UNCERTAINTY ASSESSMENT IN VOLUME AND BIOMASS ESTIMATIONS IN FOREST STANDS

Franciele Alba da Silva*, Sylvio Péllico Netto†, Alexandre Behling‡, Rubén Manso†, Kauana Engel†

*Universidade Federal do Paraná, Programa de Pós-Graduação em Engenharia Florestal, Curitiba, Paraná, Brasil – francielealba47@gmail.com
†Universidade Federal do Paraná, Programa de Pós-Graduação em Engenharia Florestal, Curitiba, Paraná, Brasil – sylviopelliconetto@gmail.com
‡University of Huddersfield, Forest Research, Northgate, Huddersfield, UK – ruben.manso@forestresearch.gov.uk

Received for publication: 29/11/2022 – Accepted for publication: 06/03/2023

Resumo
Avaliação de incertezas nas estimativas de volume e biomassa em povoamentos florestais. No Brasil, a incerteza dos inventários florestais é calculada de forma independente e a incerteza relacionada ao modelo é sistematicamente ignorada. Embora os métodos para estimar e incerteza da amostragem e do modelo de forma conjunta sejam desconhecidos no país, os mesmos, são indispensáveis para melhorar os resultados dos inventários uma vez que estimativas de alta qualidade são importantes, devido a suas grandes áreas de florestas, pelo fornecimento de produtos madeireiros e não-madeireiros, e pela referência mundial como fornecedor de serviços ambientais. Diante do exposto, este trabalho apresenta a seguinte hipótese de estudo: “Considerando a incerteza associada aos componentes do inventário: i - amostragem e ii - modelo de regressão combinado na variação total do inventário florestal resulta em estimativas mais confiáveis de volume e biomassa”. Assim, o objetivo deste estudo é avaliar as incertezas associadas à amostragem e as estimativas de volume e biomassa devidas ao modelo de regressão linear nos povoamentos de Acacia mearnsii no Brasil. Para avaliar conjuntamente estas duas fontes de incerteza, foi utilizado o estimador de variância híbrida, com uma abordagem analítica. Os resultados mostraram que, se a incerteza do modelo não for considerada, a incerteza total é subestimada em 6,51%. Nas estimativas de biomassa, a incerteza total é subestimada em 18,74%. Ignorar a incerteza nas estimativas totais pode levar à decisões equivocadas para o manejo florestal, com implicações econômicas, particularmente nas estimativas de biomassa, em que a variação associada é ainda maior do que em volume devido à natureza desta variável.

Palavras-chave: Inventário florestal, Propagação de erros, Quantificação da madeira.

Abstract
In Brazil, forest inventory variation is calculated independently, and model-related uncertainty is systematically ignored. Although methods of estimation evaluation and sampling uncertainty together are unknown in the country, they are indispensable for improving the results of forest inventories since obtaining high quality estimates is extremely important for the country, due to its large forest areas and distinguished for being a world leader in the supply of timber and non-timber forest products, and a reference as a provider of environmental services. In view of the above, this paper presents the following study hypothesis: “Considering the uncertainty associated with inventory components: i - sampling and ii - regression model combined in the total variance of the forest inventory results in more accurate volume and biomass estimates” Thus, the aim of this study was to evaluate the uncertainties associated with sampling and the linear regression model in volume and biomass estimates in Acacia mearnsii stands in Brazil. To jointly evaluate these two sources of uncertainty, the hybrid variance estimator was used, with an analytical approach. The results showed that if model uncertainty is not considered, the total uncertainty is underestimated by 6.51%. In biomass estimates, the total uncertainty is underestimated by 18.74%. Ignoring uncertainty in total estimates can lead to uninformed decisions in forest management, with economic implications, particularly in biomass estimates, where the associated variation is even greater than in volume due to the nature of this variable.

Keywords: Forest inventory, Error propagation, Wood quantification.

INTRODUCTION
The quantification of forest stocks is fundamental to assess the available forest resources, which is the basis for forest planning at national and regional level. Forest inventory is the tool that makes that quantification possible (Péllico Netto and Brea, 1997). The same typically provide two different types of metrics: those that are directly measured (e.g. basal area) and those that need to be modelled from auxiliary variables because they are difficult to obtain (e.g. volume and biomass). The latter are highly relevant for management and decision-making
purposes. These metrics can then be upscaled to the regional or national level through statistical estimators to obtain estimates for the total population.

