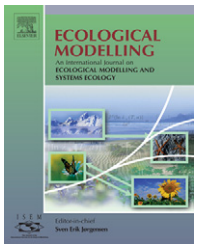
available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/ecolmodel



Improving spatial distribution estimation of forest biomass with geostatistics: A case study for Rondônia, Brazil

Marcio H. Sales[a], *Carlos M. Souza Jr.*[a,b,*], *Phaedon C. Kyriakidis*[b], *Dar A. Roberts*[b], *Edson Vidal*[a,c]

[a] *Instituto do Homem e Meio Ambiente da Amazônia, Imazon, Caixa Postal 5101, Belém, PA 66613-397, Brazil*
[b] *Department of Geography, University of California at Santa Barbara, 1832 Ellison Hall, Santa Barbara, CA 93106-4060, USA*
[c] *Departamento de Ciências Florestais, ESALQ/USP, Caixa Postal 9, 13418-900 Piracicaba, SP, Brazil*

ARTICLE INFO

Article history:
Received 6 April 2006
Received in revised form
15 February 2007
Accepted 20 February 2007
Published on line 27 April 2007

Keywords:
Biomass
Kriging with external drift
Rondônia
Brazilian Amazon

ABSTRACT

Mapping aboveground forest biomass is of fundamental importance for estimating CO_2 emissions due to land use and land cover changes in the Brazilian Amazon. However, existing biomass maps for this region diverge in terms of the total biomass estimates derived, as well as in the spatial patterns of mapped biomass. In addition, no regional or location-specific measure of reliability accompanies most of these maps. In this study, 330 one-hectare plots from the RADAMBRASIL survey, acquired over and along areas adjacent to the state of Rondônia, were used to generate a biomass map over the entire region using geostatistics. The RADAMBRASIL samples were used to generate a biomass map, along with a measure of reliability for each biomass estimate at each location, using kriging with external drift with elevation, vegetation type and soil texture considered as biomass predictor variables. Cross-validation was performed using the sample plots to compare the performance of kriging against a simple biomass estimation using the sample mean. Overall, biomass varied from 225 to 486 Mg ha^{-1}, with a local standard deviation ranging from 62 to 202 Mg ha^{-1}. Large uncertainty values were obtained for regions with low sampling density, in particular in savanna areas. The geostatistical method adopted in this paper has the potential to be applied over the entire Brazilian Amazon region to provide more accurate local estimates of biomass, which would aid carbon flux estimation, along with measures of their reliability, and to identify areas where more sampling efforts should be concentrated.

1. Introduction

There are several estimates of the spatial distribution of forest biomass in the Amazon. These estimates, however, diverge in terms of the total biomass reported and in the identification of sites with high and low biomass concentration (Houghton et al., 2001). Additionally, none of the methods used for estimating the spatial distribution of forest biomass in the Amazon explicitly incorporate the spatial correlation of biomass in their models or are accompanied by measures of reliability. Precise estimates of forest biomass spatial distribution are important for reducing uncertainty in estimating carbon emissions due to deforestation and forest degradation associated with selective logging and fires (Houghton, 2005).

The methods used for estimating spatial distribution of forest biomass in the Amazon may be classified into two

* *Corresponding author at*: Instituto do Homem e Meio Ambiente da Amazônia, Imazon, Caixa Postal 5101, Belém, PA 66613-397, Brazil.
E-mail address: souzajr@imazon.org.br (C.M. Souza Jr.).
0304-3800/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2007.02.033

groups. Methods in the first group employ interpolation or extrapolation of biomass estimates obtained in the field for extensive regions, whereas methods in the second group estimate biomass from remote sensing and environmental model data. As an example of the first group of methods, volume data from the RADAMBRASIL project were converted into total biomass, using conversion equations from Brown and Lugo (1992) and Fearnside (1997), and extrapolated based on vegetation types (Houghton et al., 2001). Another example of this group of methods was a biomass map for the Amazon, generated via interpolation from biomass measurements obtained at only 56 sites (Houghton et al., 2000). Examples of map products produced via the second group of methods are biomass estimates obtained through NDVI (Normalized Difference Vegetation Index) and models of NPP (Net Primary Production) (Potter, 1999).

In this paper, we present a promising approach, based on geostatistics, for overcoming the major problem of forest biomass estimates in the Brazilian Amazon: the lack of explicit spatial information in the interpolation models. To this respect, we modeled the spatial variability of timber volume with a semi-variogram analysis utilizing 330 1 ha forest inventory plots from the RADAMBRASIL project, for the state of Rondônia. Next, we utilized the parameters of the semi-variogram model in a kriging algorithm with external drift, to estimate the spatial distribution of the timber volume together with a reliability measure of the estimated value for each location. Finally, we applied conversion factors from Brown and Lugo (1992) and from Fearnside (1997), to convert the timber volume estimates obtained with the kriging algorithm to forest biomass, and compared the results.

2. Spatial interpolation

Spatial interpolation consists of generating estimates for unsampled sites based on known values from samples collected in the vicinity. Throughout the text, we will refer to the coordinate vector of each location using the letter **s**, to the value of the variable of interest at that location as $z(\mathbf{s})$ and to the ith sample datum as $z(\mathbf{s}_i)$. The estimated value for the variable of interest Z at any unsampled location will be referred to as $\hat{z}(\mathbf{s})$, while the samples located in the vicinity of **s** will be referred to as $z(\mathbf{s}_i)$, $i = 1, \ldots, n$.

