

A simple stem taper model with mixed effects for boreal black spruce

Geneviève Lejeune · Chhun-Huor Ung ·
Mathieu Fortin · Xiao Jing Guo ·
Marie-Claude Lambert · Jean-Claude Ruel

Received: 1 October 2007 / Revised: 16 March 2009 / Accepted: 5 June 2009 / Published online: 8 July 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract We simplified Kozak's taper model by setting the inflection point at 1.3 m (dbh) without losing accuracy and precision. The simplification was required to facilitate the estimation of the covariance parameters when using a mixed-effects method. This method was necessary to take into account the correlation among multiple diameter measurements on an individual stem. The simple stem taper model was fitted to an extended data set collected across the province of Quebec, Canada. Comparison of the predicted stem taper and the derived stem volume with those obtained using existing models showed a comparable predictive power for the simple model. Including a prediction of the tree random effects based on supplementary diameter measurements of the bole improves the predictive ability of the model around the extra diameter observation. This model offers welcome simplicity as a means of predicting tree taper at coarse resolution for planning tree harvesting.

Keywords Random effects · Black spruce · Stem taper · Variable-exponent taper equation

Introduction

Knowledge of tapering in a stem is of great value to primary log breakdown because it affects grade yield. Moreover, in the context of integrating timber harvest with product recovery, it is necessary to develop mathematical models that predict stem taper from (1) tree-level covariates such as dbh, height, and crown length, and (2) stand attributes such as those available from forest polygon maps. For tree level, several models have been proposed with more or less detail to assess the generality of taper equations as well as the effect of diverse tree attributes on taper variation. So far no study has investigated the simplification of taper equations for facilitating the quantification of current and eventual stand-level attributes on the variation of stem form. This simplification approach is justified for planning tree harvesting.

In Canada, boreal black spruce (*Picea mariana* (Mill.) B.S.P.) is a major commercial species. The use of taper equations can improve estimates of standing timber value and help schedule harvesting as a function of its current market demand. However, the existing stem taper equations are limited in their extended applications by the small number of trees used for their parameterization and by their inability to take into consideration the correlation among repeated measurements on individual trees (e.g., Newnham 1988; Bonnor and Boudewyn 1990; Sharma and Zhang 2004). Within this context, the objective of this paper was to develop a more general taper equation, which would apply over larger areas of black spruce stands. This was done by combining data sets from past studies into a larger

Communicated by T. Seifert.

G. Lejeune · X. J. Guo · J.-C. Ruel
Faculté de Foresterie et de Géomatique, Université Laval,
Quebec, QC G1K 7P4, Canada

C.-H. Ung (✉)
Natural Resources Canada, Canadian Wood Fibre Centre,
Laurentian Forestry Centre, 1055 du P.E.P.S., P.O. Box 10380,
Stn. Sainte-Foy, Quebec, QC G1V 4C7, Canada
e-mail: cung@rncan.gc.ca

M. Fortin · M.-C. Lambert
Ministère des Ressources Naturelles et de la Faune du Québec,
Direction de la recherche forestière, 2700 Einstein,
Quebec, QC G1P 3W8, Canada

database. The correlations among the data were taken into account through a mixed-effects model. Moreover, we tested the effect of measuring additional diameters along the bole over the predictive ability of the model by predicting the random effects using measured upper diameters as pre information. Like existing taper equations for black spruce, the mathematical formulation of stem taper was based on Kozak's variable-exponent equation (1988).

Data

We used data from six studies containing 947 trees and covering all ecological domains where black spruce is present: the western and eastern black spruce-moss forest, and the western and eastern balsam fir-yellow birch forest (MRNF 2003). Since the data set was largely sufficient for the parameterization, 30% of the trees from each ecological domain were randomly kept apart for validation. dbh and height ranges are shown in Table 1 for partitions and study sites. Large ranges of dbh and heights were covered, and these ranges were similar for the parameterization and validation partitions.

Ouellet (1983) measured outside bark diameter from the stump up to merchantable height (9-cm diameter outside the bark). Between both limits, outside bark diameter was measured at 0.15, 0.80, and 1.30 m, at 0.05 m below the first live whorl as well as at each third of merchantable height. The same sampling was used in Ung and Ouellet (1993) and in Bonnet and Pastor (1997). In Ung (1990), Beaumont et al. (1999) and Ruel et al. (2003), outside bark diameter was measured at 0.15, 1.30, and thereafter, at each meter up to the first live whorl. Also, diameter measurements were taken at each whorl up to the tip of the

crown. Rycabel (personal communication, 2002) used a more regular pattern with measurements at 0.15, 0.65, 1.15, 1.30, 2.15 m, and thereafter, at each meter up to the apex. For each sampled tree, total height and dbh (measured at 1.3 m) were also recorded.

Method

Three steps were followed: simplification of Kozak's equation, model specification, and model verification with wood volume comparison.

Model simplification

Let i and j be the tree and the along-bole measurement position indices, respectively, such that $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, m_i$. Kozak's original equation can then be represented as follows:

$$y_{ij} = A \cdot dbh_i^B \left(\frac{1 - \sqrt{g_{ij}}}{1 - \sqrt{p}} \right)^{C_{ij}} + \varepsilon_{ij} \quad (1a)$$

$$g_{ij} = h_{ij}/ht_i \quad (1b)$$

where y_{ij} is the inside bark diameter for section j in tree i , dbh_i is the outside bark diameter at breast height, p is the relative height of the inflection point, h_{ij} is the height of the measurement, ht_i is the total height of the tree, ε_{ij} is the random error term, and A , B and C_{ij} are parameters to be estimated. The inflection point is where the stem profile changes from neiloid to paraboloid, and its relative height is thought to be constant within a given species, regardless of tree size (Demaerschalk and Kozak 1977). Parameters A and B , which deal with the conversion from outside to inside bark diameter, are assumed to be constant within a given species, and can be estimated with any appropriate data set (e.g., Garber and Maguire 2003). Parameter C_{ij} represents the core of model 1a as it makes it possible to fit the various forms all along the bole. In fact, the bole form is considered to vary according to the relative height of the section (g_{ij}) and tree characteristics.