When it comes to the modelled metrics, there are two sources of uncertainty linked to the estimation of the total population. First, the uncertainty due to the fact only a sample of the population is known (sampling), and second, the uncertainty associated to the model errors. However, forest inventory statistics, generally consider these two main uncertainty components independently, which results in a poor uncertainty assessment – usually as a noticeable subestimation of the total uncertainty. This may have a massive impact on decision-making and, consequently, on forest management, as may end up with a distorted perception of precision that eventually leads to wrong decisions. Ignoring uncertainty compromises the data collection planning, potentially incurring higher costs than needed as well. Therefore, uncertainty assessment is an essential component to improve the quality and reveal the accuracy of estimates reported in forest inventory (KANGAS 2018, STAHL et al., 2016, IPCC, 2022).

Sampling and model uncertainties are unavoidable but can be estimated using statistical estimators, although combining sampling, and model uncertainty to calculate the total uncertainty is somewhat challenging and usually results in complex mathematical developments (STÅHL et al., 2016; KANGAS et al., 2018). Variance estimators that consider model and sampling uncertainty jointly are called hybrid estimators (CORONA et al. 2012) and significant research effort has been dedicated in recent years to the development of these methodologies to evaluate the uncertainty of the main sources of forest inventories (CUNIA 1987; GERTNER 1990; CHAVE et al., 2004; FORTIN et al., 2016; MCROBERTS and WESTFALL 2015; FORTIN et al., 2018). Efficiency gains in using methodologies that consider uncertainties jointly have already been reported for volume estimates by McRoberts and Westfall (2014) and Berger et al. (2014) for biomass estimates by Cunia (1987), Chave et al. (2004) and Molto et al. (2013).

Even with these established and studied methodologies, especially in Brazil, forest inventory precision assessments and model-related uncertainty, is systematically ignored. However, obtaining high-quality estimates is extremely important for Brazil, because it is a country with large areas of native and planted forests. Besides, the country plays an important role in the supply of timber and non-timber forest products, being a reference as a supplier of environmental services. Thus, the application of these methods to Brazilian forests would make a relevant contribution in this domain since in Brazil this theme is still unknown.

Therefore, this study aimed to i) assess the uncertainty associated with sampling and model jointly, in volume and biomass estimates and ii) present the main hybrid variance estimators to quantify uncertainty in linear models. The aims were developed under the hypothesis that quantifying the uncertainty associated with the regression model and sampling jointly, results in a more reliable assessment of volume and biomass estimates.

To achieve our aims, we used the case study of Acacia mearnsii stands in South Brazil and we fitted a linear model for the variables volume and biomass aboveground. The approaches presented in this research were the analytical method described by Cunia (1987), following the error propagation law, and Monte Carlo simulations.

MATERIAL AND METHODS

The data comes from Acacia mearnsii stands, located in Rio Grande do Sul, south of Brazil. Stands were divided in three regions, according to the stand's location in different sites: Cristal, Encruzilhada do Sul and Piratini, with the ages 10.08 years, 10.75 9.83 years, respectively.

Data collection

First phase data: Sampling
In each stand four circular plots were randomly installed, with a diameter of 22.56 m (400 m²), and another 82 plots were randomly distributed among the stands, totaling 94 temporary plots. In these plots 6,927 trees were measured in the diameter at breast height (cm) and total height (m), which were used to quantify the stem volume and biomass stocks of the components above ground, by regression models. Diameter at breast height was measured using a dendrometer tape and total height with a hypsometer (Haglöf).

Second phase data: Model
A 10 m diameter subplot (78.54 m²) was established in the center of 48 subplots because it was not possible to measure all variables of all trees in the larger plot, due to the high cost. Biomass from branches and crown, and total volume were measured for 180 trees.

All trees in the subplots were felled and evaluated for their characteristics: diameter at breast height was measured using a dendrometer tape and total crown height, and length with a tape measure. The stem volume was obtained using the Huber method and its scale was made using the Hohenadl method. Measurements were taken with a dendrometric tape along the stem at positions of 5%, 15%, 25%, 35%, 45%, 55%, 65%, 75%, 85%, and 95% of the total height.
Biomass measurements were taken for the stem (stem wood + bark) and crown (living branches and dead leaves, flowers and fruits). For each tree, these components were separated and weighed to determine green biomass with a digital scale (Portable Electronic Scale) with an accuracy of 5 g.

To determine the dry biomass of the stem and crown, samples were collected and immediately weighed using a digital balance (Hoyle) with an accuracy of 1 g. From the stem, 5 disks (2 cm thick) were collected at the positions 0%, 25%, 50%, 75%, and 95% along the total height. The samples were dried in an oven with air circulation at 100°C and after a constant weight the material was weighed on a digital scale with an accuracy of 1 g.