In linear spatial interpolation, the unknown value $z(\mathbf{s})$ of the variable of interest at an unsampled location is estimated by $\hat{z}(\mathbf{s})$ as a linear combination of the nearby samples, using:

$$\hat{z}(\mathbf{s}) = \sum_{i=1}^{n} w_i(\mathbf{s}) z(\mathbf{s}_i) \quad (1)$$

where each sample $z(\mathbf{s}_i)$, located near the location **s**, receives a weight factor $w_i(\mathbf{s})$ which reflects the importance of the sample $z(\mathbf{s}_i)$ for estimating the unknown value at location **s**. There are various types of interpolators. With most, it is assumed that samples closest to location **s** are more important than more distant samples for determining the value $z(\mathbf{s})$. What differentiate them are the rules for attributing weights to the sample data. A review of various interpolation methods may be found in other papers (e.g., Caruso and Quarta, 1998).

2.1. *Kriging with external drift (KED)*

Geostatistics is a branch of statistics that explicitly incorporates the concept of spatial correlation in the analysis of spatial data (Isaaks and Srivastava, 1989; Goovaerts, 1997). In the field of ecology and forest management, examples of geostatistical applications include modeling the spatial distribution of tree diameters (Nanos and Montero, 2002), and estimating timber stocks and diameter increments (Biondi et al., 1994).

In geostatistics, the unknown value $z(\mathbf{s})$ at any location is typically decomposed into a mean (drift) component $m(\mathbf{s})$ and a residual component $r(\mathbf{s})$:

$$z(\mathbf{s}) = m(\mathbf{s}) + r(\mathbf{s}) \quad (2)$$

Different kriging variants can be distinguished according to whether the mean component $m(\mathbf{s})$ is assumed constant or spatially variable (Goovaerts, 1997). When the mean component is constant and unknown $m(\mathbf{s}) = m$, ordinary kriging is the procedure for estimating the unknown value at any unsampled location **s**. Alternatively, the mean component can be modeled as spatially variable by expressing it as a function of auxiliary variables (predictors) that vary in space. For example, one can adopt the following linear function:

$$m(\mathbf{s}) = \sum_{a=0}^{A} b_a x_a(\mathbf{s}) \quad (3)$$

where $x_a(\mathbf{s})$ represents the value of the ath external variable at location **s**, and b_a denotes the regression coefficient associated with that external variable; by convention $b_0 = 1$.

The residuals are typically assumed spatially autocorrelated, with zero mean, and covariance function $C_R(\mathbf{h}) = C_R(\mathbf{s} - \mathbf{s}')$, where **h** denotes the separation vector between any two locations **s** and **s**′. Any covariance model must be positive-definite, and various known functions have this characteristic (e.g., Isaaks and Srivastava, 1989; Goovaerts, 1997; Gringarten and Deutsch, 2001). Once the functional form of the covariance model (e.g., spherical or exponential) is postulated, its parameters (relative nugget, sill and range) are usually inferred using non-linear least squares or maximum likelihood (Chilès and Delfiner, 1999).

The estimated value $z(\mathbf{s})$ at any unsampled location is computed using Eq. (1), subject to the following unbiasedness constraints on the kriging weights:

$$\sum_{i=1}^{n} w_i(\mathbf{s}) x_a(\mathbf{s}_i) = x_a(\mathbf{s}), \quad a = 0, \ldots, A \quad (4)$$

where $x_a(\mathbf{s}_i)$ represents the value of the ath external variable collected at the ith sample location.

The associated kriging weights are finally obtained as:

$$\begin{bmatrix} w_1(\mathbf{s}) \\ \vdots \\ w_n(\mathbf{s}) \\ -\mu_0(\mathbf{s}) \\ \vdots \\ -\mu_A(\mathbf{s}) \end{bmatrix} = \begin{bmatrix} C_R(\mathbf{s}_1-\mathbf{s}_1) & \cdots & C_R(\mathbf{s}_1-\mathbf{s}_n) & 1 & \cdots & x_A(\mathbf{s}_1) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ C_R(\mathbf{s}_n-\mathbf{s}_1) & \cdots & C_R(\mathbf{s}_n-\mathbf{s}_n) & 1 & \cdots & x_A(\mathbf{s}_n) \\ 1 & \cdots & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_A(\mathbf{s}_1) & \cdots & x_A(\mathbf{s}_n) & 0 & \cdots & 0 \end{bmatrix}^{-1} \cdot \begin{bmatrix} C_R(\mathbf{s}_1-\mathbf{s}) \\ \vdots \\ C_R(\mathbf{s}_n-\mathbf{s}) \\ 1 \\ \vdots \\ x_A(\mathbf{s}) \end{bmatrix} \quad (5)$$

where $\mu_A(\mathbf{s})$ denotes the Lagrange parameter associated with the Ath unbiasedness constraint on the weights (Goovaerts, 1997). The term $C_R(\mathbf{s}_n-\mathbf{s}_1)$ denotes the residual covariance between two sample locations $\mathbf{s}_n$ and $\mathbf{s}_1$, whereas the term $C_R(\mathbf{s}_n-\mathbf{s})$ denotes the residual covariance between the nth sampled location $\mathbf{s}_n$ and the unsampled location $\mathbf{s}$. It should be noted that KED reduces to classical OLS regression if the spatial autocorrelation in the residuals is negligible. An important detail in the application of KED is that it requires the knowledge of the residual variogram model. In this work, we estimated a variogram model from the residuals obtained from an aspatial linear regression model, as described in Section 3.

Last, the associated minimum error variance at an unsampled location $\mathbf{s}$ is given by:

$$\hat{\sigma}_E(\mathbf{s}) = \sigma_R - \sum_{i=1}^{n} w_i(\mathbf{s}) C_R(\mathbf{s}_i-\mathbf{s}) + \sum_{a=0}^{A} \mu_a(\mathbf{s}) x_a(\mathbf{s}) \quad (6)$$

where σ_R denotes the a prior (not conditional to nearby data) variance of the residual component.

The kriging estimates and associated variance values derived from the above equations account for both the auxiliary variables and the spatial correlation in the residuals. Kriging with external drift is essentially a spatial (or generalized) linear regression model (Chilès and Delfiner, 1999), and offers a convenient means for incorporating additional data into spatial interpolation, while reporting the associated uncertainty in the resulting predictions.

