NRC Research Press

ARTICLE

A general method for assessing the effects of uncertainty in individual-tree volume model predictions on large-area volume estimates with a subtropical forest illustration

Ronald E. McRoberts, Paolo Moser, Laio Zimermann Oliveira, and Alexander C. Vibrans

Abstract: Forest inventory estimates of tree volume for large areas are typically calculated by adding the model predictions of volumes for individual trees at the plot level, calculating the mean over plots, and expressing the result on a per unit area basis. The uncertainty in the model predictions is generally ignored, with the result that the precision of the large-area volume estimate is optimistic. The primary study objective was to assess the performance of a Monte Carlo based approach for estimating model prediction error that had been developed for boreal and temperate forest applications when used for a subtropical forest application. Monte Carlo simulation approaches were used because of the complexities associated with multiple sources of uncertainty, the nonlinear nature of the models, and heteroskedasticity. A related objective was to estimate the effects of model prediction uncertainty due to residual and parameter uncertainty on the large-area volume estimates for the Brazilian state of Santa Catarina. The primary conclusions were fourfold. First, the methodological approach worked well. Second, the effects of model residual and parameter uncertainty on large-area estimates of mean volume per unit area were negligible for the models and calibration datasets used for the study. Third, for the models currently in use in Santa Catarina, the effects of model residual and parameter uncertainty may be ignored when calculating large-area estimates of mean volume per unit area. Fourth, differences were negligible between estimates of the mean and standard error obtained using a single, nonspecific volume model and estimates obtained using both forest-type models and species-specific/species-group models.

Key words: allometric model, residual uncertainty, parameter uncertainty, Monte Carlo simulation, Santa Catarina, Brazil.

Résumé : L'estimation du volume des arbres d'un inventaire forestier pour un grand territoire est typiquement calculée en additionnant les prédictions, faites par un modèle, du volume des arbres individuels à l'échelle de la placette, en calculant la moyenne des placettes et en exprimant le résultat par unité de surface. L'incertitude de la prédiction du modèle est généralement ignorée avec pour résultat que la précision du volume de ce grand territoire est optimiste. Un des premiers objectifs de cette étude était d'évaluer la performance d'une approche utilisant la méthode Monte-Carlo pour estimer l'erreur de prédiction d'un modèle, développé pour être utilisé dans des forêts boréales et tempérées, lorsqu'il est utilisé en forêt subtropicale. L'approche des simulations Monte-Carlo a été utilisée en raison de la complexité associée à des sources d'incertitudes multiples, de la nature non-linéaire des modèles et de l'hétéroscédasticité. Un objectif connexe était d'estimer les effets sur l'estimation du volume pour un grand territoire, de l'État brésilien de Santa Catarina, de l'incertitude associée à la prédiction du modèle causée par l'erreur des paramètres et l'erreur résiduelle. Les principales conclusions ont été au nombre de quatre. Premièrement, l'approche méthodologique a bien fonctionné. Deuxièmement, les effets de l'erreur des paramètres et de l'erreur résiduelle sur l'estimation du volume moyen par unité de surface pour un grand territoire se sont avérés négligeables pour les modèles et les ensembles de données d'étalonnage utilisées dans cette étude. Troisièmement, pour les modèles actuellement en usage dans l'État de Santa Catarina, les effets de l'erreur résiduelle et de l'erreur des paramètres du modèle peuvent être ignorés lorsque l'on estime le volume moyen par unité de surface pour ce grand territoire. Quatrièmement, les différences étaient négligeables entre les estimations de la moyenne et de l'erreur standard obtenues en utilisant un seul modèle de volume, non spécifique, et les estimations obtenues en utilisant à la fois des modèles de type forestier et des modèles d'espèces individuelles ou de groupes d'espèces. [Traduit par la Rédaction]

Mots-clés : modèle allométrique, incertitude résiduelle, incertitude des paramètres, simulation Monte Carlo, Santa Catarina, Brésil.

1. Introduction

Forest inventory and monitoring programs routinely use statistical models to predict the volume, biomass, or carbon content for individual trees using measurements of tree attributes such as species, diameter, and height as predictor variables. The individual-tree model predictions are aggregated at the plot level, and plot-level estimates are averaged over plots to produce large-area estimates, often expressed on a per unit area basis. The uncertainty in these model predictions is routinely ignored, with the result that precision estimates are erroneously optimistic; the only question is the degree of optimism.

Uncertainty in model predictions can be attributed to four primary sources: (*i*) model misspecification, (*ii*) uncertainty in the values of the independent variables, (*iii*) residual variability around model predictions, and (*iv*) uncertainty in the model parameter estimates. Model misspecification is due to the lack of

Received 5 June 2014. Accepted 13 August 2014.

R.E. McRoberts. Northern Research Station, USDA Forest Service, Saint Paul, MN 55108, USA.
P. Moser, L. Zimermann Oliveira, and A.C. Vibrans. Universidade Regional de Blumenau, Blumenau, Santa Catarina 89030-000, Brazil.
Corresponding author: Ronald E. McRoberts (e-mail: rmcroberts@fs.fed.us).

appropriate model calibration data and (or) the lack of modeling expertise. The effects of uncertainty from this source are not considered in this study because model misspecification is typically not a problem for allometric volume models. The effects of uncertainty in values of the predictor variables, whether the result of measurement error or simply observer-to-observer variability, have been studied extensively (Gertner 1990; McRoberts et al. 1994; Berger et al. 2014) and, therefore, are also not considered in this study. Rather, this study focused only on the effects of model residual and parameter uncertainty. For a correctly specified model, residual variability is typically estimated as root mean square error, and the uncertainty in model parameter estimates is often expressed by the model parameter covariance matrix.

Multiple approaches have been used to propagate residual and parameter uncertainty to model predictions and large-area estimates. Cunia (1965, 1987) and Ståhl et al. (2014) used sampling theory and theoretical derivations of expressions for the magnitudes of uncertainty from the different sources to estimate their effects on the uncertainty of large-area estimates. Gertner (1990) and Berger et al. (2014) used an approach based on first-order Taylor series approximations to produce an additive expression for total uncertainty of the estimate of interest. Monte Carlo simulations have often been used to circumvent the mathematical and computational difficulties involved when uncertainty from multiple sources must be propagated through complex models (Gertner and Dzialowy 1984; McRoberts and Westfall 2014; Breidenbach et al. 2014).

The primary study objectives were twofold: (*i*) to assess the utility of a method for estimating the effects of model prediction error that had been developed for boreal and temperate forest applications when used for a subtropical forest application, and (*ii*) to estimate the effects on large-area volume estimates for the Brazilian state of Santa Catarina. The technical objectives were to estimate the particular effects of model residual and parameter uncertainty for species-specific/species-group, forest-type, and nonspecific allometric volume models. A Monte Carlo approach was used with models that were constructed specifically for this study so that uncertainty in model residual and parameter uncertainty could be rigorously quantified. In addition, estimates of mean volume per unit area obtained using these models constructed with local data were compared with estimates obtained with two pantropical models.