Uncertainty assessment

Spurr model in linear fashion was fitted for volume and biomass, as a function of the diameter and total height variables (Equation 1). The coefficients $\hat{\mathbf{\beta}}$ were estimated by weighted ordinary least squares.

$$\hat{y} = \beta_1 + \beta_2 x_1$$  

Where: $\hat{y}$ is the interest variables: volume and biomass, $x_1 = d^2 h$ which $d$ is diameter at breast height (cm) and $h$ is height (m).

The basic regression assumptions were met. The equations quality was evaluated by coefficient of determination ($R^2$), standard deviation error in percentage (Syx%), and Akaike information criterion (AIC).

The procedure for combining the error of the two components has been presented by Cunia (1987) (Equation 2). The estimates of the volume stock biomass in the population level combine the statistics of the second phase (coefficients of equation of biomass) with the statistics of the first phase (number of trees and their size). To calculate the error of these estimates ($S_{vw}$), consider the following procedures which, for convenience, refer to an estimator of volume and biomass aboveground per hectare $(w)$.

$$S_{vw} = [b]'[s_{zz}][b] + [z]'[s_{bb}][z]$$  

Where: $[b]$ is the vector of the model coefficients, $[b]'$ is the transposed matrix, $[s_{bb}]$ is the covariance matrix due to the model coefficients, $[z]$ is the vector of the explanatory variables, $[z]'$ is the transposed matrix, and $[s_{zz}]$ is the variance-covariance matrix due to the sampling.

The estimation of the vector $[b]'$ is denoted by $[b]^* = [b_1 \ b_2 \ \ldots \ b_m]$, $m$: number of model variables, and the estimate of the covariance matrix $[\sigma_{bb}]$ is defined as:

$$[s_{bb}] = \begin{bmatrix}
S_{b_1,b_1} & S_{b_1,b_2} & \cdots & S_{b_1,b_m} \\
S_{b_2,b_1} & S_{b_2,b_2} & \cdots & S_{b_2,b_m} \\
\vdots & \vdots & \ddots & \vdots \\
S_{b_m,b_1} & S_{b_m,b_2} & \cdots & S_{b_m,b_m}
\end{bmatrix}$$

The statistics $z_1, z_2, \ldots, z_m$ are calculated from data of sampling related to the explanatory variables. In our specific case with Spurr in linear fashion $[z]$ is described as $[z] = \sum_{i=1}^{N} \sum_{d=1}^{d_i^2} h_i$.

The estimator of the covariance matrix $[S_{zz}]$ is defined. The diagonal matrix is the variance of the stand attributes, corresponding of each variable of the first phase in the plots, and the diagonals are the covariances between them, per hectare.

$$[s_{zz}] = \begin{bmatrix}
S_{z_1,z_1} & S_{z_1,z_2} & \cdots & S_{z_1,z_m} \\
S_{z_2,z_1} & S_{z_2,z_2} & \cdots & S_{z_2,z_m} \\
\vdots & \vdots & \ddots & \vdots \\
S_{z_m,z_1} & S_{z_m,z_2} & \cdots & S_{z_m,z_m}
\end{bmatrix}$$

Where for the values of the diagonal we calculate the variance of the stand attributes $\frac{\sum_{i=1}^{n} (z_i - \bar{z})^2}{n-1}$, and the off-diagonal values are the covariance between the stand attributes $\frac{\sum_{i=1}^{n} (z_i - \bar{z})(z_i' - \bar{z}')}{n-1}$.

The calculation is done per plot and extrapolated to the hectare. After the results per plot, the mean of these values is made and these values are introduced into equation 2.
Monte Carlo simulation

Besides the analytical method, we used the Monte Carlo (MC) simulation procedure to access the uncertainty in volume and biomass estimates. The technique consists in designing a large number of repetitions of some random variables in order to reproduce the variability of a process (Fortin et al. 2018).