3. Methodology

3.1. Study area

The State of Rondônia is located in the southwestern Amazon between latitude 8° and 15°S, and longitude 60° and 65°W (Fig. 1). The annual rainfall rate is 2600 mm year^{-1}, with the rainy season lasting from November to April (Brown et al., 1995). The elevation varies from flat soils to undulating terrain between 80 and 140 m above sea level, and the predominant soil types are red-yellow latosols and red-yellow lithic podzols (Holmes, 2003). The dominant vegetation is open moist tropical forests, with the presence of palms and lianas (DNPM, 1978). The state is characterized by having a history of high deforestation rates (INPE, 2003; Roberts et al., 2002). From 1994 to 2004, 12% of Rondônia was deforested, which makes the state one of the major carbon emitters due to deforestation. Deforestation in the state is principally caused by the activities of small farmers, ranchers, and mining and timber companies (Pedlowski et al., 1997).

3.2. Data base

We used the volume data from a subset of forest inventories from the RADAMBRASIL project (DNPM, 1978) collected in 1-ha plots. The subset used is made up of all of the samples located within the state of Rondônia with a buffer of 100 km around it (Fig. 1). Plots from outside of the state were included because they are important for calculating estimates at locations near its boundaries. According to RADAMBRASIL's reports, the sampling design was a stratified random sampling limited by accessibility. The same report states that in some cases, samples were intentionally located to ensure that the maximum number topographical profiles possible were covered. Thus, we assumed that the data set was representative of the variability of wood volume over the study area.

The RADAMBRASIL forest type and soil maps were utilized as explanatory variables, because they show evidence of a relationship to forest biomass (Laurance et al., 1999; Houghton et al., 2001). We also included terrain elevation data generated by the Shuttle Radar Topographic Mission (SRTM) (http://srtm.usgs.gov/) with 90 m resolution as an explanatory variable. All of the maps were resampled to a 1 km resolution.

3.3. Volume spatial distribution modeling

The stages of modeling and calculation of the biomass estimates for Rondônia are summarized in Fig. 2. We first used the explanatory variables to establish the mean component of timber volume at sample locations. We used this model to compute sample residuals and then modeled their spatial variability via variogram analysis. We finally used the explanatory variables at all locations in kriging with external drift to interpolate timber volume at a 1 km grid over the state of Rondônia, as described in Section 2. Finally, we converted the tree volume from the forest inventories into biomass using conversions from Brown and Lugo (1992) and Fearnside (1997). The conversion details are presented in Section 3.5.

More precisely, the mean $m(\mathbf{s})$ at any location was specified using the following linear model:

$$m(\mathbf{s}) = b_0 + \mathbf{b}_1^T \text{florst}(\mathbf{s}) + \mathbf{b}_2^T \text{solo}(\mathbf{s}) + b_3 \text{SRTM}(\mathbf{s}) + \mathbf{b}_4^T \text{florst}(\mathbf{s}) \times \text{SRTM}(\mathbf{s}) + \mathbf{b}_5^T \text{solo}(\mathbf{s}) \times \text{SRTM}(\mathbf{s}) \quad (7)$$

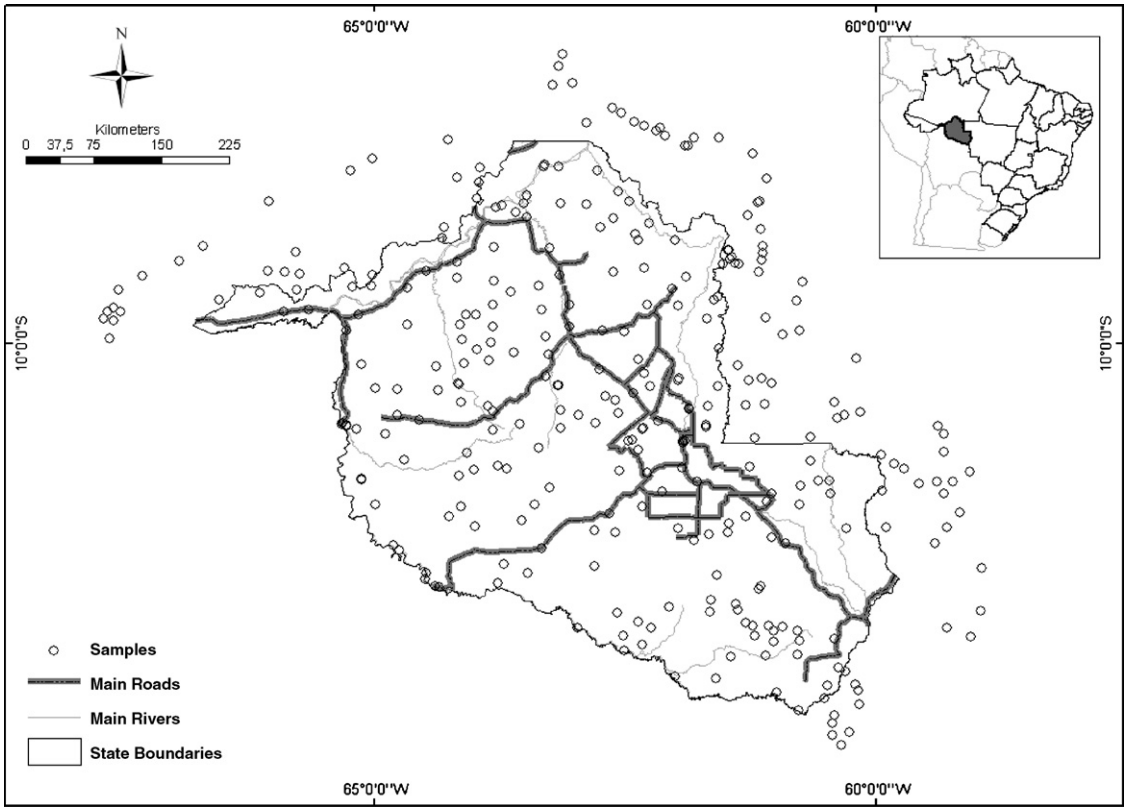


Fig. 1 – Study area and location of RADAMBRASIL forest surveys used in this study.