2. Data

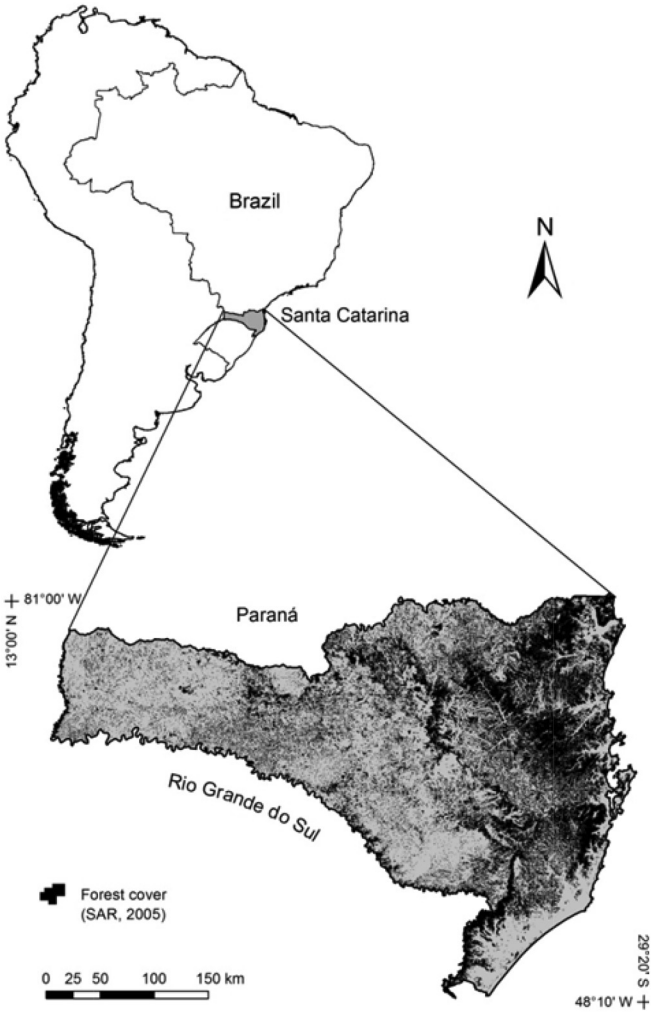
2.1. Study area

The study area was defined as the southern Brazilian state of Santa Catarina with an area of 95 346 km^2 and located between latitudes 26°S and 29°S and between longitudes 48°W and 53°W (Fig. 1). The Santa Catarina Forest and Floristic Inventory (IFFSC) defines forest land with respect to the following four criteria: (*i*) minimum area of 0.5 ha, (*ii*) continuous tree vegetation, (*iii*) minimum canopy height of 10.0 m, and (*iv*) minimum age of 15 years. Forest land covers 27.7% of Santa Catarina and includes three phytogeographic classes (Klein 1978): dense ombrophylous forest (DEN) with an area of 31 281 km^2, of which 12 633 km^2 is forest land; mixed ombrophylous forest with *Araucaria angustifolia* (Bertol.) Kuntze (MIX) with an area of 56 395 km^2, of which 12 317 km^2 is forest land; and seasonal deciduous forest (DEC) with an area of 7671 km^2, of which 1251 km^2 is forest land (Vibrans et al. 2013*a*; table 3).

2.2. Estimation dataset

Sample plot data from the IFFSC were used to estimate large-area mean volume per unit area. Sample plots were configured as clusters of four crosswise 1000 m^2 subplots (20 m × 50 m) and were located at the intersections of a 10 km × 10 km grid for DEN and MIX and at the intersections of a 5 km × 5 km grid for DEC (Vibrans et al. 2010). A total of 1074 plots was established and measured between 2007 and 2010. Ground conditions for 418 of these plots satisfied the IFFSC definition of forest land. For all trees with a diameter at breast height (dbh; 1.3 m) of at least 10.0 cm on forest-land plots, IFFSC field crews obtained observations of species, measurements of dbh using a tape measure, and visual estimates of height calibrated using heights measured with a hypsometer as references. Because relationships among volume, dbh, and height for *A. angustifolia* differ substantially from relationships for other species, and because calibration data for this species were not available, the species was omitted from all analyses. For the DEN and DEC classes, the effects were negligible because of the small numbers of plots with this species. However, for the MIX class, *A. angustifolia* represented as much as 15% of the volume. For future reference, the data described in this section are characterized as the estimation dataset because they constitute the source of information for estimating mean volume per unit area.

Fig. 1. State of Santa Catarina, Brazil, with black indicating forest cover.



2.3. Calibration dataset

Data for selected trees from the same plots as described in section 2.2 were used to construct models of relationships between individual-tree stem volume and tree attributes. For each subplot, IFFSC field crews collected additional data for one to eight trees with a dbh of at least 10 cm. Overall, trees were selected to represent the largest possible number of species and the largest possible number of diameter classes (Table 1). Tree-level data in-

Table 1. Model calibration datasets.

Phytogeographical class	Species	Species code	Sample size	DBH range (cm)	R^2 (ln–ln scale)	R^{2*} (original scale)
DEC (58 species)	*Nectandra megapotamica*	103	34	17.19–44.56	0.98	0.96
	Ocotea puberula	108	72	17.83–56.66	0.97	0.98
	All other species		182	18.46–84.35	0.96	0.95
	All species		288	17.19–84.35	0.97	0.95
DEN (205 species)	*Cedrela fissilis*	23	47	15.92–52.20	0.98	0.97
	Ocotea puberula	108	30	15.92–37.88	0.93	0.98
	Piptocarpha angustifolia	120	36	16.55–43.93	0.98	0.97
	Nectandra oppositifolia	183	37	14.96–42.97	0.98	0.98
	Alchornea triplinervia	197	40	13.37–49.34	0.95	0.93
	Virola bicuhyba	213	46	17.51–63.03	0.97	0.93
	Hieronyma alchorneoides	580	38	18.14–44.88	0.97	0.98
	Miconia cinnamomifolia	693	100	16.55–44.25	0.94	0.93
	Tapirira guianensis	696	32	15.60–50.29	0.89	0.83
	All other species		797	11.46–73.53	0.95	0.94
	All species		1203	11.46–73.53	0.93	0.91
MIX (104 species)	*Cedrela fissilis*	23	30	18.78–53.48	0.94	0.94
	Clethra scabra	32	47	17.51–49.02	0.87	0.89
	Matayba elaeagnoides	82	33	20.37–43.48	0.91	0.93
	Ocotea puberula	108	59	17.19–54.11	0.98	0.98
	Prunus myrtifolia	123	34	10.82–59.52	0.98	0.98
	All other species		425	10.82–76.71	0.95	0.96
	All species		628	10.82–76.71	0.95	0.96
Total (367 species)	All species		2119	10.82–84.35	0.94	0.92

Note: Class abbreviations: DEC, seasonal deciduous forest; DEN, dense ombrophylous forest; MIX, mixed ombrophylous forest. DBH is diameter at breast height.

cluded species, dbh measured using a tape, and total height measured using a hypsometer. In addition, for each selected tree, climbers measured circumferences at heights above the ground at 0.3 m, 1.0 m, 1.3 m, 2.0 m, and thereafter at 1.0 m intervals until the beginning of the crown (Vibrans et al. 2010). Stem volumes were calculated using Smalian's formula (Avery and Burkhart 2002) as the sum over sections of the product of mean crosssectional area and section length and were characterized as volumes over bark to the top of the stems. For future reference, these data are characterized as the calibration dataset because they were used to construct and calibrate the individual-tree volume models.

