper equation and a total volume-ratio model to *P. occidentalis* stem data from La Sierra, Dominican Republic, employing a nonlinear mixed-effects modeling approach. 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 from *P. occidentalis,* is the log with a top diameter of 14 cm. After comparing the accuracy and precision of each of the models tested based on robust evaluation criteria in an independent validation data set, we concluded that the bias and precision of the nonlinear mixed taper model is superior to the volume-ratio model for that portion of the bole. Therefore the model represented in eqn. 20 should be used to estimate inside-bark volume content of *P. occidentalis*, up to the different utilization standards studied, namely top inside diameters of 14 cm, 8 cm and 4 cm.

Due to the fact that the country is mostly mountainous and soils are very fragile for agriculture production, it is clear that forest production should probably be the most important economic activity in the study region and in the country. The soils in the study area are generally poor and shallow, with great extensions of terrain severely degraded and eroded due to deforestation and migratory agriculture. These areas were once populated by our endemic species *P. occidentalis*, which is the only species in the country capable of prospering in these types of soils. Forest stands not only provide cover to keep soils stable and habitat for wildlife, but they also produce wood and generate rural employment, bringing along important benefits to the social ecosystem.

To maintain and increase *P. occidentalis* populations, so that the species continues to serve its role as the only one capable of growing in the shallow and infertile soils in the country, as well as continue being the main commercial tree species, it becomes imperative that forest management in the Dominican Republic dispose of the tools developed in the present study. They will permit the prediction of cumulative bole volume to any predefined upper stem diameter on *P. occidentalis* standing trees, which is essential for estimating current inventory levels and making informed decisions regarding the management of this forest resource. They have allowed the determination of the productive capacity of the three main ecological

zones where the species is intensively managed. Further processing of the information on volume content suggest that the amount of inputs committed and silvicultural concepts applied to the management of these forest stands should be more intense in the humid zone and gradually become more extensive from the intermediate to the dry zone.

The productive capacity of the humid zone calls for dedicating its fragile soils to the production of wood, especially *P. occidentalis*, mixed with plantings of different broadleaf species to help in the protection of water ways and steep terrain. Forest stands of the species in the dry zone have worse shape and lower commercial value; nonetheless, they should be kept under the cover of *P. occidentalis*. To manage the species, clear cutting as a reproduction method should be avoided because of soil conditions. These shallow and poor soils would favor the occupancy of the terrain by very aggressive shrub species, impeding the successful establishment of regeneration. Establishing new plantations in this zone requires that a big effort be put into site preparation, both in terms of resources and of costs.

Acknowledgements

References

Avery TE, Burkhart HE (2002). Forest measurements (5th edn). McGraw Hill, New York, USA, pp. 453.

Bailey RL, Clutter JL (1970). Volume tables for old-field Loblolly pine plantations in the Georgia Piedmont. Forest Research Council, Georgia, USA, pp. 18.

Banco Central de la República Dominicana (2004). Annual report. Gobierno Dominicano, Santo Domingo de Guzman, República Dominicana, pp. 2056.

Bueno-López SW, Bevilacqua E (2010). Modeling stem increment in individual *Pinus occidentalis* Sw. trees in La Sierra, Dominican Re-

public. Forest Systems 19 (2): 170-183. [online] URL: http://revistas.inia.es/index.php/fs/article/view/1312/1289

Bueno-López SW, Bevilacqua E (2011). Developing a diameter-distribution prediction system for *Pinus occidentali*, s Sw. in La Sierra, Dominican Republic. Revista Chapingo, Serie Ciencias Forestales y del Ambiente XVII (1): 115-132. [online] URL: http://www.academia.edu/1213948/Developing_a_Diameter-Distribution_Prediction_System_for_Pinus_occidentalis_Sw._in_La_Sierra_Dominican_Republic

Burkhart HE, Walton SB (1985). Notes: incorporating crown ratio into taper equations for loblolly pine trees. Forest Science 31(2): 478-484.

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

Calama R, 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 WG (1977). Sampling Techniques (3rd edn). John Wiley and Sons, New York, USA, pp. 428.