By only addressing outside bark diameter, which is required for wood product estimates at the preharvest planning stage and during harvest for optimum bucking, Eq. 1a can be simplified by dropping not only A but also B . Further simplification was carried out on term C_{ij} , which represents the variable exponent. In the original model, C_{ij} involved a complex linear function that included seven parameters and incorporated g_{ij} under many forms, such as linear, square, inverse, square root, and logarithmic. First, C_{ij} was relocated as an exponent of both g_{ij} and p . Then, we reduced the linear function to three parameters as follows:

Table 1 Total height and dbh ranges in the database

Partitions and sites	Number of trees	dbh (cm)	Height (m)
Parameterization			
Chibougamau-Quévillon	437	7.7–32.9	5.7–22.8
Saint-Camille	14	8.6–23.0	9.7–17.5
Alma	20	6.6–25.0	8.9–19.2
Rouyn-Noranda	127	8.8–32.2	8.7–25.1
Baie-Comeau	45	9.9–26.4	9.7–19.5
Validation			
Chibougamau-Quévillon	212	6.3–31.1	7.2–23.1
Saint-Camille	5	11.4–25.2	11.7–19.4
Alma	12	7.7–24.2	9.1–18.5
Rouyn-Noranda	51	10.1–30.2	9.0–25.3
Baie-Comeau	24	9.7–26.5	8.8–22.3

$$y_{ij} = dbh_i \left(\frac{1 - g_{ij}^{C_{ij}}}{1 - p_i^{C_{ij}}} \right) + \varepsilon_{ij} \tag{2a}$$

$$C_{ij} = a + b \cdot g_{ij} + c \cdot g_{ij}^2 \tag{2b}$$

$$p_i = \frac{1.3}{ht_i} \tag{2c}$$

where a , b and c are the general parameters of the model. Through Eq. 2c, we lose the classical notion of inflection point by setting it at breast height, or 1.3 m. Although this is a strong assumption, preliminary trials revealed no major bias. Like most stem taper equations, Eq. 2a is mathematically consistent. At the apex of the tree, diameter y_{ij} converges to 0. At breast height, the ratio involving g_{ij} and p_i yields 1 and y is equal to dbh_i plus an error term so it is equal to dbh in expectation.

Model specification

When parameterizing a stem taper model, regular statistical estimators, such as ordinary or nonlinear least squares, rely on the assumptions of independently and normally distributed error terms with homogeneous variances (Steel et al. 1997, § 7.10). Besides, stem taper data are often autocorrelated as we can reasonably assume that diameter measurements are not entirely independent within the same tree. Departures from the assumption of independent errors result in biased statistical inferences (Sullivan and Clutter 1972), which hinder the selection of appropriate explanatory variables (Gregoire et al. 1995). Over the last three decades, the mixed-effects method has become increasingly popular in the forestry literature for this kind of analysis. This method makes it possible to relax the assumption of homogeneous variances and independent error terms. Recently, it has been used for modeling stem taper (e.g., Garber and Maguire 2003; Leites and Robinson 2004; Trincado and Burkhart 2006).

In general terms, model 2a can be expressed as follows:

$$y_{ij} = f(\boldsymbol{\beta}, \boldsymbol{\delta}_i, \mathbf{x}_{ij}) + \varepsilon_{ij} \tag{3}$$

where \mathbf{x}_{ij} is a vector of explanatory variables; $\boldsymbol{\delta}_i$ is a vector of unobserved tree random effects for tree i ; $\boldsymbol{\beta}$ is a vector of unknown fixed-effect parameters; and ε_{ij} is the residual error term. The vector of tree random effects and the vector of within-tree residual error terms (ε_i) are both assumed to be multivariate normally distributed, i.e., $\boldsymbol{\delta}_i \stackrel{i.i.d.}{\sim} \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma})$ and $\varepsilon_i \stackrel{i.i.d.}{\sim} \text{MVN}(\mathbf{0}, \mathbf{R}_i)$. The variance–covariance matrix of the within-tree error terms (\mathbf{R}_i) is usually set to $\mathbf{R}_i = \sigma^2 \mathbf{I}_{m_i}$ under the assumption of independence. However, the distribution of ε_i becomes multivariate when \mathbf{R}_i is assumed to have a correlation structure.

Due to the nonlinear form of Eq. 3, estimating the variance–covariance matrix $\boldsymbol{\Sigma}$ for the tree random effects is tedious. One way to obtain a more convenient form is to approximate model 3 through a first-degree Taylor expansion around the random effects (Pinheiro and Bates 2000, p. 312):

$$y_{ij} = f(\mathbf{x}_{ij}, \boldsymbol{\delta}_i, \boldsymbol{\beta}) + \varepsilon_{ij} \cong f(\mathbf{x}_{ij}, \mathbf{0}, \boldsymbol{\beta}) + \mathbf{z}_{ij} \boldsymbol{\delta}_i + \varepsilon_{ij} \tag{4}$$

$$\mathbf{z}_{ij} = \left. \frac{\partial f(\mathbf{x}_{ij}, \boldsymbol{\delta}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\delta}_i^T} \right|_{\boldsymbol{\delta}_i = \mathbf{0}}$$

where \mathbf{z}_{ij} is a matrix of partial derivatives of the model with respect to the random effects.