3. Methods

3.0. Volume models

Allometry refers to relationships between size and shape, and allometric models are often used to characterize biological phenomena. Allometric models are often of the general form:

$$Y = \beta_0 X_1^{\beta_1} \ldots X_p^{\beta_p} + \varepsilon \quad (1)$$

where Y denotes the response variable, $\mathbf{X}$ denotes a vector of predictor variables, $\boldsymbol{\beta}$ is a vector of parameters to be estimated, and ε is a random residual error. Allometric forms for individual-tree volume and biomass models have been used for boreal and temperate applications in China (Xiang et al. 2011), Europe (Zianis et al. 2005), the United States of America (Jenkins et al. 2003), and Norway (Breidenbach et al. 2014) and extensively for tropical applications (Brown et al. 1989; Chave et al. 2005; Basuki et al. 2009; Návar 2009).

For this study, an allometric model of the relationship between individual-tree stem volume (V) as the response variable for the ith tree and dbh and height (ht) as the predictor variables was formulated as

$$V_i = \beta_0 \cdot \mathrm{dbh}_i^{\beta_1} \cdot \mathrm{ht}_i^{\beta_2} + \varepsilon_i \quad (2)$$

where the βs are as defined for eq. 1, and ε_i is a random residual with mean 0. Before model fitting, natural logarithmic (ln) transformations of the response and predictor variables were calculated, and the model was reformulated as

$$\ln(V_i) = \alpha_0 + \alpha_1 \cdot \ln(\mathrm{dbh}_i) + \alpha_2 \cdot \ln(\mathrm{ht}_i) + \varepsilon_i \quad (3)$$

where the αs are the parameters to be estimated. Note that the ε_i values of eq. 3 are not necessarily the same as the ε_i values of eq. 2. The advantages of the transformations were that the model could be expressed in linear form, which facilitates estimation of the parameters, and heteroskedasticity was removed, thereby eliminating the necessity of weighted regressions. On the original scale, predictions were calculated as

$$\hat{V}_i = \exp\left[\hat{\alpha}_0 + \hat{\alpha}_1 \cdot \ln(\mathrm{dbh}_i) + \hat{\alpha}_2 \cdot \ln(\mathrm{ht}_i) + \frac{\hat{\sigma}_{\mathrm{res}}}{2}\right] \quad (4)$$

where $\hat{\sigma}_{\mathrm{res}}$ is the residual standard deviation on the ln–ln scale, and the term $\hat{\sigma}_{\mathrm{res}}/2$ compensates for bias that accrues when transforming from the ln–ln scale back to the original scale (Baskerville 1972). For each of the DEN, MIX, and DEC forest-type classes, species-specific models were constructed for individual species with at least 30 calibration observations, and species-group models were constructed for the aggregation of observations for species with fewer than 30 observations; forest-type models were constructed for aggregations of observations for the DEN, MIX, and DEC classes; and a nonspecific model was constructed for the aggregation of all observations.

The quality of model fit on the ln–ln scale was assessed using R^2 calculated as

$$R^2 = \frac{\sum_{i=1}^{n_{\mathrm{cal}}} (V_i - \bar{V})^2 - \sum_{i=1}^{n_{\mathrm{cal}}} (V_i - \hat{V}_i)^2}{\sum_{i=1}^{n_{\mathrm{cal}}} (V_i - \bar{V})^2} \quad (5)$$

where n_{cal} is the number of elements in the calibration dataset; V_i and $\hat{V}_i$ are the observation and model prediction, respectively, for the ith tree in the calibration dataset; and $\overline{V}$ is the mean over all n_{cal} observations. For application to the nonlinear model of eq. 2 on the original scale, the metric of eq. 5 was also calculated, but it was denoted R^{2*} and was characterized as a pseudo R^2 because the assumptions underlying R^2 are not completely satisfied when using nonlinear models (Anderson-Sprecher 1994). Quality of fit was also assessed by graphing observations versus model predictions. In the absence of a lack of fit of the model to the data, the points should lie along the 1:1 line, with intercept 0 and slope 1.

3.1. Stratified estimation

Stratified estimation is required when different portions of the population are sampled with different intensities; for this study, DEN and MIX were sampled using a 10 km × 10 km grid, and DEC was sampled using a 5 km × 5 km grid. Stratified estimation can also be used to improve precision by using strata that separate population units into more homogeneous groups. For this study, the DEC portion of the population was used as a separate stratum to accommodate its greater sampling intensity, and the DEN and MIX portions of the population were separated into individual strata as a potential means of increasing precision.

The essence of stratified estimation is to assign population units to strata, calculate within-strata sample plot means and variances, and then calculate the population estimate as a weighted average of the within-strata estimates, in which the weights are proportional to the strata sizes. Stratified estimation requires accomplishment of two tasks: (*i*) calculation of the strata weights as the relative proportions of the population area corresponding to strata, and (*ii*) assignment of each plot to a single stratum on the basis of the stratum assignment of the plot center. For this study, the population was the portion of the study area that satisfied the IFFSC definition of forest land. Because systematic sampling was used in the study area, the sample size per stratum was considered random rather than fixed, as would be the case for stratified sampling. Therefore, the following poststratified estimators that accommodated this randomness were used:

$$(6a) \qquad \hat{\mu} = \sum_{h=1}^{H} w_h \hat{\mu}_h$$

and

$$(6b) \qquad \hat{\mathrm{Var}}(\hat{\mu}) = \sum_{h=1}^{H} \left[w_h \cdot \frac{\hat{\sigma}_h^2}{n} + (1 - w_h) \cdot \frac{\hat{\sigma}_h^2}{n^2} \right]$$

where

$$(6c) \qquad \hat{\mu}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{h_i}$$

$$(6d) \qquad \hat{\sigma}_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{h_i} - \hat{\mu}_h)^2$$

n is the total sample size; $h = 1, \ldots, H$ denotes strata; and for the hth stratum, y_{h_i} is the observation of the ith plot, w_h is the weight, and $\hat{\mu}_h$ and $\hat{\sigma}_h^2$ are the sample estimates of the within-strata means and variances, respectively (Cochran 1977).

