Form factors vs. regression models in volume estimation of *Pinus taeda* L. stem

Fatores de forma vs. modelos de regressão na estimação volumétrica do fuste de *Pinus taeda* L.

Carlos Roberto SANQUETTA; Marilia do Carmo DOLCI; Ana Paula Dalla CORTE; Mateus Niroh INOUE SANQUETTA; Allan Libanio PELISSARI

Abstract

Volume is an important information to know the available potential of a forest stand, although direct volume determination is expensive and often impractical. Thus, precise and accurate indirect methods are fundamental. Three volume estimation methods of *Pinus taeda* stem were compared: mean form factor, form factor by diameter class, and volume models fitted by regression. A total of 146 trees were used for this study, in which 96 were used for fitting and estimation, and 50 for validation, collected in the middle of the Paraná State, Brazil. Fitting statistics (\(R^2\)adj and Syx%) were used to evaluate regression models. Bias, precision and accuracy statistics, t-test and residue graphical analysis were used to judge estimation methods. Estimation by mean form factor (diameter class) and the Spurr model were equivalent, being superior to the others examined. It was concluded that the use of mean form factor by diameter class has satisfactory statistical performance, robustness and simplicity in volume estimation.

Additional keywords: accuracy; precision; validation; volumetry.

Introduction

Quantification of the wood volume existing in a forest stand is essential for its management and wood commercialization. However, direct determination of volume is expensive, since it requires rigorous sampling by destructive techniques and/or procedures that require time and resources (Sanquetta et al., 2014). Therefore, precise and accurate indirect methods are fundamental. Regression models based on diameter at breast height (dbh) and tree height have been extensively tested and applied.

The artificial form factor (based on dbh) is also widely used to estimate stem volume, where, generally, a mean value is used and multiplied by the cross-sectional area and tree height. However, its indiscriminate use is constantly subject to criticism, and its use is only recommended in situations where there are no volume equations and expeditious calculations (Miguel et al., 2010), restricted to local conditions, are going to be made. Any extrapolation beyond these limits is considered risky. Despite this, form factor is still widely used today in practice due to its simplicity.

The *Pinus* genus is the second most planted in Brazil, with 1.6 million hectares, and the Paraná State has the largest *Pinus* stands area (IBA, 2016).
Currently, the *Pinus taeda* L. species is the most important commercially, which is used for several purposes, such as cellulose and long fiber paper, packaging, lumber and panels. *Pinus* stands out in Brazil’s southern region due to its high volume increment, reaching mean annual increment (MAI) in volume of 40 m³ ha⁻¹ year⁻¹ (Ferreira, 2005; Kohler, 2013).

Quantifying the volume of *Pinus taeda* managed stands in a precise and accurate manner is always one of the most relevant activities for companies that cultivate this species, as yield gains demand their estimation with methods that are appropriate, robust and, as far as possible, simple to apply in practice. The aim of this study was to evaluate the performance of form factor by diameter class in volume estimations compared with regression volume models in a *Pinus taeda* stand located in the middle of Paraná State, Brazil.

Table 1 – Frequency of trees sampled by diameter class for estimation the total stem volume with bark of *Pinus taeda* L. stem in the middle of Paraná State, Brazil.

<table>
<thead>
<tr>
<th>Diameter class (cm)</th>
<th>For fitting</th>
<th>For validation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 9.0</td>
<td>7</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>9.0 – 12.9</td>
<td>11</td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>13.0 – 16.9</td>
<td>10</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>17.0 – 20.9</td>
<td>14</td>
<td>5</td>
<td>19</td>
</tr>
<tr>
<td>21.0 – 24.9</td>
<td>12</td>
<td>6</td>
<td>18</td>
</tr>
<tr>
<td>25.0 – 28.9</td>
<td>16</td>
<td>5</td>
<td>21</td>
</tr>
<tr>
<td>29.0 – 32.9</td>
<td>11</td>
<td>6</td>
<td>17</td>
</tr>
<tr>
<td>33.0 – 36.9</td>
<td>7</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>37.0 – 40.9</td>
<td>7</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>&gt; 40.9</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>96</strong></td>
<td><strong>50</strong></td>
<td><strong>146</strong></td>
</tr>
</tbody>
</table>

In order to evaluate individual total volume estimation quality, three different methods were tested, namely:

**Mean artificial form factor application**

Mean artificial form factor for all 96 trees (\(\bar{f}\)) was calculated based on the diameter at 1.30 m (\(dbh\)), as follows:

\[
\bar{f} = \frac{\sum_{i=1}^{n} f_i}{n}, \text{ where } n = 96
\]