where florst(**s**) and solo(**s**) are column vectors of binary values representing the vegetation type and soil texture classes, respectively, at location **s**. The term SRTM(**s**) represents terrain elevation—the only continuous predictor variable. The terms florst(**s**) × SRTM(**s**) and solo(**s**) × SRTM(**s**) represent the interaction between forest type and elevation, and between soil texture and elevation, respectively. Therefore, except for b_3, all $\boldsymbol{b}$s are vectors of regression coefficients matching each column vector of binary variables in Eq. (7). For example, the term $\boldsymbol{b}_1^{\mathrm{T}}\mathrm{florst}(\mathbf{s})$ is the vector notation for the sum $b_{11}\mathrm{florst}(\mathbf{s})_2 + b_{21}\mathrm{florst}(\mathbf{s})_3 + b_{31}\mathrm{florst}(\mathbf{s})_4 + b_{41}\mathrm{florst}(\mathbf{s})_5$, where each $\mathrm{florst}(\mathbf{s})_i$ is 0 or 1 depending on the forest type in location **s**. The regression coefficients were estimated using ordinary least squares (OLS). Detailed explanation on this classical regression approach can be found in the literature (e.g., Draper and Smith, 1998).

The residuals were then obtained using $r(\mathbf{s}_i) = z(\mathbf{s}_i) - \hat{m}(\mathbf{s}_i)$, $i = 1, \ldots, n$, which supplies the differences between the sample data and the estimated mean component at the sample locations. Next, to model the spatial variability of the residuals, we first calculated experimental variograms of the resulting residuals along four different directions to assess the possible existence of anisotropy (the phenomenon in which the variogram behavior depends on direction). No anisotropy was detected, however, and we thus considered only the omnidirectional experimental residual variogram.

Next, a positive-definite variogram model (with nested structures) was selected based on its average squared error in adjusting the residuals between that model and the sample variogram values. For more details on positive-defined models, see Gringarten and Deutsch (2001). We used weighted least squares with the number of samples in each distance class as weights, to determine the parameters of the nested structure: the nugget effect, partial sills and ranges. The nugget effect is a discontinuity at the origin of the variogram, generally attributed to small scale variability or measurement error. A sill is a variance value that bounds the variogram of the variable, while the range is the distance at which the variogram reaches the sill. It should be noted here that a somewhat more general (but computationally more demanding) approach to residual variogram fitting is maximum likelihood (typically Gaussian) fitting. This latter method is often advocated due to the experimental variogram of the residuals being biased, particularly at large lag distances. We did not opt for this method in this work, due to the relative robustness of kriging with respect to variogram model misspecification. Future work will address this issue in a more robust way. For more details on variogram fitting procedures, one can consult Chilès and Delfiner (1999).

Once the residual variogram was modeled, the next step was to obtain the weights for kriging with external drift at each grid cell using Eq. (5). To do this, we used the residual variogram model and the elevation, forest type and soil texture values. Finally, we utilized the resulting weights in Eq. (1) to compute the volume estimate at each grid cell.

3.4. *Conversion of timber volume to biomass*

The volume maps obtained by kriging were converted into total biomass maps by applying two methodologies for con-