From Eq. 4, the variance for the vector of within-tree diameters can be estimated as:

$$\text{var}(\mathbf{y}_i) = \mathbf{Z}_i \boldsymbol{\Sigma} \mathbf{Z}_i^T + \mathbf{R}_i \equiv \mathbf{V}_i \tag{5a}$$

$$\mathbf{Z}_i = \begin{bmatrix} \mathbf{z}_{i1} \\ \mathbf{z}_{i2} \\ \dots \\ \mathbf{z}_{im_i} \end{bmatrix} \tag{5b}$$

where matrix \mathbf{Z}_i consists of the appropriate \mathbf{z}_{ij} stacked and T is a matrix transposition. In addition to random effects, a correlation structure and a variance function can be specified for \mathbf{R}_i . Details on correlation structures and variance functions are available in Davidian and Giltinan (1995), Littell et al. (2006), and Pinheiro and Bates (2000, § 5.3.3). In a general context, the covariance matrix \mathbf{R}_i can be expressed as (Pinheiro and Bates 2000, p. 205):

$$\mathbf{R}_i = \boldsymbol{\Gamma}_i^{1/2} \boldsymbol{\Phi}_i \boldsymbol{\Gamma}_i^{1/2} = \begin{bmatrix} \sigma_{i1}^2 & \rho_{i12} \sigma_{i1} \sigma_{i2} & \dots & \rho_{i1m_i} \sigma_{i1} \sigma_{im_i} \\ \rho_{i21} \sigma_{i2} \sigma_{i1} & \sigma_{i2}^2 & \dots & \rho_{i2m_i} \sigma_{i2} \sigma_{im_i} \\ \dots & \dots & \dots & \dots \\ \rho_{i1m_i} \sigma_{im_i} \sigma_{i1} & \rho_{i2m_i} \sigma_{im_i} \sigma_{i2} & \dots & \sigma_{im_i}^2 \end{bmatrix} \tag{6}$$

where $\boldsymbol{\Gamma}_i$ is a diagonal matrix whose elements are the residual variances σ_{ij}^2 and $\boldsymbol{\Phi}_i$ is a correlation matrix whose off-diagonal elements are the correlations $\rho_{ijj'}$.

In this case study, the parameters of model 2a were expressed through a mixed model adjustment as linear functions of one or many fixed effects as well as of the single tree random effect:

$$a_i = \beta_1 + \beta_2 \frac{ht_i}{dbh_i} + \beta_3 ht_i + \delta_{1,i} \tag{7a}$$

$$b_i = \beta_4 + \delta_{2,i} \tag{7b}$$

$$c_i = \beta_5 + \beta_6 \frac{ht_i}{dbh_i} + \delta_{3,i} \tag{7c}$$

where β_1 , β_2 , β_3 , β_4 , β_5 , and β_6 are the elements of the column vector of fixed-effect parameters ($\boldsymbol{\beta}$), and $\delta_{1,i}$, $\delta_{2,i}$, and $\delta_{3,i}$ are the elements of the vector of tree-level random effects ($\boldsymbol{\delta}_i$). We assumed that this vector had an

unstructured covariance structure, i.e., there was no structure specified for the 3×3 matrix Σ . Moreover, matrix \mathbf{R}_i was modeled through a first-order autoregressive correlation structure and a power-of-the-mean variance function, which provided the best fit. The power-of-the-mean variance function and the continuous autoregressive correlation structure that were fitted are defined as follows:

$$\text{var}(\varepsilon_{ij}|\delta_i) = \sigma_{ij}^2 = \sigma^2 (E(y_{ij}|\delta_i))^{2\theta} \tag{8a}$$

$$\text{corr}(\varepsilon_{ij}, \varepsilon_{ij'}|\delta_i) = \rho_{ijj'} = \rho^{|h_{ij}-h_{ij'}|} \tag{8b}$$

where $E(y_{ij}|\delta_i)$ is the mean predicted value conditional on the tree random effects, and σ^2 , θ and ρ are parameters to be estimated. Note that Eqs. 8a and 8b completely specify matrices Γ_i and Φ_i in Eq. 6. The adequacy of the random effects, the variance function, and the correlation structure were assessed through likelihood ratio tests and parsimony criteria (AIC and BIC).

The parameterization of nonlinear mixed models relies on likelihood estimation. Because the maximization of the likelihood function involves a complex integral, some approximations have been proposed in the statistical literature. Some of them consist in linearizing the model through a first-order Taylor expansion around the expected value zero of the random effects as shown in Eq. 4 and applying the linear mixed-model theory (e.g., Lindstrom and Bates 1990; Vonesh and Carter 1992). Model (2a) with all the aforementioned features was parameterized using the nlme package (Pinheiro et al. 2008) available in R software (R Core team 2008), which is based on this type of approximation.

Model validation

In addition to likelihood ratio tests and parsimony criteria (AIC and BIC), normalized residuals were used to check the fit of the model. Considering the variance–covariance matrix of the error terms, a vector of within-tree normalized residuals (\mathbf{r}_i) can be approximated from the linear mixed-model theory (cf., Pinheiro and Bates 2000, p. 239) as

$$\mathbf{r}_i = \left(\hat{\mathbf{V}}_i^{1/2}\right)^T (\mathbf{y}_i - \mathbf{f}(\mathbf{X}_i, \mathbf{0}, \hat{\boldsymbol{\beta}})) \tag{9}$$

Normalized residuals can be seen as standardized residuals for mixed models because they account for the modeled covariances. If the mixed model is properly parameterized with regard to its covariance features, we can expect the elements of \mathbf{r}_i to be independently and normally distributed with mean 0 and variance 1, i.e., $\mathbf{r}_i \stackrel{\text{i.i.d.}}{\sim} N(\mathbf{0}, \mathbf{I})$ (SAS Institute 2008). These normalized residuals were used to check residual correlations.

Once model 2a was parameterized, mean predictions were generated for each observation of the validation partition. Because the random effects do not occur linearly in the model, these predicted values did not correspond to the mathematical expectation for the population, i.e., $E[y_{ij}] \neq f(\mathbf{x}_{ij}, \mathbf{0}, \hat{\boldsymbol{\beta}})$. In fact, the population-averaged predictions, i.e., the mathematical expectation of y_{ij} without prior knowledge of the random effects, were obtained by integrating the marginal prediction of the model with respect to the random effects (e.g., McCulloch and Searle 2001, p. 65). In this case study, these population-averaged predictions would be obtained through $E[y_{ij}] = \int f(\mathbf{x}_{ij}, \delta_i, \hat{\boldsymbol{\beta}}) \cdot \text{pr}(\delta_i) \cdot d\delta_i$. However, there was no closed-form solution to this integral and therefore, it was computed numerically.