3.2. Simulating uncertainty

This study focused on estimating the effects of model prediction uncertainty from two sources, model residual and parameter uncertainty, on estimates of large-area mean volume per unit area.

3.2.1. Residual uncertainty

Residual uncertainty was assessed on the original scale where the models were applied. For calibration datasets for which model parameters were estimated, a four-step procedure was used to estimate the heteroskedastic residual uncertainty: (*i*) the pairs $(V_i, \hat{V}_i)$ were ordered with respect to the model prediction, $\hat{V}_i$; (*ii*) the pairs were aggregated into groups of size at least 25; (*iii*) within each group, g, the mean of the observations, $\overline{V}_g$, the mean of the predictions, $\overline{\hat{V}}_g$, and the standard deviation, $\hat{\sigma}_g$, of the residuals, $\varepsilon_i = V_i - \hat{V}_i$, were calculated; and (*iv*) the relationship between the group standard deviations and the group prediction means was represented using the following model:

$$(7) \qquad \hat{\sigma}_g = \gamma_1 \cdot \overline{\hat{V}}_g^{\gamma_2} + \varepsilon_g$$

where the γs are parameters to be estimated.

3.2.2. Parameter uncertainty

Parameter uncertainty for linear models depends on residual uncertainty, the calibration dataset sample size, and the distribution of the values of the independent variables. For this study, uncertainty in the linear model parameter estimates on the ln–ln scale was assessed using a three-step Monte Carlo approach: (*i*) the transformed dataset was aggregated into $c = 2$, 5, and 10 dbh size classes; (*ii*) each dbh size class was resampled with replacement until the original class size, n_{class}, was achieved; and (*iii*) the model was fit to the resampled data and the parameters were estimated. Steps (*i*)–(*iii*) were replicated 5000 times. The resulting distribution of parameter estimates represented the uncertainty in estimates of the linear model parameters on the ln–ln scale.

The effects of sample size on model parameter estimates were estimated using the same three-step Monte Carlo approach with two modifications. First, resampling within each dbh class was terminated when the achieved within-class sample size reached n_{class}/τ, where $\tau = 2$, 4, 8, and 16. Second, the effects of sample size were estimated only for the forest-type and nonspecific models because most of the species-specific calibration datasets were too small to reduce further.

3.2.3. Uncertainty in large-area volume estimates

A four-step Monte Carlo simulation procedure was used to estimate the effects of uncertainty in model residual and parameter uncertainty on the uncertainty of large-area estimates of mean volume per unit area.

Step 1 For the kth replication and for each model set (species-specific and species group, forest type, nonspecific) a set of model parameter estimates, $\hat{\beta}^k$, was randomly selected from the joint distribution constructed in section 3.2.2.

Step 2 For the ith tree on the jth plot in the estimation dataset, a volume observation was simulated using the parameter estimates from step 1 as

$$(8) \qquad V_{ij}^k = \hat{\beta}_0^k \cdot d_{ij}^{\hat{\beta}_1^k} \cdot h_{ij}^{\hat{\beta}_2^k} + \varepsilon_{ij}$$

where ε_{ij} is a residual randomly selected from a Gaussian $(0, \hat{\sigma}_i)$ distribution for which $\hat{\sigma}_i$ was calculated using eq. 7 with $\hat{\beta}_0^k \cdot d_{ij}^{\hat{\beta}_1^k} \cdot h_{ij}^{\hat{\beta}_2^k}$ as the value of the predictor variable.

Step 3 The total volume for the jth plot in the estimation dataset was calculated as $V_j^k = \sum_{i=1}^{n_j} V_{ij}^k$ where n_j is the number of trees on the plot.

Step 4 The mean and variance of the mean over all plots for the kth replication were estimated as

$$\bar{V}^k = \frac{1}{n_{\text{plot}}} \sum_{j=1}^{n_{\text{plot}}} V_j^k \quad (9)$$

and

$$\hat{\text{V}}\text{ar}(\bar{V}^k) = \frac{1}{n_{\text{plot}}(n_{\text{plot}} - 1)} \sum_{j=1}^{n_{\text{plot}}} (V_j^k - \bar{V}^k)^2 \quad (10)$$

where n_{plot} is the number of plots.

Steps 1–4 were replicated, and the mean and variance over replications were estimated as per Rubin (1987)

$$\hat{\mu} = \frac{1}{n_{\text{rep}}} \sum_{k=1}^{n_{\text{rep}}} \bar{V}^k \quad (11)$$

and

$$\hat{\text{V}}\text{ar}(\hat{\mu}) = \left(1 + \frac{1}{n_{\text{rep}}}\right) W_1 + W_2 \quad (12)$$

where $W_1 = \frac{1}{n_{\text{rep}} - 1} \sum_{k=1}^{n_{\text{rep}}} (\bar{V}^k - \hat{\mu})^2$ is the among-simulations variance, $W_2 = \frac{1}{n_{\text{rep}}} \sum_{k=1}^{n_{\text{rep}}} \hat{\text{V}}\text{ar}(\bar{V}^k)$ is the mean within-simulation variance, and n_{rep} is the number of replications. The replications continued until $\hat{\mu}$ and $\hat{\text{V}}\text{ar}(\hat{\mu})$ stabilized.

In step 3, spatial correlations among observations in the calibration datasets for trees in close physical proximity were ignored because the proportions of such trees in the calibration datasets were very small. In addition, Berger et al. (2014), Breidenbach et al. (2014), and McRoberts and Westfall (2014) all reported that the effects of this spatial correlation were negligible.

3.3. Analyses

The four-step simulation procedure was implemented using the species-specific/species-group models, the forest-type models, and the nonspecific model. Simulation estimates were obtained both with and without the inclusion of simulated residual uncertainty and both with and without the inclusion of simulated parameter uncertainty. Simulation estimates were also obtained for different calibration dataset sample sizes, with both residual and parameter uncertainty included.

The models used for this study were constructed specifically so that the residual and parameter uncertainty could be rigorously quantified. In practice, the IFFSC uses models constructed using the same datasets but of different mathematical forms (Vibrans et al. 2013*b*, 2013*c*). Therefore, stratified estimates obtained using the models constructed for this study without the inclusion of residual and parameter uncertainty were compared with estimates obtained using the IFFSC models.