Where:

\[
f_i = \frac{v_i}{v_i \text{ cylinder}} = \text{ Each individual's form factor;}
\]

\[
v_i = \text{ Real volume, obtained through strict measurement (m}^3); \text{ and}
\]

\[
v_{i \text{ cylinder}} = \text{ Cylinder volume, calculated from the}
\]

\[
dbh \text{ (m}^3) = v_{i \text{ cilindro}} = \frac{\pi \/dbh_i^2}{40.000} h_i.
\]

Therefore, the volume of each tree \(i\) was estimated through:

\[
\hat{v}_i = \frac{\pi \/dbh_i^2}{40.000} h_i \bar{f}.
\]

Where: \(h_i = \text{ Tree total height (m)}.

**Artificial form factor by diameter class**

A mean artificial form factor was calculated for each diameter class, as follows:

\[
\bar{f}_j = \frac{\sum_{i=1}^{n_j} f_i}{n_j}
\]

Where:

\(f_i = \text{ Each individual's form factor, as previously defined; and}
\)

\(n_j = \text{ Number of trees of the diameter class } j.
\)

Therefore, the volume of each tree \(i\) was estimated through:

\[
\hat{v}_i = \frac{\pi \/dbh_i^2}{40.000} h_i \bar{f}_j.
\]

In this case, the form factor to be applied to obtain the volume was the mean value of the class to which the tree belongs.

**Material and methods**

Data for this study were collected in pure and unthinned commercial stands of *Pinus taeda* with different ages, ranging from 5 to 15 years old. 146 trees were sampled by Smalian’s method (equation 1), based on the stand diameter distribution. The total real volumes of these trees (including bark) were obtained after sampling. Of all trees sampled, 96 trees were separated for fitting equations and artificial form factors, whereas 50 trees were used for validation (Table 1).

\[
v_i = \frac{(g_1 + g_2)}{2} l_{12} + \frac{(g_2 + g_3)}{2} l_{23} + \ldots + \frac{(g_{n-2} + g_{n-1})}{2} l_{n-2n-1} + \frac{g_n}{3} l_n
\]

Where:

\(v_i = \text{ Real volume of each tree (m}^3);\)

\(g_1, g_2, \ldots, g_n = \text{ Cross sections, taken along the stem (m}^2); \text{ and}
\)

\(l_2, l_3, \ldots, l_n = \text{ Length of each log measured in the tree (m)}.

In this case, the form factor to be applied to obtain the volume was the mean value of the class to which the tree belongs.
Volume regression models

Seven mathematical models were tested for volume estimation in relationship to dbh and tree height (Table 2), in which fittings were made by ordinary least squares method, whose logarithmic discrepancy of logarithmic models (11 to 14) was corrected by Meyer’s Correction.

Table 2 – Regression models fitted for *Pinus taeda* L. total stem volume estimation in the middle of Paraná State, Brazil.

<table>
<thead>
<tr>
<th>Denomination</th>
<th>Model</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>( \hat{v}_i = \beta_0 + \beta_1 (dbh) )</td>
<td>(8)</td>
</tr>
<tr>
<td>Kopezky-Gehardt</td>
<td>( \hat{v}_i = \beta_0 + \beta_1 (dbh^2) )</td>
<td>(9)</td>
</tr>
<tr>
<td>Spurr</td>
<td>( \hat{v}_i = \beta_0 + \beta_1 (dbh^2, h) )</td>
<td>(10)</td>
</tr>
<tr>
<td>Husch</td>
<td>( \ln(\hat{v}_i) = \beta_0 + \beta_1 ln(dbh) )</td>
<td>(11)</td>
</tr>
<tr>
<td>Husch modified</td>
<td>( \ln(\hat{v}_i) = \beta_0 + \beta_1 ln(dbh^2) )</td>
<td>(12)</td>
</tr>
<tr>
<td>Schumacher-Hall</td>
<td>( \ln(\hat{v}_i) = \beta_0 + \beta_1 ln(dbh) + \beta_2 ln(h) )</td>
<td>(13)</td>
</tr>
<tr>
<td>Spurr log</td>
<td>( \ln(\hat{v}_i) = \beta_0 + \beta_1 ln(dbh^2) )</td>
<td>(14)</td>
</tr>
</tbody>
</table>

Where: \( \ln \) = natural logarithm; \( \hat{v}_i \) = estimated tree volume (m³); \( h \) = tree total height (m); \( dbh \) = diameter at breast height (cm); and \( \beta_0, \beta_1 \) and \( \beta_2 \) = regression coefficients.