Let \hat{y}_{ij} be the population-averaged predicted values for section j of tree i . The validity of the model was assessed through the average bias and the root mean square error (RMSE), which were computed as follows:

$$\text{Bias} = \frac{\sum_{i=1}^n \sum_{j=1}^{m_i} (y_{ij} - \hat{y}_{ij})}{\sum_{i=1}^n m_i} \tag{10a}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n \sum_{j=1}^{m_i} (y_{ij} - \hat{y}_{ij})^2}{\sum_{i=1}^n m_i}}. \tag{10b}$$

Relative biases and RMSE were obtained by dividing Eqs. 10a and 10b by the average observed diameter. Average biases and RMSE were calculated by classes of tenths of relative height as is done in most of the studies on the stem profile modeling (e.g., Garber and Maguire 2003; Sharma and Zhang 2004). Moreover, the trees were grouped into 5-cm dbh classes to check if bias was related to tree size.

Predicted stem tapers were also used to compute tree-level volume estimates. Estimated and observed volumes for trees of the validation data set were compared with assessed volume prediction bias. Observed volumes were calculated using two outside bark round-end diameters: a 0-cm limit, which provided the total volume, and a 9.1-cm limit, which is considered as the merchantable limit in the province of Quebec. The volumes were calculated with Smalian’s formula (Avery and Burkhart 1983, p. 30). Biases and RMSE were then computed.

Finally, like Lappi (2006) and Kublin et al (2008), the simplified model was also used to evaluate the number and best location of supplementary diameter measurements. According to the mixed model theory (Littell et al. 2006), additional diameters along the bole can be considered as prior knowledge of the stem taper, and can be used to predict the vector of tree random effects as:

$$\hat{\boldsymbol{\delta}}_i = \hat{\Sigma}_i^T \hat{\mathbf{V}}_i^{-1} (\mathbf{y}_i - \mathbf{f}(\mathbf{X}_i, \mathbf{0}, \hat{\boldsymbol{\beta}})).$$

Using the predicted $\hat{\delta}_i$ is known to reduce the variance of the model predictions (e.g., Hall and Bailey 2001). To test this effect in this case study, the validation data set was divided into ten relative height classes again. Like Trincado and Burkhart (2006), the root mean square error (RMSE) within each relative height class was obtained using the predicted random effect parameter. The evaluation was conducted with two different scenarios. The first used only one supplementary diameter at absolute heights of 0.0001 m (stump height), 3.5, 7, and 10 m. The second used two supplementary diameters at absolute heights of 3.5 m, with 7 m and at 3.5 m with 10 m.

Results

We parameterized the model several times keeping the fixed-effect specification constant. We first included the variance function, then the random effects and finally the correlation structure. According to AIC and BIC statistics, every additional covariance feature, i.e., the random effects, the variance function, and the correlation structure significantly improved the maximum likelihood of the model (Table 2). The maximum log-likelihood (LLK) value increased from -6,237 for the model including the variance function only to -4,175 for the complete model (model 5 in Table 2).

The normalized residuals were normally distributed with homogeneous variance. Empirical correlations were computed for height classes along the bole. The residuals exhibited low but persistent correlations for some distances between the diameter measurements (Fig. 1). Most correlations did not exceed the [-0.2, 0.2] range. The parameter estimation yielded the following parameter estimates (all significant at $\alpha = 0.01$):

$$\hat{a}_i = 0.8918 + 0.1575 \frac{ht_i}{dbh_i} - 0.0244ht_i + \delta_{1,i} \tag{11a}$$

$$\hat{b}_i = 2.9929 + \delta_{2,i} \tag{11b}$$

$$\hat{c}_i = -2.6326 + 1.0313 \frac{ht_i}{dbh_i} + \delta_{3,i} \tag{11c}$$

Table 2 Maximum log-likelihood (LLK) and information criteria (AIC and BIC) of the different models

Model	Covariance features	LLK	AIC	BIC
1	Variance function	-6,237	12,489	12,542
2	Model 1 + random effect $\delta_{1,i}$	-5,748	11,513	11,573
3	Model 2 + random effect $\delta_{2,i}$	-4,876	9,775	9,848
4	Model 3 + random effect $\delta_{3,i}$	-4,492	9,012	9,105
5	Model 4 + correlation structure	-4,175	8,380	8,479

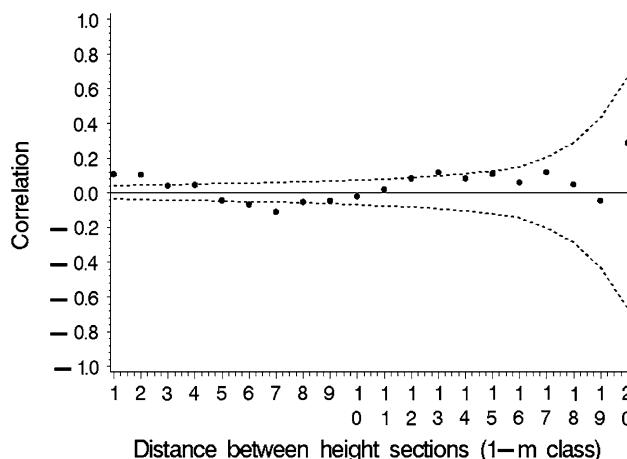


Fig. 1 Empirical correlations against distance classes between height sections along the bole (dashed lines represent the 95% confidence intervals under the null hypothesis $H_0: \text{Corr}(r_{ij}, r_{ij'}) = 0$)

$$\hat{\Sigma} = \begin{bmatrix} 0.0116 & -0.0575 & 0.0554 \\ -0.0575 & 0.6669 & -0.7280 \\ 0.0554 & -0.7280 & 1.0325 \end{bmatrix} \tag{11d}$$

$$\hat{\sigma}^2 = 0.0706 \tag{11e}$$

$$\hat{\theta} = 0.2204 \tag{11f}$$

$$\hat{\rho} = 0.3356 \tag{11g}$$

Note that the parameter estimates in Eqs. 11e, 11f, and 11g refer to the correlation structure and the variance function. The 95% confidence intervals of the fixed-effect parameters are annexed to this paper.