Because acquisition of appropriate data for constructing models is costly and laborious, particularly for tropical and subtropical forests, both Brown et al. (1989) and Chave et al. (2005) constructed pantropical individual-tree volume and biomass models for general application. However, few, if any, comparisons of estimates based on those models and estimates obtained based on models constructed using large, carefully acquired, local datasets are known to have been reported. Therefore, stratified estimates of mean volume per unit area obtained using the nonspecific model constructed for this study were compared with estimates based on pantropical models reported by Brown et al. (1989) and Chave et al. (2005). Because all plots in the estimation dataset for this study were in Holdridge's (1967) moist life zone, only pantropical models for moist forests were considered. In particular, model 3 for moist forests from Brown et al. (1989, table 2)

$$\ln(\text{AGB}) = \beta_0 + \beta_1 \cdot \ln(\text{dbh}^2 \cdot \text{ht}) + \varepsilon \quad (13)$$

was selected, as was model IA for moist forests from Chave et al. (2005, table 2)

$$\ln(\text{AGB}) = \beta_0 + \beta_1 \cdot \ln(\text{dbh}) + \beta_2 \cdot \ln(\text{ht}) + \beta_3 \cdot \ln(\rho) + \varepsilon \quad (14)$$

where AGB denotes aboveground biomass and ρ denotes wood specific gravity. For both models, the parameter estimates reported for these models were used, and the reported values of $\hat{\sigma}_{\text{res}}$ were used to compensate for bias when transforming back to the original AGB scale. For use in the Chave et al. (2005) model of eq. 14 and for converting from AGB to volume for both models, the mean wood specific gravity of $\rho = 0.70$ reported by Chave et al. (2006, fig. 2) for the Brazilian Atlantic forest was used.

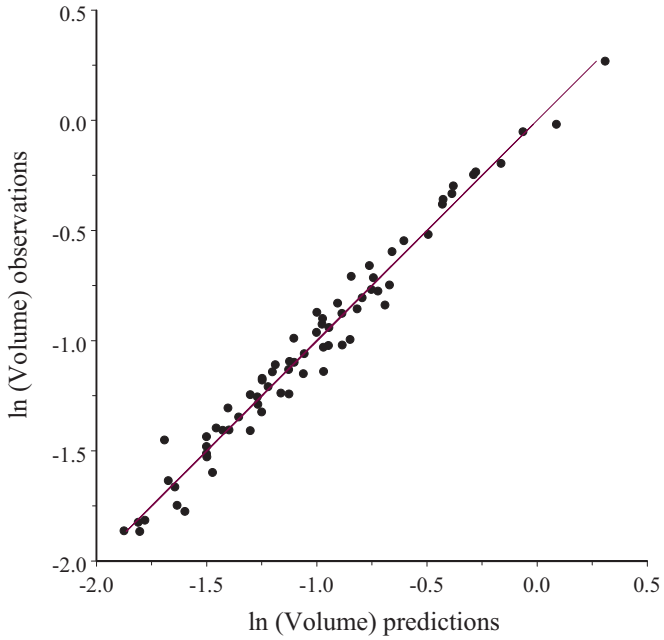


Fig. 2. Individual-tree observations versus model predictions on the ln–ln scale for *Ocotea puberula* in the dense ombrophylous phytogeographical class.

4. Results and discussion

4.1. Evaluating the method

The fits of the models to the data on both the ln–ln and original scales produced large R^2 and R^{2*} values (Table 1; Figs. 2 and 3). These values are comparable to those obtained for similar studies in different locations for different forest types (Chave et al. 2005; Zianis et al. 2005; McRoberts and Westfall 2014) and justify the decision not to include model misspecification as a source of model prediction uncertainty.

The heteroskedastic residual uncertainty was quantified using estimated relationships between residual standard deviations and volume model predictions as described in section 3.2.1. The approach was somewhat arbitrary, but the relationships were well

Fig. 3. Individual-tree observations versus model predictions on the original (m^3) scale for *Ocotea puberula* in the dense ombrophylous phytogeographical class.

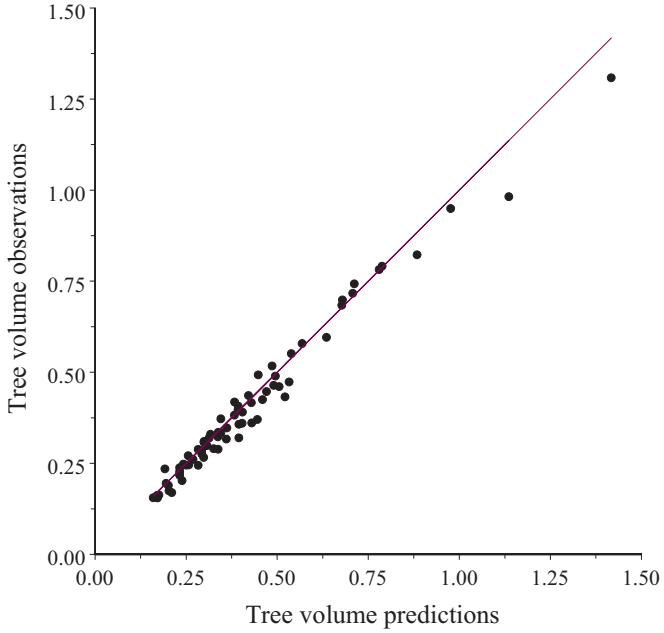
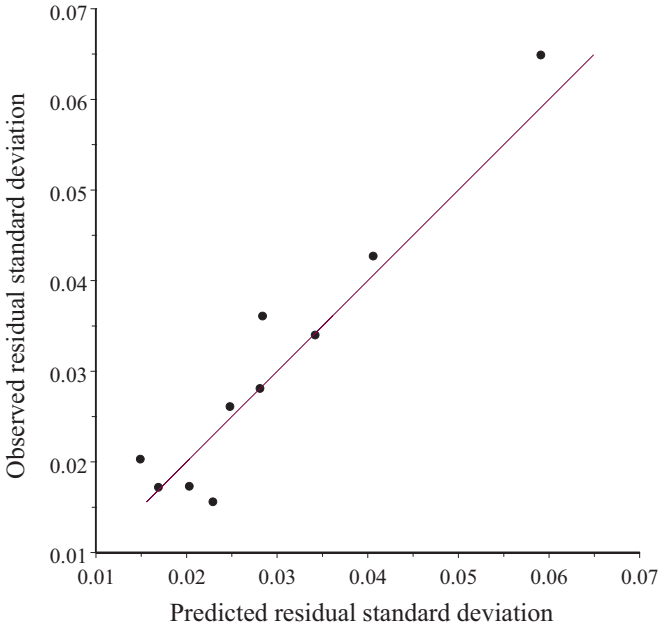


Fig. 4. Group observed versus predicted residual standard deviations on the original (m^3) scale for *Ocotea puberula* in the dense ombrophylous phytogeographical class.



estimated (Fig. 4), and estimates of the parameters of the model expressed by eq. 7 were remarkably consistent over calibration datasets.