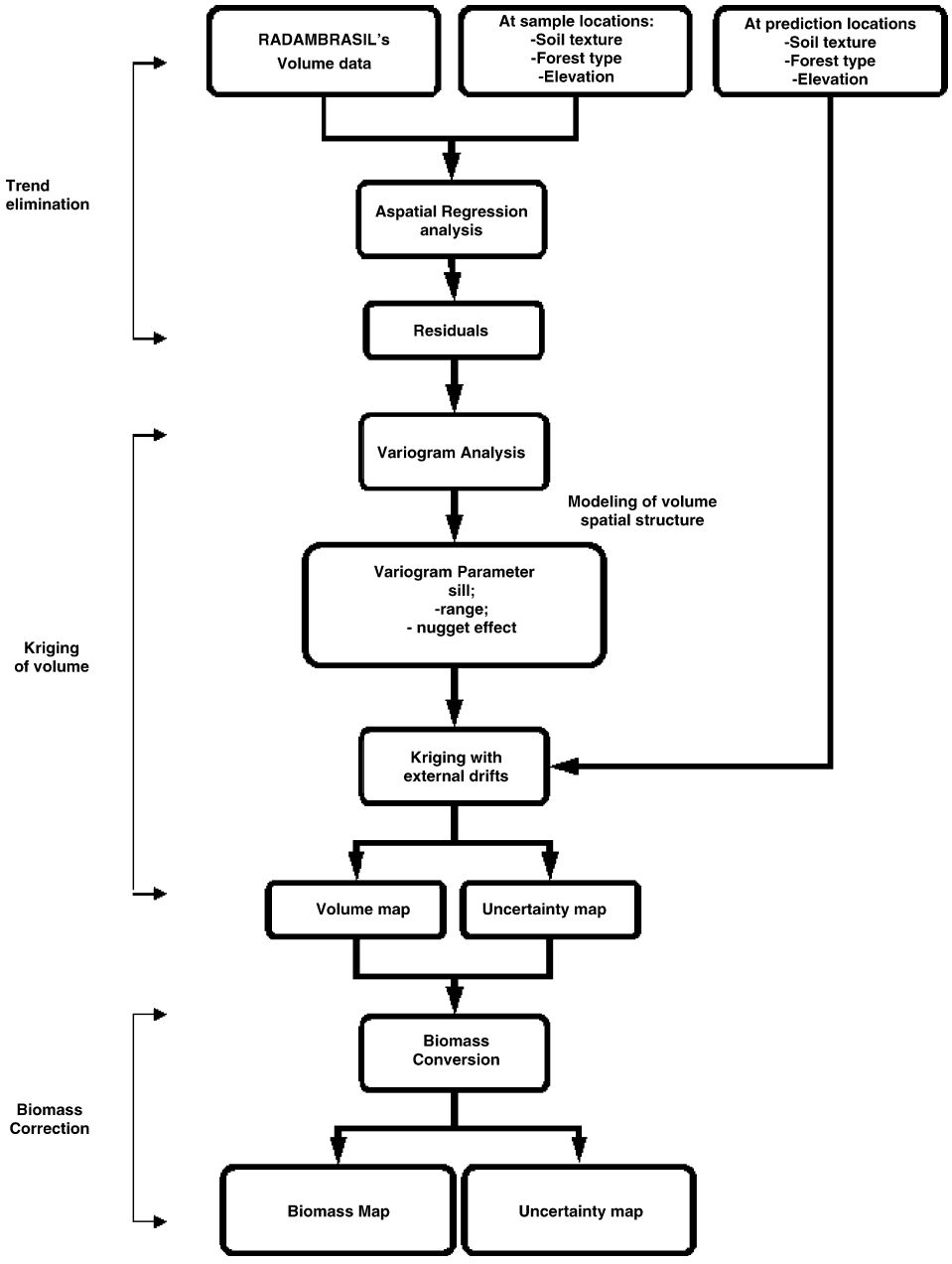


Fig. 2 – Methodology used for modeling residual spatial correlation and estimating forest biomass spatial distribution in Rondônia.

verting volume to biomass: that of Brown and Lugo (1992), modified by Houghton et al. (2001), and that of Fearnside (1997). Fearnside's conversion includes additional components of biomass not included in Brow and Lugo's. We applied both methods to evaluate which one is most appropriate, based on the range of biomass estimates observed in the literature. The Brown and Lugo equation (1992), modified by Houghton et al. (2001), is given by:

$$\mathrm{AGB(BL)} = \mathrm{SB} \times \mathrm{BEF} \times (1 + (0.09 + 0.21)) \quad (8)$$

where SB is the stem-wood biomass, and is given by:

$$\mathrm{SB} = \mathrm{Volume} \times \mathrm{VEF} \times \mathrm{WD} \quad (9)$$

with VEF (Volume Expansion Factor) being the correction factor for including trees below a diameter of less than 30 cm of DBH (define DBH) and WD being wood density. All of the above terms pertain to local values, but we have dropped the dependence on the coordinate vector **s** for notational simplicity. We used VEF = 1.25 for dense forest, and 1.50 for all other vegetation types. The value used for WD was assigned as 0.69 $\mathrm{Mg\,m^{-3}}$ for all forests, and 0.404 $\mathrm{Mg\,m^{-3}}$ for savannas, as suggested

by Barbosa and Fearnside (2004). The value of BEF (Biomass Expansion Factor) in Eq. (7) is utilized to compensate for the difference in stem biomass between small and large trees, and was given by:

$$\mathrm{BEF} = \exp(3.213 - 0.506\ln(\mathrm{SB})), \quad \text{if } \mathrm{SB} < 190,$$
$$\mathrm{BEF} = 1.74, \quad \text{if } \mathrm{SB} > 190 \tag{10}$$

In Eq. (8), SB × BEF is the conversion of volume to aboveground biomass originally proposed by Brown and Lugo (1992). The constants 0.09 and 0.21 in the formula were introduced by Houghton et al. (2001) to include belowground biomass and dead aboveground biomass.

Fearnside (1997) suggested the following modification in Brown and Lugo's equation (1992):

$$\mathrm{AGB(Fs)} = \mathrm{SB} \times \mathrm{BEF} \times (1 + \mathrm{CF}) \tag{11}$$

where CF = 96.2%, which represents the sum of various correction factors (lianas = 5.3%, trees smaller than 10 cm DBH = 12%; tree form factor = 15.6%; for trees between 30 and 31.8 cm DBH = 3.6%; hollow trees = −6.6%; bark = −0.9%; palms = 2.4%, belowground biomass = 33.6%; dead aboveground biomass soil = 31%; and other components = 0.2%).

In this paper, the correction factors from Eqs. ((8)–(11)) were assumed constant, i.e., not spatially variable. Therefore, the biomass estimates and the standard deviations of the associated prediction errors were obtained by multiplying the correction factors by the volume estimates and the associated prediction error standard deviation maps, respectively.

3.5. *Validation*

We applied cross-validation to assess the performance of kriging in estimating the spatial distribution of timber volume in relation to the estimates obtained using a simple sample average. Cross-validation is an iterative process where for each step a sample is hidden from the data set and the others are used for generating an estimate for the excluded sample value. This process is repeated for each sample, and the cross-validation error $z(\mathbf{s}_i) - \hat{z}(\mathbf{s}_i)$ is estimated at each sample location $\mathbf{s}_i$. We utilized the root mean squared error as a performance assessment criterion for the two methods, given by:

$$\mathrm{RMSE_K} = \sqrt{\frac{\sum_{i=1}^{n} [z(\mathbf{s}_i) - \hat{z}_K(\mathbf{s}_i)]^2}{n}} \tag{12}$$

$$\mathrm{RMSE_A} = \sqrt{\frac{\sum_{i=1}^{n} [z(\mathbf{s}_i) - \hat{z}_A(\mathbf{s}_i)]^2}{n}} \tag{13}$$

where $\hat{z}_K(\mathbf{s}_i)$ denotes the kriging estimate and $\hat{z}_A(\mathbf{s}_i)$ denotes the sample average estimate at each location (note that the latter is constant within each search neighborhood).