The estimates of the fixed-effect parameters and the variance-covariance matrix Σ were used in turn to generate a population-averaged mean predicted stem profile for each tree of the validation data set. The biases and RMSE were calculated for each 5-cm dbh class. Results are shown in Table 3. Two diameter classes (2.5–7.5 and 27.5–32.5 cm) were omitted from this table because they had too few trees.

The model tends to underestimate the diameters in the lower and upper sections, whereas the mid-section diameters are overestimated. For sections below 0.8 in relative height, most biases are smaller than or close to 0.4 cm in absolute value, and the RMSE ranges from 0.2 to 1.1 cm (Table 3). These biases are relatively low, with absolute relative values ranging from 0 to 6.5% which are similar to those obtained by Garber and Maguire (2003) and Sharma and Zhang (2004). For the upper sections, the biases are larger and range from 0.5 to 0.7 cm like those obtained by Garber and Maguire (2003) and Sharma and Zhang (2004). The pattern is similar for all diameter classes. With an average bias of 0.1 cm for all the validation data pooled, the model predictions can be considered nearly unbiased on average (Table 3).

Table 3 Biases and root mean square errors (RMSE) of the mean stem profile predictions with respect to the validation data set (relative biases and RMSE between parentheses)

Diameter class and relative height	Number of observations	Bias (cm)	RMSE (cm)
7.5–12.5 cm			
0.0 < g_{ij} < 0.1	118	-0.17 (-1.4%)	0.77 (6.4%)
0.1 < g_{ij} < 0.2	54	0.08 (0.8%)	0.27 (2.5%)
0.2 < g_{ij} < 0.3	40	0.07 (0.7%)	0.25 (2.6%)
0.3 < g_{ij} < 0.4	41	0.00 (0.0%)	0.36 (3.8%)
0.4 < g_{ij} < 0.5	33	-0.02 (-0.3%)	0.32 (3.7%)
0.5 < g_{ij} < 0.6	22	-0.01 (-0.1%)	0.39 (5.0%)
0.6 < g_{ij} < 0.7	24	0.13 (1.9%)	0.50 (7.4%)
0.7 < g_{ij} < 0.8	23	0.21 (4.0%)	0.52 (9.6%)
0.8 < g_{ij} < 0.9	17	0.40 (9.8%)	0.48 (11.8%)
0.9 < g_{ij} < 1.0	14	0.20 (13.3%)	0.30 (20.0%)
Overall	386	0.01 (0.1%)	0.52 (5.6%)
12.5–17.5 cm			
0.0 < g_{ij} < 0.1	253	0.06 (0.3%)	0.73 (4.4%)
0.1 < g_{ij} < 0.2	80	0.10 (0.8%)	0.44 (3.3%)
0.2 < g_{ij} < 0.3	105	0.20 (1.5%)	0.51 (3.8%)
0.3 < g_{ij} < 0.4	96	-0.08 (-0.7%)	0.60 (5.1%)
0.4 < g_{ij} < 0.5	115	-0.06 (-0.5%)	0.66 (5.8%)
0.5 < g_{ij} < 0.6	87	-0.12 (-1.3%)	0.59 (6.0%)
0.6 < g_{ij} < 0.7	80	0.05 (0.6%)	0.59 (6.7%)
0.7 < g_{ij} < 0.8	53	0.47 (6.3%)	0.73 (9.9%)
0.8 < g_{ij} < 0.9	54	0.61 (11.9%)	0.76 (14.8%)
0.9 < g_{ij} < 1.0	24	0.45 (20.5%)	0.48 (21.7%)
Overall	947	0.10 (0.8%)	0.64 (5.3%)
17.5–22.5 cm			
0.0 < g_{ij} < 0.1	222	-0.05 (-0.2%)	1.01 (4.5%)
0.1 < g_{ij} < 0.2	52	0.18 (1.0%)	0.69 (3.7%)
0.2 < g_{ij} < 0.3	139	0.30 (1.7%)	0.70 (4.0%)
0.3 < g_{ij} < 0.4	67	0.12 (0.7%)	0.64 (4.0%)
0.4 < g_{ij} < 0.5	124	-0.07 (-0.5%)	0.72 (5.1%)
0.5 < g_{ij} < 0.6	63	0.10 (0.7%)	0.88 (6.6%)
0.6 < g_{ij} < 0.7	75	-0.22 (-2.1%)	0.83 (8.1%)
0.7 < g_{ij} < 0.8	92	0.27 (3.0%)	0.71 (8.0%)
0.8 < g_{ij} < 0.9	76	0.72 (12.1%)	1.07 (18.1%)
0.9 < g_{ij} < 1.0	46	0.61 (23.6%)	0.76 (29.3%)
Overall	956	0.14 (1.0%)	0.84 (5.6%)
22.5–27.5 cm			
0.0 < g_{ij} < 0.1	53	-0.18 (-0.7%)	1.34 (4.8%)
0.1 < g_{ij} < 0.2	20	0.47 (2.1%)	0.92 (4.0%)
0.2 < g_{ij} < 0.3	38	0.26 (1.2%)	0.98 (4.7%)
0.3 < g_{ij} < 0.4	19	0.07 (0.4%)	0.89 (4.6%)
0.4 < g_{ij} < 0.5	26	-0.19 (-1.1%)	0.97 (5.7%)
0.5 < g_{ij} < 0.6	26	-0.39 (-2.6%)	0.66 (4.3%)
0.6 < g_{ij} < 0.7	14	-0.37 (-2.9%)	0.71 (5.6%)
0.7 < g_{ij} < 0.8	37	0.07 (0.7%)	0.79 (8.2%)
0.8 < g_{ij} < 0.9	30	0.41 (6.1%)	0.61 (9.1%)