The Monte Carlo simulation procedure for quantifying parameter uncertainty was designed to mimic the manner in which trees were selected for the calibration datasets. In particular, trees were not selected in a completely random manner, but rather were selected randomly within dbh size classes. The simulation approach divided the dbh range for each calibration dataset into 2, 5, and 10 dbh size classes and then randomly selected trees with replacement from within each class. Because differences in esti-

Fig. 5. Simulated uncertainty in model parameter estimates (β_2 and β_3) on the ln–ln scale.

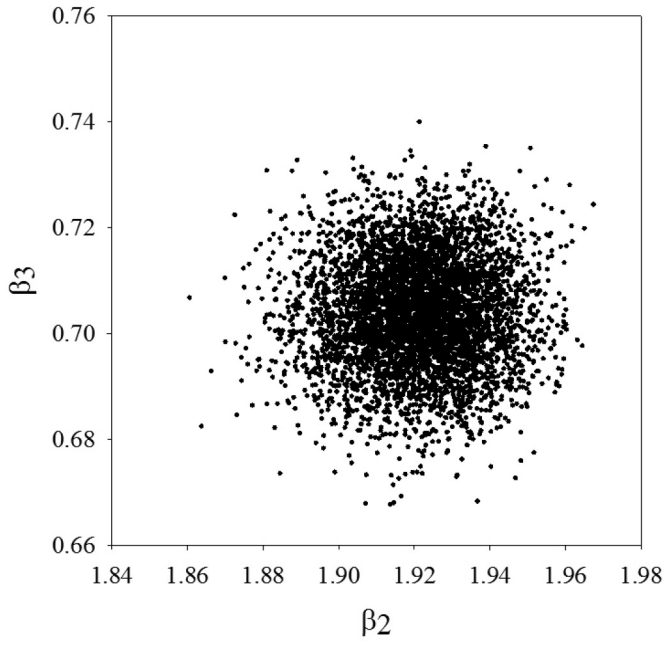


Fig. 6. Mean volume per unit area ($m^3 \cdot ha^{-1}$) and standard error of means ($m^3 \cdot ha^{-1}$) versus number of simulations.

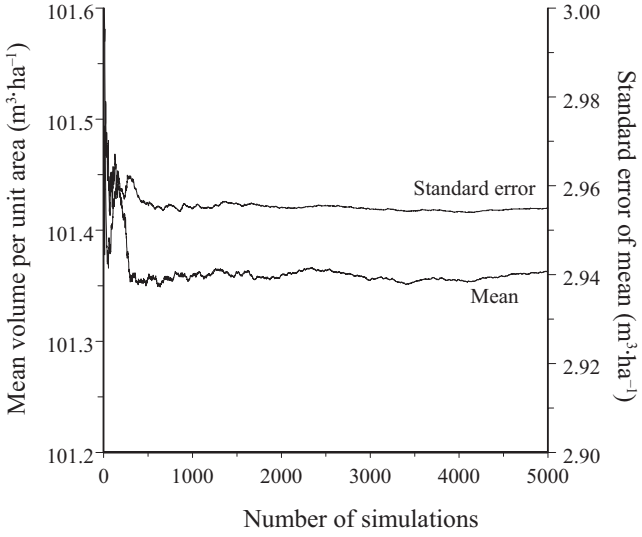


Table 2. Stratified estimates.

Stratum	Weight	Sample size	Mean	SE
DEC	0.0424	78	83.72	4.91
DEN	0.4283	197	99.81	3.87
MIX	0.5294	143	102.99	4.60
Total	1.0000	418	101.15	2.64

Note: SE is standard error.

mates of means and standard errors (SE) using different numbers of size classes were negligible, only results for 10 dbh size classes are reported. The distribution of estimates appropriately reflected parameter uncertainty (Fig. 5).

For all combinations of levels of residual and parameter uncertainty and for assessing the effects of calibration dataset sample

sizes, 5000 replications of the simulation procedure were sufficient for estimates of both means and SEs to stabilize (Fig. 6).

4.2. Stratified estimation

Stratified estimation was necessary to accommodate the unequal sampling intensities. The use of three strata with no residual and parameter uncertainty incorporated into the estimates produced $\hat{\mu}$ = 101.15 $m^3 \cdot ha^{-1}$, with SE = 2.64 $m^3 \cdot ha^{-1}$ (Table 2). The analyses with three strata split the DEN and MIX classes into separate strata for purposes of increasing precision. If these two classes had not been split and the analyses had been based on only two strata, the estimate would have been $\hat{\mu}$ = 101.41 $m^3 \cdot ha^{-1}$, with SE = 2.64 $m^3 \cdot ha^{-1}$. Therefore, the separation of the DEN and MIX classes into separate strata did not accomplish the intended purpose of decreasing the SE and thereby increasing precision. Nevertheless, analyses using three strata were retained for the remaining analyses because within-stratum estimation is facilitated with no meaningful disadvantage accrued. If the different sampling intensities had been ignored and simple random sampling rather stratified estimators had not been used, the estimates would have been $\hat{\mu}$ = 97.90 $m^3 \cdot ha^{-1}$, with SE = 2.59 $m^3 \cdot ha^{-1}$. The reason for the smaller estimate of the mean was that for the simple random sampling estimators, the effective weight for the DEC class with the small within-stratum mean was the larger plot proportion, 0.1866, rather than the smaller area proportion, 0.0424. The smaller estimate of the SE is attributed to the same reason and also to the use of the larger overall sample size rather than the smaller within-strata sample sizes.

The SEs of the stratified estimates with no residual or parameter uncertainty incorporated represented sampling variation that could not be reduced further apart from larger sample sizes or different sampling designs. The effects of residual and parameter uncertainty were assessed with respect to the effects of sampling variation. In particular, the relevant question was whether the effects of residual and parameter uncertainty were meaningful relative to sampling variation.

4.3. Effects of uncertainty

For the calibration datasets and models constructed for this study, the effects of residual and parameter uncertainty were negligible relative to sampling variability (Table 3). In addition, estimates of means and SEs were nearly indistinguishable, regardless of whether the species-specific/species-group, forest-type, or nonspecific models were used. Several factors contributed to the latter phenomenon. First, the similarity among the R^{2*} values for species-specific/species-group, forest-type, and nonspecific models indicated that the differences in the models were slight; therefore, estimates using the different models should be expected to be similar. Second, the proportion of individual trees in the estimation dataset whose volumes were predicted using species-specific models, as opposed to species-group models, was small, ranging proportionally from 0.13 to 0.18, depending on the forest type. Furthermore, because the calibration datasets for the species-group models included more than 60% of all the calibration observations for the particular forest type, the forest-type models should be expected to be quite similar to the species-group models. Thus, estimates for the forest-type models should again reasonably be expected to be similar to estimates for the species-specific/species-group models. The small differences in SEs were attributed to the same two factors. These results suggest that if only state wide volume estimates are desired, then there is little advantage to constructing and using the species-specific/species-group or the forest-type models. McRoberts and Westfall (2014) reported similar results for temperate forests in Minnesota in the USA. However, one advantage may be that species-specific/species-group models produce more accurate large-area estimates for individual species. This was the case for *Alchornea triplinervia* (Spreng.) Müll. Arg., the most common species in DEN, for which estimates of mean volume per unit area obtained using the species-specific/species-group models and the forest-type models differed proportionally by approximately 0.2.