Fitting quality evaluation was based on the following criteria: adjusted coefficient of determination \( R^2aj \); standard error of estimate in percentage \( Syx% \); and residual graphical analysis (Sanquetta et al., 2014). Bias, precision and accuracy statistics were also calculated for comparison between models (Pretzsch, 2009).

\[
\text{Bias: } \bar{\text{e}}\% = \frac{\bar{\text{e}} \times 100}{V}
\]

(15)

where: \( \bar{\text{e}} = \frac{\sum_{i=1}^{n}(\hat{v}_i - v_i)}{n} \) (16)

\[
\text{Precision: } s_e\% = \frac{s_e \times 100}{V}
\]

(17)

where: \( s_e = \sqrt{\frac{\sum_{i=1}^{n}(\hat{v}_i - \bar{v} - v_i)^2}{n-1}} \) (18)

\[
\text{Accuracy: } m_v\% = \frac{m_v \times 100}{V}
\]

(19)

where: \( m_v = \sqrt{\frac{\sum_{i=1}^{n}(\hat{v}_i - v_i)^2}{n-1}} \) (20)

Where:
- \( \hat{v}_i \) = estimated value (m³);
- \( v_i \) = real value (m³);
- \( \bar{v} \) = mean real value (m³); and
- \( n \) = number of observations.

Data validation was conducted to provide an independent database, in order to assess estimation quality. For this purpose, 50 trees were used. In addition, to statistically evaluate the estimate results, paired t-test of observed and estimated values was used with 95% probability level for both fitting and validation.

Results and discussions

The mean form factor, calculated for the 96 trees sampled and used for volume estimation, was 0.47, with a coefficient of variation equal to 17.87%. Highest mean and also highest artificial form factor variation were verified in trees with shortest diameter. There was a slight trend of form factor reduction with diameter classes increase (Table 3).

Mainardi et al. (1996) evaluated *Pinus taeda* form factor in Rio Grande do Sul State, stating variation according to age in relationship to total height and dbh increments. Trees that were four years old had a form factor of 0.79, whereas the form factor dropped to 0.5 at 16 years old, demonstrating tapering trend and, consequently, form factor value reduction. Drescher et al. (2001) found the same behavior for 20 years-old *Pinus elliottii* Engelm.

Younger trees usually have highly tapered trunks, so that the tree base diameter is much larger than dbh. Thus, tree volume may be very close or even higher than reference cylinder volume (Kohler, 2013).

The volume regression models tested showed high coefficient of determination values, above 0.90, (except for the first model), and variable model estimation standard errors from 11 to 36%. Some models showed poor fitting, such as numbers 1, 2, 4 and 5 (Table 4), which, despite high \( R^2aj \) values, also resulted in high \( Syx% \) values, indicating poor performance. Model 3 had the highest coefficient of determination and the lowest standard error of estimate.

Mean form factor use had a positive bias of around 4% in the estimates, both with fitting data and validation data. Precision and accuracy levels were reasonable (12 to 15%), although estimated mean volume was significantly different from the real value at 95% probability level (Table 5).
Table 3 – Mean artificial form factors by diameter class in trees sampled in a *Pinus taeda* L. stand in the middle of Paraná State, Brazil.

<table>
<thead>
<tr>
<th>Diameter class (cm)</th>
<th>Form factor</th>
<th>Value</th>
<th>Coefficient of variation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 9.0</td>
<td></td>
<td>0.69</td>
<td>26.57</td>
</tr>
<tr>
<td>9.0 – 12.9</td>
<td></td>
<td>0.45</td>
<td>7.68</td>
</tr>
<tr>
<td>13.0 – 16.9</td>
<td></td>
<td>0.48</td>
<td>8.86</td>
</tr>
<tr>
<td>17.0 – 20.9</td>
<td></td>
<td>0.45</td>
<td>6.65</td>
</tr>
<tr>
<td>21.0 – 24.9</td>
<td></td>
<td>0.45</td>
<td>7.27</td>
</tr>
<tr>
<td>25.0 – 28.9</td>
<td></td>
<td>0.47</td>
<td>10.07</td>
</tr>
<tr>
<td>29.0 – 32.9</td>
<td></td>
<td>0.46</td>
<td>7.24</td>
</tr>
<tr>
<td>33.0 – 36.9</td>
<td></td>
<td>0.43</td>
<td>6.99</td>
</tr>
<tr>
<td>37.0 – 40.9</td>
<td></td>
<td>0.44</td>
<td>8.26</td>
</tr>
<tr>
<td>&gt; 40.9</td>
<td></td>
<td>0.41</td>
<td>-</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>0.47</td>
<td>17.87</td>
</tr>
</tbody>
</table>