Table 3 continued

Diameter class and relative height	Number of observations	Bias (cm)	RMSE (cm)
0.9 < g_{ij} < 1.0	29	0.50 (18.3%)	0.74 (27.2%)
Overall	292	0.07 (0.4%)	0.94 (5.8%)
All diameter classes			
0.0 < g_{ij} < 0.1	660	-0.03 (-0.2%)	0.93 (4.9%)
0.1 < g_{ij} < 0.2	212	0.18 (1.2%)	0.57 (3.7%)
0.2 < g_{ij} < 0.3	330	0.26 (1.6%)	0.68 (4.3%)
0.3 < g_{ij} < 0.4	230	0.05 (0.3%)	0.68 (5.0%)
0.4 < g_{ij} < 0.5	306	-0.07 (-0.5%)	0.70 (5.4%)
0.5 < g_{ij} < 0.6	207	-0.07 (0.6%)	0.70 (6.0%)
0.6 < g_{ij} < 0.7	200	-0.09 (-0.9%)	0.71 (7.4%)
0.7 < g_{ij} < 0.8	215	0.24 (2.8%)	0.74 (8.9%)
0.8 < g_{ij} < 0.9	183	0.59 (10.4%)	0.87 (15.3%)
0.9 < g_{ij} < 1.0	113	0.50 (20.7%)	0.66 (27.4%)
Overall	2,656	0.10 (0.8%)	0.76 (5.7%)

The volume estimates show average biases that range from 0.1 to 3.2 dm³ in absolute values (Table 4). These biases are relatively small, representing less than 1% of the average observed volumes. RMSE tends to increase with the diameter. However, the relative root mean square error is rather constant, ranging from 5.4 to 10%. For most diameter classes, bias and RMSE are slightly larger for the merchantable volume.

When only one supplementary diameter measurement was considered at stump height (0.0001 m), 3.5, 7, and

Table 4 Biases and root mean square errors (RMSE) of volume estimates with respect to the validation data set (relative biases and RMSE between parentheses)

Diameter class and volume	Number of trees	Bias (dm ³)	RMSE (dm ³)
7.5–12.5 cm			
Total volume	50	-0.1 (-0.2%)	3.2 (5.5%)
Merchantable volume		-0.2 (-0.5%)	3.6 (10.0%)
12.5–17.5 cm			
Total volume	120	0.3 (0.2%)	8.5 (6.5%)
Merchantable volume		0.3 (0.3%)	9.2 (8.1%)
17.5–22.5 cm			
Total volume	104	1.3 (0.5%)	16.7 (6.3%)
Merchantable volume		1.6 (0.6%)	17.2 (6.9%)
22.5–27.5 cm			
Total volume	25	-3.2 (-0.8%)	25.6 (6.2%)
Merchantable volume		-2.8 (-0.7%)	25.5 (6.4%)
Overall			
Total volume	304	0.8 (0.4%)	14.9 (7.6%)
Merchantable volume		0.9 (0.5%)	15.3 (8.6%)

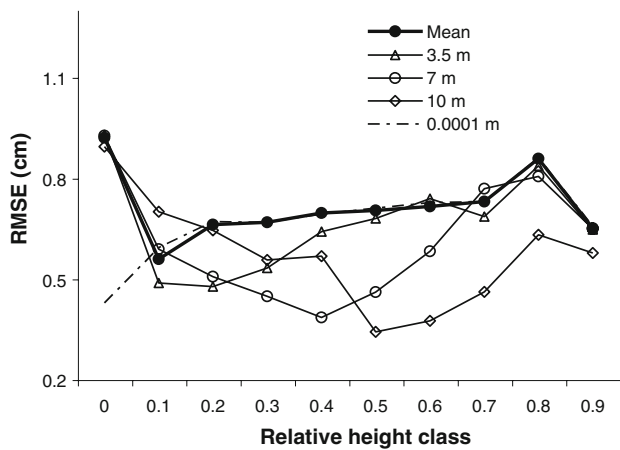


Fig. 2 Root mean square error [RMSE (cm)] for the mean and calibrated responses using one supplementary diameter measurement

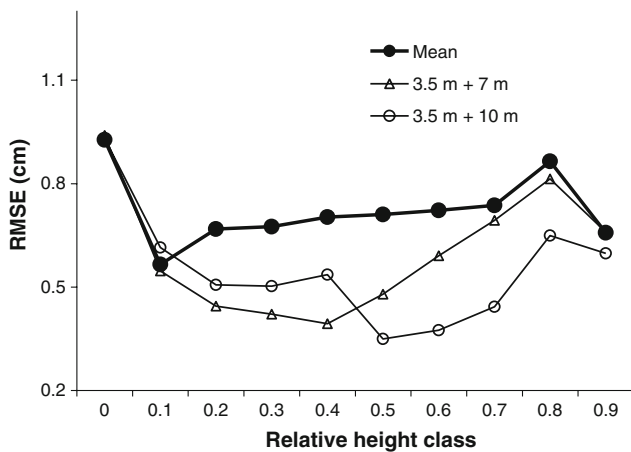


Fig. 3 Root mean square error [RMSE (cm)] for the mean and calibrated responses using two supplementary diameter measurements

10 m each time, Fig. 2 shows that: (1) the corrected curve is better than the mean response, (2) the correction cannot improve the accuracy of the whole stem; instead, the part around the supplementary diameter measurement is improved significantly, (3) the correction can hardly improve the prediction if the supplementary diameter is measured at the very bottom of the bole. Thus, the best placement depends on which part of the stem is considered to be more important. When two supplementary diameter measurements were considered, Fig. 3 shows that: (1) in general, the prediction is better than supplemented by only one extra diameter measurement, and (2) compared with the mean response, the improvement is less significant.

Discussion

The stem taper equation obtained represents a simplified version of Kozak’s original variable-exponent equation. It

has three general parameters when the outside bark diameter is considered as the dependent variable. The use of random effects with a variance function and a correlation structure represents two advantages over the ordinary least square estimators. First, if the covariance features are adequate, the estimators for the vector of fixed-effect parameters and the variance components of the model are thought to be unbiased, although this property is derived from linear theory and has not been extended to the non-linear case (Pinheiro and Bates 2000, p. 314). Therefore, the appropriate explanatory variables either at the tree scale or the stand scale can be selected according to their significance level. In this case study, tree height and the height/dbh ratio (h_i/dbh_i) were significant variables in the prediction of stem taper.