The quantity $(\mathrm{RMSE_A} - \mathrm{RMSE_K})/\mathrm{RMSE_A}$ was used to assess the relative reduction of the estimation error $\mathrm{RMSE_K}$ obtained by kriging in relation to the error $\mathrm{RMSE_A}$ obtained by simple averaging.

4. Results

4.1. *Residuals and variogram analysis*

Table 1 shows the OLS-derived coefficients of the regression model in Eq. (7) and their respective standard errors. The regression coefficients for the indeterminate and clay-rich soil texture classes show negative values, suggesting that sandy soils accumulate more volume than the indeterminate and clay-rich classes (Table 1). The interaction between terrain elevation and soil texture presents positive coefficients for the indiscriminate and clay-rich classes, which means that for higher elevations, more clay-rich soils accumulate more volume. When the other variables of the model are kept constant, the model returns higher estimates of volume in clay-rich soils than in sandy soils at elevations greater than 27 m, which represents a larger portion of the state. The same holds true for indiscriminate soils, beginning at an elevation of 135 m (80% of the state). This result is in line with what was expected, given that clay-rich soils have a higher capacity for retaining nutrients than sandy soils. This result is also in agreement with the results found by Laurance et al. (1999), in a study of the relation between soils and biomass in the central Amazon, where they found that the clay gradient for soils is the soil variable with the greatest explanatory power for determining forest biomass. Similarly, the interaction between vegetation type and elevation for the classes of open and dense moist

Table 1 – Results of the aspatial regression model for volume

Predictor variables	Coefficients	Standard deviation
Soil texture		
Sandy (omitted class)	–	
Indiscriminate	−18.715	20.090
Clay-rich	−4.0110	15.132
Forest type		
Stacional forests (omitted class)	–	
Ombrophyle open forests	53.715	20.842
Ombrophyle dense forests	64.956	21.404
Pioneers	23.484	51.351
Savannas	40.938	36.044
Elevation	0.108	0.084
Interaction: elevation × soil texture		
Indiscriminate—elevation	0.143	0.085
Clay—elevation	0.151	0.068
Interaction: elevation × forest type		
Ombrophyle open—elevation	−0.262	0.077
Ombrophyle dense—elevation	−0.266	0.088
Pioneers—elevation	−0.248	0.264
Savannas—elevation	−0.181	0.126
Intercept	66.171	22.204

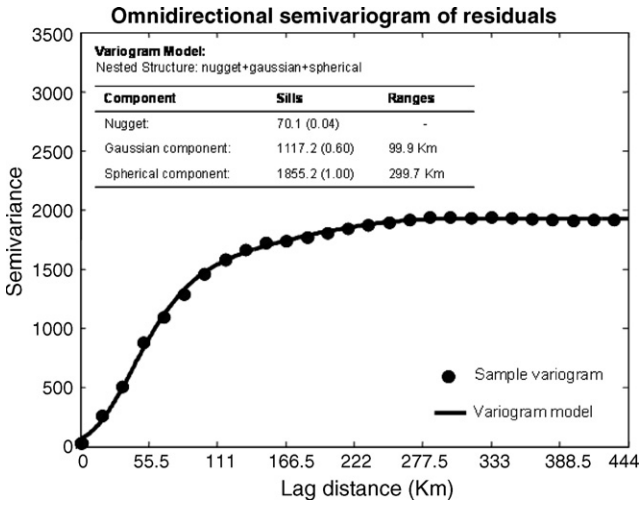


Fig. 3 – Omni-directional sample residual variogram and parameters of the fitted model.

tropical forest are associated with negative coefficients, indicating that the volume in these areas is less in higher elevation areas. The model yields greater volume estimates for seasonal forests than for open and dense moist forests beginning at altitudes of 208 m (36%) and 250 m (23%), respectively.

The directional variograms of the residuals did not show a strong pattern of anisotropy, and consequently we considered an isotropic model for the spatial variability of residual timber volume resulting from the application of Eq. (7). To do this, we adjusted a composite variogram function comprised of a nugget component, a Gaussian structure and a spherical structure to fit the sample residual variogram. The final omni-directional variogram for residual timber volume is shown in Fig. 3, and presents a total range of 300 km, and an estimated relative nugget of only 4%. This implies that the residual timber volume has a strong spatial correlation. The variogram model indicates that the spatial correlation of residual volume decreases rapidly with distance of up to 100 km, where the variogram reaches 60% of the sill under Gaussian behaviour and that correlation decreases more slowly up to the 300 km limit, and is almost zero after this distance. Nevertheless, we retained all samples in the neighborhood of the entire variogram range of 300 km, in the computation of the estimates.

4.2. *Kriging*

Fig. 4 shows the maps of timber volume estimates and their standard error. The estimated timber volume varied from 50 to 150 m^3/ha (1st and 99th percentiles), with standard error varying from 18 to 43 m^3/ha. The discontinuous aspect of the maps of the external or predictor variables used to build the mean component resulted in a predominantly discontinuous behavior for the volume and standard error maps, mainly in regions with low sampling density, such as the extreme north of the state (Fig. 4). In some regions, however, it is possible to perceive a more continuous behavior, mainly in the uncertainty map, which is a result of the residual variogram model; this effect is most visible at locations with greater sampling density. The standard error estimates were high in savanna areas, which are low sampling density regions in the RADAM samples.

Fig. 5 presents the maps of total biomass and their standard error, resulting from application of the equations of Brown and Lugo and Fearnside, described in Section 3.4. With the Brown and Lugo equation, biomass estimates in Rondônia varied from 225 to 486 Mg ha^{-1} (1st and 99th percentiles), with standard error varying from 62 to 202 Mg ha^{-1}. According to the Fearnside equation, biomass varied from 340 to 733 Mg ha^{-1}, with standard error varying from 94 to 306 Mg ha^{-1}. The Fearnside corrections resulted in a 62% increment in relation to estimates obtained using the Brown and Lugo's (1992). Although the Fearnside corrections (1997) included several components not considered by Brown and Lugo (1992), the latter generated estimates much closer to those found in the literature for different types of forest in the Amazon; see, for example, Laurance et al. (1999), Houghton et al. (2001), and Cummings et al. (2002).