To assess the uncertainty with MC simulation, sequential steps are followed:
(i) For a given realization, random deviations are generated to account for the model-related uncertainty.
(ii) The random deviation value is added to the model coefficients and
(iii) Values of volume and above ground biomass in the plots are obtained $\hat{Y}_b$ with coefficients generated in (ii).
(iv) We ran $B$ simulations ($B = 10$).
(v) To distinguish the contribution of each source of uncertainty to the total uncertainty we use the hybrid estimator, presented in the equation

$$\hat{V}(\hat{Y}_b) = \frac{1}{B} \sum_{b=1}^{B} (\hat{Y}_b - \hat{Y}_{b5})^2 + \frac{1}{B} \sum_{b=1}^{B} \hat{V}_d(\hat{Y}_b)$$

Where: $\hat{V}(\hat{Y}_b)$ is the total variance, $\hat{Y}_b$ is an estimate of the mean volume or biomass for each realization, $\hat{Y}_{b5}$ is the mean for all realizations, $\hat{V}_d(\hat{Y}_b)$ is the contribution of the sampling (mean variance) and $B$ is the number of realization.

RESULTS

The equation estimated using the regression model used in this study are presented in Table 1. The equations showed satisfactory values of good fit results for volume and biomass. The residual plots showed no bias in the estimates for either volume or biomass (Figure 1).

Table 1. Statistics of volume and biomass equation fitting of Acacia mearnsii stands.
Tabela 1 - Estatísticas de ajuste da equação de volume e biomassa para plantios de Acacia mearnsii.

<table>
<thead>
<tr>
<th>Variável</th>
<th>Equation</th>
<th>$R^2$</th>
<th>AIC</th>
<th>Syx(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume (m$^3$)</td>
<td>$\hat{v} = 0.004278 + 0.00003724 d^2 h$</td>
<td>0.98</td>
<td>-11524</td>
<td>7.8</td>
</tr>
<tr>
<td>Biomassa (kg)</td>
<td>$\hat{b} = 0.7696 + 0.02729 d^2 h$</td>
<td>0.96</td>
<td>1267.1</td>
<td>14.04</td>
</tr>
</tbody>
</table>

R$^2$adj: coefficient of determination; AIC: Akaike's information coefficient; Syx(%): standard error of the estimate;

For the experimental condition, the mean wood volume in the forest was 320.366 m$^3$.ha$^{-1}$ with a total variance of 45.8 m$^3$.ha$^{-1}$. Of this total variance, the sampling contribution was 93.49%, while the variance due to the model contributed 6.51% to the total variance (Table 2). The mean for biomass was 230,736.5 kg.ha$^{-1}$. Sampling contributed 81.26% of the total variance and the biomass model contribution to the total variance was 18.74%. This means that if one does not consider the error propagation due to the volume equation, the total uncertainty of the inventory would be underestimated by 6.51%. For the biomass, the uncertainty would be underestimated by...
18.74% (Figure 2). The results were shown to be similar for both approaches, MC simulations and analytical methods.

Table 2. Uncertainty of volume and biomass estimates associated with inventory components: i - sampling and ii- regression model.

<table>
<thead>
<tr>
<th>Mean</th>
<th>Sampling</th>
<th>Model</th>
<th>Under (%)</th>
<th>Total</th>
<th>Ic 95 ±</th>
</tr>
</thead>
<tbody>
<tr>
<td>ψ</td>
<td>An</td>
<td>MC</td>
<td>An</td>
<td>MC</td>
<td>An</td>
</tr>
<tr>
<td>(m³·ha⁻¹)</td>
<td>42.81</td>
<td>43.28</td>
<td>2.98</td>
<td>3.05</td>
<td>6.52</td>
</tr>
<tr>
<td>w</td>
<td>22.720.945</td>
<td>22.720.945.0</td>
<td>4.940.54</td>
<td>51.24.850.8</td>
<td>17.97</td>
</tr>
<tr>
<td>(kg·ha⁻¹)</td>
<td>230.736.</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

An is the analytical results, MC is the Monte Carlo simulation results, Sampling is the uncertainty due to the sampling, model is the uncertainty due to the model, under (%) is the uncertainty underestimated in case that consider just the uncertainty due to sampling, total is the total uncertainty and IC is the confidente intervals.

Figure 2. Proportion of the total uncertainty in volume and biomass estimations of Acacia mearnsii stands.

The smaller confidence intervals of the volume and biomass estimates occur due to an underestimate of the model uncertainty, as shown by the red confidence interval in figure 1. We can have narrow intervals that induce us to think that the estimate is good, while in fact, the variance estimator has a large variance. This can be clearly seen in the length of the value of the confidence interval for the mean.

Figure 3. Confidence interval for the mean volume and biomass estimates considering the uncertainty associated to the forest inventory components: i - sampling and ii- regression model. Icc is the CI considering sampling+model uncertainty; Ica is the CI considering just sampling uncertainty.