Second, the mixed-effects method ensures some consistency among model predictions through the estimation of distinct error components. In this case study, four error components were found to be significant. Three were related to tree level, i.e., the three random effects ($\delta_{1,i}$, $\delta_{2,i}$, and $\delta_{3,i}$), whereas the last one was the residual error, which is associated with diameter measurements (ε_{ij}). The result obtained is more satisfying than what would be obtained using least square estimators in which the single error term is assumed to be entirely independent from one section to the other along the bole. Actually, such a least square-based model is likely to have a lower maximum log-likelihood when compared with a model with a variance function only (model 1 in Table 2), which in turn has a much lower maximum log-likelihood when compared with the proposed model (model 5 in Table 2).

An alternative to the simplification of Kozak’s equation is a linearization through a logarithmic transformation, which reduces the magnitude of the numerical problem of estimating its parameters. However, the logarithmic transformation induces a bias when the predicted values have to be back-transformed (Duan 1983). Moreover, the log-transformed diameter did not respect the assumption of normality in this case study. For these reasons, the non-linear approach was preferable.

In terms of accuracy, the proposed model shows biases that are similar in range to those of other studies. Biases in the upper sections are particularly large, as in Garber and Maguire (2003) and Sharma and Zhang (2004). In our context, the low reliability in the upper bole section can be explained by the cylindrical assumption whose effect is exaggerated by the lack of diameter measurements in the upper bole sections. Also, the model is constrained to pass through the tip of the tree, and to be equal to the dbh at 1.3 m. Consequently, there is less flexibility near these points than anywhere else along the bole. However, the effects of these upper section biases on volume calculation are small. Actually, biases on the estimated volumes are

relatively low (Table 4), and the numerical integration of the taper function yields a reliable mean prediction for the different volumes of a particular tree.

The mathematical consistency of the model, i.e., $y_{ij} = 0$ when $h_{ij} = ht_i$ and $y_{ij} = dbh_i$ when $h_{ij} = 1.3$, brings up some considerations about the heteroscedastic pattern of the residuals. In fact, the variance of the prediction error increases with the increase in the predicted value. This statement is logical, since we can reasonably expect the diameters at but swell to be more variable than those at the tip of the tree. However, because of the mathematical consistency of the model, the variance of the prediction error should decrease as sections get closer to breast height. For instance, there should not be a great variability in diameter predictions 10 cm above or below breast height. This statistical constraint has not been taken into account in this case study in a direct manner because it would have required a complex variance function, while a good compromise between complexity and efficiency is required for operational applications.

The evaluation of the supplementary diameter measurements showed that estimating the tree random effects improves the predictive capability of the simplified model. The corrected curve is better than the mean response. However, the correction cannot improve the accuracy of the whole stem; instead, the part around the extra diameter measurement is improved significantly. So the best placement depends on which part of the stem is considered more important. The prediction can be minimally improved by adding diameter measurements at both ends of the bole.

The model does not include any plot covariates because information at the plot level was not available. Consequently, the model relies on the assumption of within-plot tree independence. This can be a strong assumption, considering that some studies have already demonstrated that some plot attributes may have an effect on stem taper (e.g., Garber and Maguire 2003; Sharma and Zhang 2004). We also tested some plot random effects in the model, but convergence could not be achieved because of the complexity of the covariance features. However, the empirical correlation (an indirect evaluation of this plot random effect) was close to 0 (analysis not shown), indicating that the variance of the potential stand-level random effects might be low in this case study. This low random effect may be explained by the inclusion of the height/dbh ratio in the model. The ratio is a surrogate of tree density at the plot scale.

Conclusion

We have shown that a simplified version of Kozak's equation can be applied to stem taper data of boreal black spruce using the mixed-effects method with a proposed correlation structure. The simplified equation was obtained by fixing the inflection point at dbh. The proposed correlation structure makes it possible to apply the mixed-effects method for adequately addressing the variability of stem taper at the tree scale. The low validation error based on the wide gradient of collected black spruce stem taper data provides the first evidence that this simplified variable-exponent taper equation can reliably predict black spruce tree taper over large areas. Related to the mean response, predicting the random effects through additional diameter measurements improves the predictive capability of the obtained model around the extra diameter observation.

To further improve the model, the first step will be to develop a variance function that is more consistent with the model. This function should induce a decrease in the variance as the sections get closer to breast height. The second step will be to link tree scale with plot scale. Plot scale is represented by forest polygon attributes based on aerial photos or remote sensing. Also it is necessary to clarify the limits of applicability of the proposed taper equation because the assumption of the model may be violated in other species and at other observations. Nevertheless, the model offers welcome simplicity as a means of predicting tree taper at coarse resolution from standard tree attribute measurements and from eventual polygon attributes available on any forest map that can be used for the primary log breakdown.

Acknowledgments The first author received financial support from NSERC, Grant No. 234774. Dr. Steen Magnussen of the Canadian Forest Service helped to make the paper more concise. The comments and suggestions of two anonymous reviewers have contributed to clarifying its scope.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix

The 95% confidence intervals of the parameters are given in the following table (Table 5).