4.3. *Validation*

Table 2 shows the RMSE values obtained from the cross-validation procedure. The accuracy of the methods varied as

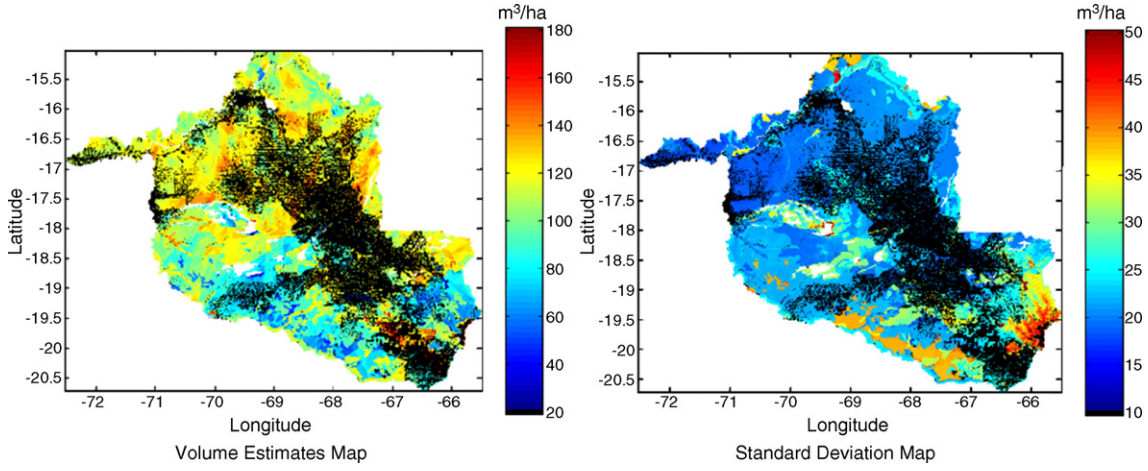


Fig. 4 – Volume estimates and standard errors.

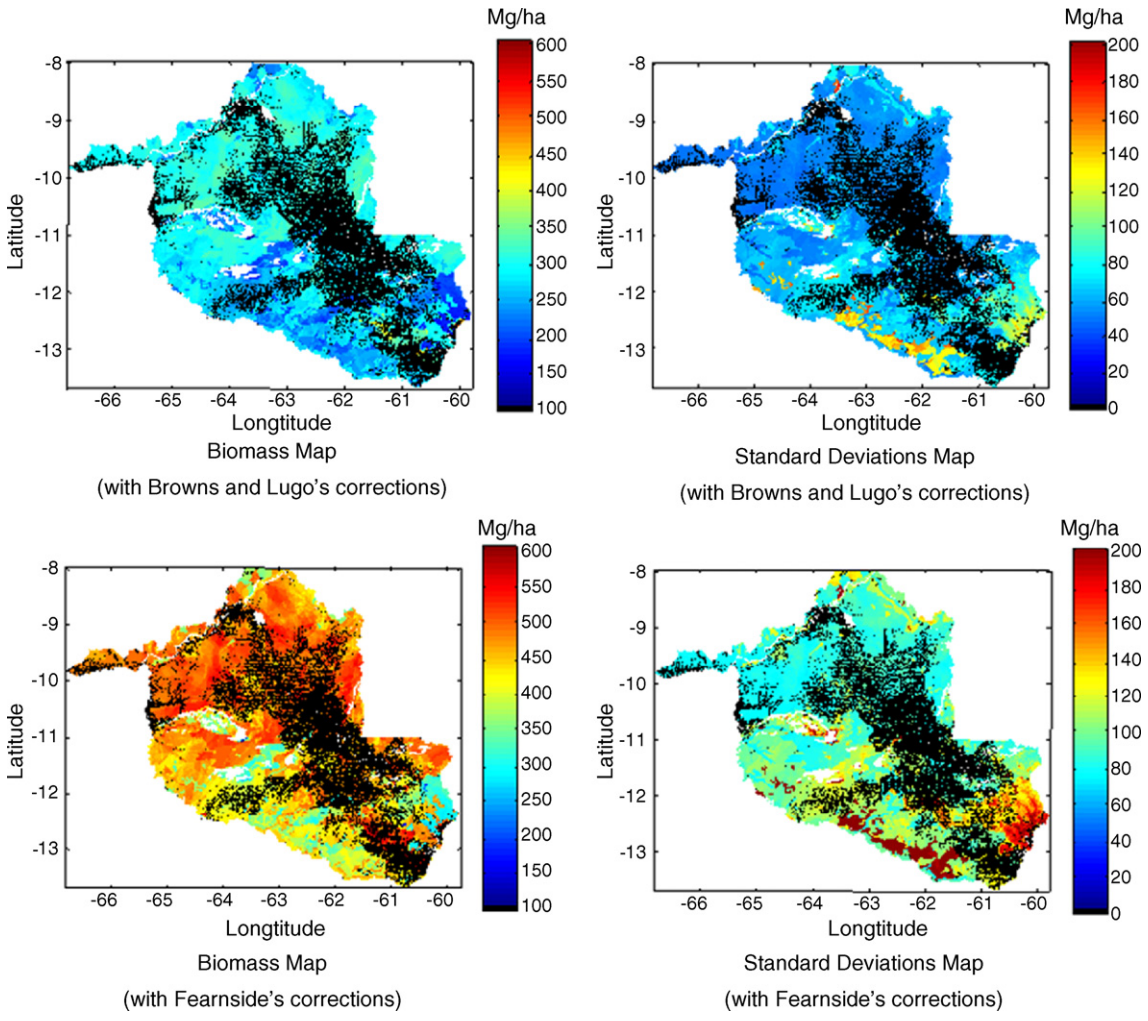


Fig. 5 – Biomass estimates and standard errors.