The effects of differences in calibration dataset sample sizes were manifested in differences in parameter uncertainty. For the nonspecific calibration dataset for this study, the effects of additional parameter uncertainty resulting from reducing sample sizes by factors as large as 4.0 were negligible relative to the effects of sampling variability on estimates of means and SEs (Table 4). However, when sample sizes for the nonspecific dataset were reduced by factors of 8.0 and 16.0, estimates of the SEs increased proportionally by 0.14 and 0.33, respectively. For the forest-type calibration datasets, the effects were nearly negligible when the calibration datasets were reduced by factor of 4.0 or less. However, for reduction by factors of 8.0 and 16.0, SEs increased proportionally by 0.28 and 2.01, respectively. For both datasets, the reduction of sizes of the calibration datasets also produced substantial differences in estimates of the mean.

These results are generally similar to those from a study by McRoberts and Westfall (2014) who reported that the effects of residual and parameter uncertainty were negligible for calibration dataset sizes of 100 or greater and $R^{2*} \geq 0.95$. For the McRoberts and Westfall (2014) study, the effects of residual and parameter uncertainty were negligible when the forest-type calibration datasets were reduced by factors of 4.0, resulting in dataset sample sizes of approximately 100, 300, and 400. The larger samples sizes necessary for negligible effects for this study can at least be partially attributed to smaller forest-type R^{2*} values that were in the range $0.91 \leq R^{2*} \leq 0.96$.

4.4. Comparisons to external models

Estimates for mean volume per unit area obtained for the species-specific/species-group models constructed for this study were $\hat{\mu}$ = 101.15 $m^3 \cdot ha^{-1}$, with SE = 2.64 $m^3 \cdot ha^{-1}$, whereas estimates obtained using the IFFSC models were $\hat{\mu}$ = 100.81 $m^3 \cdot ha^{-1}$, with SE = 2.66 $m^3 \cdot ha^{-1}$. The difference between the two estimates of the mean is much less than a single SE and for practical purpose may be considered negligible. Although the mathematical forms of the models differ, the similarity in the estimates suggests that the effects of residual and parameter uncertainty would be similar. Therefore, the effects of uncertainty from these two sources may be considered to be negligible, not only for the models constructed for this study, but also for the current IFFSC models.

Estimates obtained using the nonspecific model constructed for this study were compared with estimates obtained using the Brown et al. (1989) and Chave et al. (2005) pantropical, nonspecific models for moist forests. The estimate of mean volume per unit area obtained using the nonspecific model for this study was $\hat{\mu}$ = 102.02 $m^3 \cdot ha^{-1}$, with SE = 2.67 $m^3 \cdot ha^{-1}$. Following conversion from AGB to volume, the estimate obtained using the Brown et al. (1989) model was $\hat{\mu}$ = 80.09 $m^3 \cdot ha^{-1}$, with SE = 2.31 $m^3 \cdot ha^{-1}$. The plot-level mean deviation was nearly −0.25, the overall mean and SE were underestimated proportionally by 0.13 and 0.18, respectively, and the estimate of the mean was considerably more than two SEs from the estimate for the nonspecific model for this study. Thus, the Brown et al. (1989) model would not have been an acceptable substitute for the local models. Following conversion from AGB to volume, the estimates obtained using the Chave et al. (2005) model were $\hat{\mu}$ = 107.52 $m^3 \cdot ha^{-1}$, with SE = 3.09 $m^3 \cdot ha^{-1}$. The Chave et al. model tended to overestimate volume for large volume plots but had a mean deviation of less than 0.03. The overall mean and SE were overestimated proportionally by 0.05 and 0.6, respectively, and the estimate of the mean was within two SEs of the estimate for the nonspecific model for this study. Therefore, the Chave et al. (2005) model would have been marginally acceptable as a substitute for the local models, although the precision estimate was considerable less.

Table 3. Effects of uncertainty by source on estimates of mean volume per unit area.

Source of uncertainty		Models					
		Species-specific and (or) species group		Forest type		Nonspecific	
Parameter	Residual	Mean ($m^3 \cdot ha^{-1}$)	SE ($m^3 \cdot ha^{-1}$)	Mean ($m^3 \cdot ha^{-1}$)	SE ($m^3 \cdot ha^{-1}$)	Mean ($m^3 \cdot ha^{-1}$)	SE ($m^3 \cdot ha^{-1}$)
No	No	101.15	2.64	102.63	2.67	102.02	2.67
No	Yes	101.15	2.64	102.63	2.67	102.01	2.68
Yes	No	101.36	2.69	102.68	2.71	102.00	2.70
Yes	Yes	101.36	2.69	102.68	2.71	102.01	2.70

Note: SE is standard error.

Table 4. Effects of calibration dataset sample size on estimates of mean volume per unit area.

	Models			
	Forest type		Nonspecific	
Sample size proportion	Mean ($m^3 \cdot ha^{-1}$)	SE ($m^3 \cdot ha^{-1}$)	Mean ($m^3 \cdot ha^{-1}$)	SE ($m^3 \cdot ha^{-1}$)
1.0000	102.68	2.71	102.01	2.70
0.5000	102.75	2.75	102.01	2.73
0.2500	104.32	2.92	102.03	2.79
0.1250	106.83	3.48	103.71	3.07
0.0625	115.07	5.45	104.82	3.60

Note: SE is standard error.

5. Conclusions

Four conclusions may be drawn from the study. First, the general method using Monte Carlo simulations was relatively easy to implement, accommodated uncertainty from multiple sources, and can readily be extended to include uncertainty from additional sources such as diameter and height measurement error. Second, the effects of model residual and parameter uncertainty on the uncertainty of large-area estimates of mean volume per unit area were negligible relative to the effects of sampling variability. However, the effects of reducing the sample sizes of the calibration datasets used for this study by factors greater than 4.0 on parameter uncertainty and indirectly on estimates of mean volume were potentially substantial. Third, the similarity of results for the models constructed for this study and the models currently in use by the IFFSC indicates that the adverse effects of model residual and parameter uncertainty on IFFSC model predictions may be ignored for large-area estimation. However, reduction of the effects of sampling variability via use of stratified and model-assisted estimators that incorporate auxiliary information may substantially increase the relative effects of residual and parameter uncertainty. Fourth, estimates of both the mean and the standard error differed very little for the species-specific/species-group, the forest-type, and the nonspecific models. This result suggests that the nonspecific model may be adequate for many applications.