Table 4 – Fitting statistics of regression models used to estimate *Pinus taeda* L. stem total volume in the middle of Paraná State, Brazil.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\beta_0$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$R^2_{adj}$</th>
<th>$Syx%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.472558</td>
<td>0.040430</td>
<td>-</td>
<td>0.8552</td>
<td>36.13</td>
</tr>
<tr>
<td>2</td>
<td>-0.102343</td>
<td>0.000906</td>
<td>-</td>
<td>0.9320</td>
<td>25.69</td>
</tr>
<tr>
<td>3</td>
<td>0.006017</td>
<td>0.000035</td>
<td>-</td>
<td>0.9957</td>
<td>11.22</td>
</tr>
<tr>
<td>4</td>
<td>-9.131972</td>
<td>2.559242</td>
<td>-</td>
<td>0.9350</td>
<td>25.20</td>
</tr>
<tr>
<td>5</td>
<td>-9.131972</td>
<td>1.279621</td>
<td>-</td>
<td>0.9350</td>
<td>25.20</td>
</tr>
<tr>
<td>6</td>
<td>-9.725396</td>
<td>1.878337</td>
<td>0.956779</td>
<td>0.9912</td>
<td>12.52</td>
</tr>
<tr>
<td>7</td>
<td>-9.719753</td>
<td>0.944064</td>
<td>-</td>
<td>0.9919</td>
<td>12.55</td>
</tr>
</tbody>
</table>

Table 5 – Bias, precision and accuracy of *Pinus taeda* L. total individual volume estimation models in the middle of Paraná

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean volume</th>
<th>Fitting</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(m$^3$)</td>
<td>Bias</td>
<td>Precision</td>
</tr>
<tr>
<td>Mean f</td>
<td>0.4734*</td>
<td>4.12</td>
<td>14.78</td>
</tr>
<tr>
<td>Classes f</td>
<td>0.4527$_{NS}$</td>
<td>-0.44</td>
<td>10.14</td>
</tr>
<tr>
<td>1</td>
<td>0.4547$_{NS}$</td>
<td>0.00</td>
<td>35.93</td>
</tr>
<tr>
<td>2</td>
<td>0.4547$_{NS}$</td>
<td>0.00</td>
<td>25.56</td>
</tr>
<tr>
<td>3</td>
<td>0.4547$_{NS}$</td>
<td>0.00</td>
<td>11.16</td>
</tr>
<tr>
<td>4</td>
<td>0.4375*</td>
<td>-3.78</td>
<td>25.91</td>
</tr>
<tr>
<td>5</td>
<td>0.4375*</td>
<td>-3.78</td>
<td>25.91</td>
</tr>
<tr>
<td>6</td>
<td>0.4419*</td>
<td>-2.80</td>
<td>13.31</td>
</tr>
<tr>
<td>7</td>
<td>0.4419*</td>
<td>-2.80</td>
<td>13.41</td>
</tr>
<tr>
<td>Real</td>
<td>0.4547</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Volume estimation with mean form factors by diameter class generated low negative bias values, that is, with negligible underestimate (less than 1%), both using fitting and validation data. Estimates with this technique were not statistically different from real values at 95% probability (Table 5). Residue graphical analysis with estimation made from mean form factor showed underestimation for trees of lower diameter class and overestimations for larger trees (Figure 1a). Residue graphical distribution behaved in a balanced way, both for fitting and validation data (Figure 1b).
Among regression models tested, those of numbers 1, 2 and 3 did not show bias in the fitting. There were negative tendencies in the validation, notably with the estimation of models 1 and 2. Precision and accuracy values were high, indicating poor fitting, a fact that was also proved by residue graphical analysis (Figures 1c, d). The paired t-test for comparison between estimated and real values indicated statistical difference in validation, which confirms that the models 1 and 2 do not produce satisfactory estimates.

The model 3 (Spurr) had low bias and lower precision and accuracy values, 11 and 9%, respectively, for fitting and validation. Residuals with this model (Figure 1e) behaved properly and there was no significant difference between estimates and real values.