Dobler G, Peralta LE, Debord LT, Torres JG (1995). Investigación y manejo de especies maderables de uso comun en La Sierra: Guía técnica. Plan Sierra Inc., San José de las Matas, Republica Dominicana, pp. 269.

Garber SM, Maguire DA (2003). Modeling stem taper of three central Oregon species using nonlinear mixed effects models and autoregressive error structures. Forest Ecology and Management 179 (1-3): 507-522. - doi: 10.1016/S0378-1127 (02)00528-5

Gaztelurrutia M, Montero G (2001). Modelo de simulación de claras en masas de *Pinus silvestris* L. Monografías. Investigación Agraria N. 3, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, Madrid, Spain, pp. 114.

Gregoire TG, 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. - doi: 10.2307/1400563

Gregoire TG, Schabenberger O (1996b). A nonlinear mixed-effects model to predict cumulative bole volume of standing trees. Journal of Applied Statistics 23: 257-271. - doi: 10.1080/02664769 624233

Huang S, 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 TW, Kershaw JA (2002). Forest mensuration (4th edn). Wiley, New York, USA, pp. 456.

Jones RH (1990). Serial correlation or random subject effects? Communications in Statistics - Simulation and Computation 19 (3): 1105-1123. - doi: 10.1080/03610919008812907

Kozak A (1988). A variable-exponent taper equa-

tion. Canadian Journal of Forest Research 18: 1363-1368. - doi: 10.1139/x88-213

Kozak A (2004). My last words on taper equations. Forest Chronicle 80: 507-515. - doi: 10.5558/tfc80507-4

Leites LP, Robinson AP (2004). Improving taper equations of loblolly pine with crown dimensions in a mixed-effects modeling framework. Forest Science 50: 204-212.

Lindstrom MJ, Bates DM (1990). Nonlinear mixed-effects models for repeated measures data. Biometrics 46: 673-687. - doi: 10.2307/2532087

Littell RC, Henry PR, Ammerman CB (2006). Statistical analysis of repeated measures data using SAS procedures. Journal of Animal Science 76: 1216-1231. [online] URL: http://www.journalofanimalscience.org/content/76/4/1216

Montalvo JM, Bueno SW, Torres JG (2001). Ecuación para el cálculo de volumen del árbol para *Pinus occidentalis* Sw. en el Plan Sierra,

Republica Dominicana. Universidad Autonoma Chapingo, Mexico. Revista Chapingo, Serie ciencias forestales y del ambiente 7 (001): 49-53. [online] URL: http://redalyc.uaemex.mx/pdf/629/62970107.pdf

Perez DN, Burkhart HE, Stiff CT (1990). A variable-form taper function for *Pinus oocarpa* Schiede in central Honduras. Forest Science 36 (1): 186-191.

SAS Institute Inc (2006). SAS System for regression (3th edn). SAS Institute Inc., Cary, NC, USA. pp. 157.

Schabenberger O, Pierce FJ (2002). Contemporary statistical models for the plant and soil sciences. CRC Press, Boca Raton, FL, USA, pp. 738.

Tasissa G, Burkhart HE (1998). Modelling thinning effects on ring specific gravity of loblolly pine (*Pinus taed*a L.). Forest Science 44 (2): 212-223.

Trincado G, Burkhart HE (2006). A generalized

approach for modeling and localizing stem profile curves. Forest Science 52 (6): 670-682.

Van der Schaaf CL, Burkhart HE (2007). Comparison of methods to estimate Reineke's maximum size-density relationship species boundary line slope. Forest Science 53 (3): 435-442.

West PW, Ratkowsky DA, Davis AW (1984). Problems of hypothesis testing of regressions with multiple measurements from individual sampling units. Forest Ecology and Management 7: 207-224. - doi: 10.1016/0378-1127(84)90068-9

Yang Y, Huang S, Trincado G, Meng SX (2009). Nonlinear mixed-effects modeling of variable-exponent taper equations for lodgepole pine in Alberta, Canada. European Journal of Forest Research 128: 415-429. - doi: 10.1007/s10342-009-0286-2