Table 5 The 95% confidence intervals of the parameters

Parameters	Estimates	95% confidence intervals
a_1	0.8918	[0.8133, 0.9703]
a_2	0.1575	[0.0895, 0.2256]
a_3	-0.0244	[-0.0270, -0.0219]
b_1	2.9929	[2.9058, 3.0800]
c_1	-2.6326	[-3.0134, -2.2518]
c_2	1.0313	[0.6302, 1.4323]
$\text{var}(\delta_{1,i})$	0.0116	[0.0100, 0.0140] ^a
$\text{var}(\delta_{2,i})$	0.6669	[0.5350, 0.8312] ^a
$\text{var}(\delta_{3,i})$	1.0325	[0.8043, 1.3255] ^a
$\text{cov}(\delta_{1,i}, \delta_{2,i})$	-0.0575	[-0.0653, -0.0478] ^b
$\text{cov}(\delta_{1,i}, \delta_{3,i})$	0.0554	[0.0390, 0.0690] ^b
$\text{cov}(\delta_{2,i}, \delta_{3,i})$	-0.7280	[-0.7551, -0.6920] ^b
σ^2	0.0706	[0.0577, 0.0863] ^a
θ	0.2204	[0.1929, 0.2868]
ρ	0.3356	[0.2978, 0.3757]

^a Confidence intervals derived from confidence intervals on the standard deviations

^b Confidence intervals derived from confidence intervals on the correlations

References

- Avery TE, Burkhart HE (1983) Forest measurements, 3rd edn. McGraw-Hill Book Company, Princeton, NJ
- Beaumont JF, Ung C-H, Bernier-Cardou M (1999) Relating site index to ecological factors in black spruce stands: tests of hypotheses. For Sci 45(4):484–491
- Bonnet P-A, Pastor M (1997) Utilisation de la dimension fractale du houppier pour la vérification de la loi d'auto-éclaircie. Rapport de stage non publié. Centre de foresterie des Laurentides, Sainte-Foy, Québec
- Bonnor GM, Boudewyn P (1990) Taper-volume equations for major tree species of the Yukon territory. Information Report BC-X-323, Forestry Canada, Pacific Forestry Centre, Victoria, BC
- Davidian M, Giltinan DM (1995) Nonlinear models for repeated measurement data. Chapman & Hall, Boca Raton, FL
- Demaerschalk JP, Kozak A (1977) The whole-bole system: a conditioned dual-equation system for precise prediction of tree profiles. Can J Res 7:488–497. doi:10.1139/x77-063
- Duan N (1983) Smearing estimate: a nonparametric retransformation method. J Am Stat Assoc 78:605–610. doi:10.2307/2288126
- Garber SM, Maguire DA (2003) Modeling stem taper of three central Oregon species using nonlinear mixed effects models and autoregressive error structures. For Ecol Manage 179:507–522
- Gregoire TG, Schabenberger O, Barrett JP (1995) Linear modelling of irregularly spaced, unbalanced, longitudinal data from permanent-plot measurements. Can J Res 25:137–156
- Hall DB, Bailey RL (2001) Modeling and prediction of forest growth variables based on multilevel nonlinear mixed models. For Sci 47(3):311–321
- Kozak A (1988) A variable-exponent taper equation. Can J Res 18:1363–1368. doi:10.1139/x88-213
- Kublin E, Augustin NH, Lappi J (2008) A flexible regression model for diameter prediction. Eur J For Res 127:415–428. doi:10.1007/s10342-008-0225-7
- Lappi J (2006) A multivariate, non parametric stem-curve prediction method. Can J Res 36:1017–1027. doi:10.1139/X05-305
- Leites LP, Robinson AP (2004) Improving taper equations of loblolly pine with crown dimensions in a mixed-effects modeling framework. For Sci 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, Milliken GA, Stroup WW, Wolfinger RD, Schabenberger O (2006) SAS for mixed models, 2nd edn. SAS Institute Inc, Cary, NC
- McCulloch CE, Searle SR (2001) Generalized, linear, and mixed models. Wiley, New York
- MRNF (2003) Vegetation zones and bioclimatic domains in Québec. Gouvernement du Québec. Ministère des Ressources naturelles et de la Faune. Available at: <http://www.mrn.gouv.qc.ca/english/forest/quebec/quebec-environment-zones.jsp>
- Newnham RM (1988) A variable-form taper function. Information Report PI-X-83, Canadian Forestry Service, Petawawa National Forest Institute, Petawawa, ON
- Ouellet D (1983) Biomass equations for black spruce in Quebec. Information Report LAU-X-60E, Canadian Forestry Service, Laurentian Forestry Centre, Sainte-Foy, QC
- Pinheiro JC, Bates DM (2000) Mixed-effects models in S and S-PLUS. Springer, New York
- Pinheiro JC, Bates DM, DebRoy S, Sarkar D, the R Core team (2008) The nlme package. Available at: <http://cran.stat.sfu.ca/>. Revised on June 23, 2008
- R Core team (2008) The R project for statistical computing. Available at: <http://www.r-project.org/>. Revised on June 23, 2008
- Ruel J-C, Horvath R, Ung C-H, Munson A (2003) Comparing height growth and biomass production of black spruce trees in logged and burned stands. For Ecol Manage 193:371–384
- SAS Institute (2008) SAS online doc 9.1.3. SAS Institute Inc. Available at: <http://support.sas.com/onlinedoc/913/docMainpage.jsp> (last revised on July 31, 2008)
- Sharma M, Zhang SY (2004) Variable-exponent taper equations for jack pine, black spruce, and balsam fir in eastern Canada. For Ecol Manage 198:39–53
- Steel RGD, Torrie JH, Dickey DA (1997) Principles and procedures of statistics: a biometrical approach. McGraw-Hill, New York
- Sullivan AD, Clutter JL (1972) A simultaneous growth and yield model for loblolly pine. For Sci 18:76–86
- Trincado G, Burkhart HE (2006) A generalized approach for modeling and localizing stem profile curves. For Sci 52:670–682
- Ung C-H (1990) Tarifs de cubage paramétrés: application à l'épinette noire de Lebel-sur-Quévillon. Can J Res 20:1471–1478. doi:10.1139/x90-194
- Ung C-H, Ouellet D (1993) Croissance et branchaison de l'épinette noire *Picea mariana* (Mill.) B.S.P. In: Ung C-H (ed) Les modèles de croissance forestière et leurs utilisations. Colloque international tenu les 18 et 19 novembre 1993, Centre municipal des Congrès, Québec (Canada) Canadian Forest Service, Laurentian Forestry Centre, Sainte-Foy, Q, pp 45–57
- Vonsh EF, Carter RL (1992) Mixed-effects nonlinear regression for unbalanced repeated measures. Biometrics 48:1–17. doi:10.2307/2532734