Figure 3. Intervalo de confiança para a média das estimativas do volume e biomassa considerando a incerteza associada aos componentes do inventário florestal; i- amostragem e ii- modelo de regressão. Icc é o IC considerando a incerteza da amostragem+modelo; Ica considera apenas a incerteza devido a amostragem.
DISCUSSION

We assessed model and sampling-related uncertainty in volume and biomass estimates. According to our results, forest inventory estimates should take into account these sources of uncertainty because, when important sources of uncertainty are overlooked, such as uncertainties related to the model parameters, the total uncertainty is underestimated.

Since the total variance is used to construct confidence intervals for the means, per unit area, this underestimated variance directly affects the confidence interval results (SAARELA et al. 2017). Specifically, the confidence interval is a measure of precision of estimates, associated with a certain confidence level, and determines the lower and upper limits within which one expects to find the parametric value of the estimated variable (PÉLICO NETTO and BRENA, 1997). Thus, when uncertainty is not evaluated in forest inventory estimates, the total variance is not correctly represented and one may have narrow confidence intervals, that is, with underestimated values, which induce to believe that the estimate is adequate, while in fact estimates around the mean have a large variance (SAARELA et al. 2017). Consequently, forest managers may have a misleading perception of precision, with potential impacts on decision-making and forest management (KANGAS et al. 2018).

In order to provide reliable estimates of uncertainty, variance estimators should not only be unbiased, but also accurate (FORTIN et al. 2016). As a result of our research, we tested two hybrid variance estimators, and both produced the same results for volume and biomass. The hybrid variance estimator presented in analytical form by Cunia (1987) has lower computational demand, compared to the MC methodology. Some papers highlight the idea of replacing some time-consuming and computational efforts of MC simulation by choosing the appropriate analytical approach (GU et al. 2021). In our specific situation, for linear models, the analytical method can be uniquely used with the concern of using a concise and accurate hybrid variance estimator. According to Fortin et al. (2018) analytical approach is a good way to assess the uncertainty in linear or non-linear models that apply in a non-complex system, e.g., systems with more than one equation, or at a lower hierarchical level.

The larger uncertainty from the biomass equation are due to the fact that biomass varies greatly among trees and biomass components. This leads to the conclusion that no biomass model can account for the tree-level biomass variability as precisely as a volume model does with volume variability (VORSTER et al. 2020). It seems that ignoring the uncertainty due to the model in population-level biomass estimates can lead to a more severe underestimation of the total uncertainty than in the case of volume.

The results presented in this paper are in accordance with other surveys regarding the amount of uncertainty due to the volume and biomass models. Cunia (1987) found that percentage to be as high as 34.86% of the total variance in mixed forests of the New York State Forest area. In another study, Chave et al. (2004) found an error greater than 20% due to the model in aboveground biomass estimates in tropical rainforests of Central Panama. Our results are with the uncertainty contribution due to the model of 18%, lower than the cited research. We can attribute this to the fact that our data are from experimental forest plantations, while the other authors worked exclusively with native forests. In another study, Magalhaes and Seifert (2015) also used forest plantations for estimates of above and below-ground biomass and carbon stocks of Mecrusse forests in Mozambique, and the variance contributed was approximately 10% of the uncertainty in the estimate of total biomass.

For the volume, Gertner (1990) concluded that the variance due to the model parameters amounted to 6.76% of the total variance for Pinus resinosa in Minnesota, EUA. Kangas (1996) compared Taylor series expansion, Monte Carlo simulations and recursive modelling methods to assess uncertainty in volume predictions and the percentage of the error due to the model was 3.53% to data from the Finnish National Forest Inventory (NFI) on Scots pine (Pinus sylvestris L.) sample units. Berger et al. (2014) used Monte Carlo simulations and an analytical method in the context of the Austrian NFI. The authors found that the error due to the model was 5.6% of the total error for all species. McRoberts and Westfall (2015) observed that the contribution of the variance due to the model was 10% of the total variance when using DBH as the sole predictor in their volume model. Based on these values, the uncertainty due to the volume equation is generally around 10%, which is in line with our work, that resulted in at 6.5% to the total uncertainty contribution due to the model.

CONCLUSION

- The uncertainty associated with the model is smaller than the sampling uncertainty. However, we have seen that ignoring this source of error can lead to uninformed decisions in forest management, particularly in biomass estimates, where the associated variance is even greater than in volume due to the nature of that variable.
- Assessing the uncertainty is, consequently, essential to implement management strategies and policy interventions and should be included as part of the estimates of any forest inventory.
• Evaluating and understanding uncertainty in population mean estimates is the fastest and cheapest way to improve the reliability in decision making.

REFERENCES