Models 4 and 5 show the same behavior aforementioned, with negative bias both in fitting and validation, with higher accuracy and precision values (21 to 25%). In addition, estimates statistical analysis against real values showed a significant difference, which implies that such models do not produce satisfactory results. Higher residue amplitude corroborates with a worse fitting of these models compared to model 3 (Figures 1f, g).

Models 6 and 7 showed negative biases in both fitting and validation, besides comparatively lower precision and accuracy values compared to models 1, 2, 4 and 5. However, estimates compared to real data indicated that such models do not produce satisfactory estimates, which was also revealed by residue analysis (Figures 1h, i).

In short, use of mean form factors by diameter class and the Spurr model were equivalent, yielding unbiased, precise and accurate estimates of the stem volume of the species in question. Use of mean form factor and other regression models generate inadequate volume estimates. Schumacher-Hall model (regression model 6) individual analysis, which is almost always unbeatable in volume model tests, had biased, inaccurate and imprecise estimates, losing in performance against the Spurr model and estimation with form factors by diameter class.

One of the most adopted methods to estimate tree commercial volume is the so-called form factor. It corrects conicity by relating stem volume and the...
volume of a regular solid, the cylinder in this case, calculated from the cross-sectional area and height (Thaines et al., 2010). In the past, it was common to calculate total volume by cylindrical volume equation corrected by a fixed form factor (Rolim et al., 2006; Cerdeira, 2012).

Campos & Leite (2009) discussed form factor use to estimate tree volume, where, according to the authors, it is usual to use mean form factor, although there are several observations for its use. Rocha et al. (2010) stated that mean form factor should be used by observing stand characteristics, such as species, site, spacing, thinning and age, since not observing these characteristics may result in inaccurate volume estimates.

Miguel et al. (2010) commented that the form factor can be used for a quick estimation of individual volume or stand, when the crops have similar characteristics. Kohler (2013) mentioned that the literature recommends using variable form factors by age and diameter class to estimate volumes, given the shape variations with such variables. This suggestion is in line with this study, as it was verified that there was an important numerical difference in the form factor between diameter classes.

The main motivation for using the form factor in the estimation of tree stem volumes is its simplicity. However, simplifying also increases the risk of estimation errors. Volume estimation from volume regression models fitted for the location of the study area generally result in higher precision (Cerdeira, 2012). Although volume equations fitted by linear regression usually generate more precise and accurate estimates than using the mean form factor, this is not always the case (Souza & Jesus, 1991; Miranda et al., 2015).

One possibility of improving form factor volume estimation would be to stratify trees by diameter class and to apply such estimation factors, since several studies have shown that the form factor varies with age (Kohler, 2013) and diameter class (Mainardi et al., 1996). This would guarantee simplicity without much use of statistics, and results comparable to those generated by the regression models will be obtained.

Another possibility would be to model the form factor in relation to dbh (Drescher et al., 2001; Souza et al., 2008), although it would also require to use regression, which is very accessible today, but not always used by researchers, who prefer simpler methods. In addition, unsatisfactory results with the fittings were reported in studies with Pinus taeda (Souza et al., 2008) using the approach of this study.

Simple and robust methods are required to estimate volume. Wood volume estimates in forest inventories mainly involve the use of volume equations (Oliveira et al., 2009; Móra et al., 2014). Taper functions are also quite useful for this purpose, although being relatively more complex than volume equations. Their use is justified when assortment is calculated, although their estimates may not have statistical significance in relation to those obtained with traditional equations (Móra et al., 2014).

It has been proved in this study that the direct application of form factor by diameter class ensures as good results as those of the best volume models fitted by regression, with the advantage of its simplicity. Similar studies with Araucaria angustifolia (Bertol.) Kuntze (Sanqueta et al., 2016) corroborate with the findings of this research.

Conclusions

The use of mean form factor by diameter class generated accurate and precise estimates of the total individual volume of the species studied.

Spurr volume model fitted by linear regression generated estimates equally satisfactory to those produced from the application of form factors by diameter classes.

Despite apparently good fittings, not all volume models fitted by regression produced adequate estimates.

The use of form factors by diameter class is simple and robust, and may be safely used in analogous conditions to this study.

The use of a general mean form factor results in biased, inaccurate and imprecise individual total volume estimations, and is not recommended.

References


Cerdeira ALN (2012) Modelos para quantificação do volume de diferentes sortimentos em plantio de Eucalyptus urophylla x Eucalyptus grandis. 66p. Monografia (Graduação em Engenharia Florestal) - Universidade de Brasília.