References

Anderson-Sprecher, R. 1994. Model comparisons and R^2. Am. Stat. **48**(2): 113–117. doi:10.2307/2684259.

Avery, T.E., and Burkhart, H.E. 2002. Forest measurements. 5th edition. McGraw-Hill, New York.

Baskerville, G.L. 1972. Use of logarithmic regression in the estimation of plant biomass. Can. J. For. Res. **2**(1): 49–53. doi:10.1139/x72-009.

Basuki, T.M., van Laake, P.E., Skidmore, A.K., and Hussin, Y.A. 2009. Allometric models for estimating aboveground biomass in tropical lowland *Dipterocarp* forests. For. Ecol. Manage. **257**: 1684–1694. doi:10.1016/j.foreco.2009.01.027.

Berger, A., Gschwantner, T., McRoberts, R.E., and Schadauer, K. 2014. Effects of measurement errors on individual tree stem volume estimates for the Austrian National Forest Inventory. For. Sci. **60**: 14–24. doi:10.5849/forsci.12-164.

Breidenbach, J., Antón-Fernández, C., Petersson, H., Astrup, P., and McRoberts, R.E. 2014. Quantifying the contribution of biomass model errors to the uncertainty of biomass stock and change estimates in Norway. For. Sci. **60**: 25–33.

Brown, S., Gillespie, A.J.R., and Lugo, A.E. 1989. Biomass estimation methods for tropical forests with application to forest inventory data. For. Sci. **35**(4): 881–902.

Chave, J., Andalo, C., Brown, S., Cairns, M.A., Chambers, J.Q., Eamus, D., Fölster, H., Fromard, F., Higuchi, N., Kira, T., Lescure, J., Nelson, B.W., Ogawa, H., Puig, H., Riéra, B., and Yamakura, T. 2005. Tree allometry and improved estimation of carbon stocks and balance in tropical forests. Oecologia, **145**: 87–99. doi:10.1007/s00442-005-0100-x.

Chave, J., Muller-Landau, H.C., Baker, T.R., Easdale, T.A., ter Steege, H., and Webb, C.O. 2006. Regional and phylogenetic variation of wood density across 2456 neotropical tree species. Ecol. Appl. **16**(6): 2356–2367. doi:10.1890/1051-0761(2006)016[2356:RAPVOW]2.0.CO;2.

Cochran, W.G. 1977. Sampling techniques. 3rd edition. Wiley, New York. p. 134.

Cunia, T. 1965. Some theory on reliability of volume estimates in a forest inventory sample. For. Sci. **11**(1): 115–128.

Cunia, T. 1987. On the error of continuous forest inventory estimates. Can. J. For. Res. **17**(5): 436–441. doi:10.1139/x87-075.

Gertner, G.Z. 1990. The sensitivity of measurement error in stand volume estimation. Can. J. For. Res. **20**(6): 800–804. doi:10.1139/x90-105.

Gertner, G.Z., and Dzialowy, P.J. 1984. Effects of measurement error on an individual tree-based growth projection system. Can. J. For. Res. **14**(3): 311–316. doi:10.1139/x84-057.

Holdridge, L.R. 1967. Life zone ecology. Tropical Science Center, San Jose, California.

Jenkins, J.C., Chojnacky, D.C., Heath, L.S., and Birdsey, R.A. 2003. National-scale biomass estimators for United States tree species. For. Sci. **49**(1): 12–35.

Klein, R.M. 1978. Mapa fitogeográfico de Santa Catarina. Herbário Barbosa Rodrigues, Itajaí, Brazil.

McRoberts, R.E., and Westfall, J.A. 2014. The effects of uncertainty in model predictions of individual tree volume on large area volume estimates. For. Sci. **60**: 34–43. doi:10.5849/forsci.12-141.

McRoberts, R.E., Hahn, J.T., Hefty, G.J., and Van Cleve, J.R. 1994. Variation in forest inventory field measurements. Can. J. For. Res. **24**(9): 1766–1770. doi:10.1139/x94-228.

Návar, J. 2009. Allometric equations for tree species and carbon stocks for forests of northwestern Mexico. For. Ecol. Manage. **257**: 427–434. doi:10.1016/j.foreco.2008.09.028.

Rubin, D.B. 1987. Multiple imputation in nonresponse surveys. Wiley, New York.

Ståhl, G., Heikkinen, J., Petersson, H., Repola, J., and Holm, S. 2014. Adapting uncertainty assessments from sample based forest inventories to include the effects of model errors. For. Sci. **60**: 3–13.

Vibrans, A.C., Sevegnani, L., Lingner, D.V., Gasper, A.L., and Sabbagh, S. 2010. Inventário florístico florestal de Santa Catarina (IFFSC): aspectos metodológicos e operacionais. Pesquisa Florestal Brasieira, **30**: 291–302.

Vibrans, A.C., McRoberts, R.E., Moser, P., and Nicoletti, A.L. 2013*a*. Using satellite image based maps and ground inventory data to estimate the area of the remaining Atlantic forest in the Brazilian state of Santa Catarina. Remote Sens. Environ. **130**: 87–95. doi:10.1016/j.rse.2012.10.023.

Vibrans, A.C., Moser, P., Lingner, D.V., Maçaneiro, J.P., and Silveira e Silva, L. 2013*b*. Amostragem dos remanescentes da Floresta Estacional Decidual em Santa Catarina. *In* Inventário Florístico Florestal de Santa Catarina. Vol. II. *Edited by* A.C. Vibrans, L. Sevegnani, A.L. Gasper, and D.V. Lingner. Floresta Estacional Decidual, Edifurb, Blumenau.

Vibrans, A.C., Moser, P., Lingner, D.V., Maçaneiro, J.P., and Silveira e Silva, L. 2013*c*. Amostragem dos remanescentes da Floresta Ombrófila Mista em Santa Catarina. *In* Inventário Florístico Florestal de Santa Catarina. Vol. III. *Edited by* A.C. Vibrans, L. Sevegnani, A.L. Gasper, and D.V. Lingner. Floresta Ombrófila Mista, Edifurb, Blumenau.

Xiang, W., Liu, S., Deng, X., Shen, A., Lei, S., Tian, D., Zhao, M., and Peng, C. 2011. General allometric equations and biomass allocation of *Pinus massoniana* trees on a regional scale in southern China. Ecol. Res. **26**: 697–711. doi:10.1007/s11284-011-0829-0.

Zianis, D., Muukkonen, P., Mäkipää, R., and Mencuccini, M. 2005. Biomass and stem volume equations for tree species in Europe. Silva Fennica Monographs 4. Tammer-Paino Oy, Tampere, Finland.