a function of the sampling density. Fig. 6 shows the relative reduction in RMSE for kriging with external drift in relation to simple averaging, according to the number of samples located within a 300 km radius from any estimation location. The model does not reduce RMSE in locations with less than 50 samples in a radius of 300 km. But for locations with more than 110 samples within this distance radius, the reduction rises to approximately 27%. This effect may have been caused by errors in the coordinates, since, as seen previously, the spatial variability model adopted for residual volume attaches great importance to samples located near the estimation location. Thus, small errors in the coordinates of these samples may generate significant errors in the estimates. However, a large number of samples – so long as they are not clustered – means that smaller weights will be given to all the samples. This implies that the estimated value at each location is comprised of small contributions from samples at different locations, canceling the error effect due to coordinate misspecification. The graph of Fig. 6 suggests a growing rate of error decrease starting with 110 samples in a radius of 300 km, but the data were not sufficient to observe at which point this rate would stabilize. This result implies that our method significantly improves estimates for a sample intensity beyond 110 samples for each 300 km, what is equivalent to a sample intensity of approximately one 0.36 ha/km^2, or a sampling intensity of approximately 0.36%.

Table 2 – RMSE using the sample average and the kriging prediction within each neighborhood by sample size within 300 km from prediction locations

Method	RMSE					
	$N > 0$ (any sample size)	$N > 30$ samples	$N > 50$ samples	$N > 70$ samples	$N > 90$ samples	$N > 110$ samples
Kriging	48.06	48.06	48.02	46.35	41.81	35.40
Simple mean	48.79	48.79	48.79	48.79	48.79	48.79

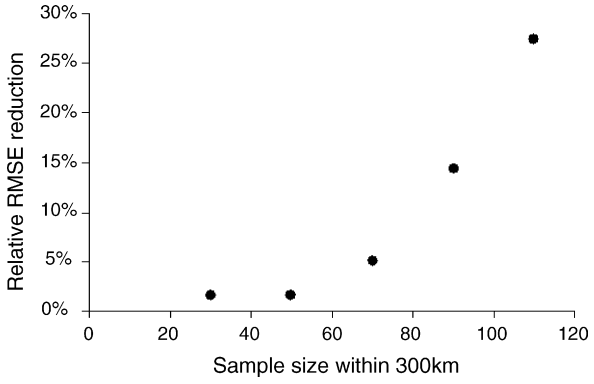


Fig. 6 – Relative reduction in RMSE when comparing kriging with simple averaging, by sample size within a 100 km radius from prediction locations.

5. Discussion and conclusions

Forest biomass spatial distribution estimation for Amazônia is important to derive more accurate rates of carbon emission due to deforestation. Here, we provide a promising approach to improve the carbon emission estimates of this region using geostatistics. We found through variogram analysis that 96% of the variation in residual volume is spatially structured, that correlation being most significant up to a distance of 300 km. This implies that geostatistical methods are more appropriate for generating estimates of biomass distribution than the use of simple average or statistical models that ignore spatial correlation. In Rondônia, biomass values obtained from volume estimates derived by kriging, varied from 225 to 486 $Mg\,ha^{-1}$ using the Brown and Lugo corrections and from 340 to 733 $Mg\,ha^{-1}$ using the Fearnside corrections; the estimates generated by the first method were more similar to results from field measurements in the Amazon. Forest biomass in the state is more concentrated in the region to the north, where there is a predominance of open forests in clay-rich soils. The uncertainty maps reveal that a greater sampling effort is needed for estimating forest biomass accurately in savanna and pioneer forest regions in the state.

The methods applied in this study may be used for generating estimates for the spatial distribution of forest biomass and its uncertainty for the entire Amazon. As already stated, such maps have the potential for improving estimates of carbon flow to the atmosphere as a result of deforestation, which are frequently computed using only data of the annual area of deforested forests combined with estimates of average biomass per hectare in the region (Houghton, 2005), ignoring regional fluctuations and trends of forest biomass in the deforested areas. The use of such maps, together with detailed information on the spatial profile of deforested areas may improve estimates of carbon emissions, principally in deforested regions whose biomass concentration is quite different from the average (Houghton, 2005). Additionally, the uncertainty map offers a means for evaluating the reliability of estimates, allowing the identification of regions with major uncertainties to prioritize data collection or apply other means for estimating biomass.

Biomass estimation in the Amazon is still hindered by barriers that must be overcome. The scarcity of data is still one of the largest of these barriers. In this paper, we track one of the potential sources of errors in biomass estimation by explicitly accounting for the spatial variability of timber volume in interpolation models. Uncertainties derived from field measurements and the application of the biomass conversion factors were not taken into account. Consequently, there is a great need to understand the uncertainty related to the different correction factors applied, which are averages extrapolated to the entire Amazon, derived from data of only a few and small regions. The use of other continuous predictor variables such as fraction images derived from remotely sensed optical imagery could improve the performance of the methodology presented here.

REFERENCES

Barbosa, R., Fearnside, P., 2004. Wood density of trees in open savannas of the Brazilian Amazon. For. Ecol. Manage. 199, 115–123.

Biondi, F., Myers, D.E., Avery, C.C., 1994. Geostatistically modeling stem size and increment in an old-growth forest. Can. J. For. Res. 24, 1354–1368.

Brown, F., Martinelli, L.A., Thomas, W.W., Moreira, M.Z., Ferreira, C.A.C., Victoria, R.A., 1995. Uncertainty in the biomass of Amazonian forests: an example from Rondonia. For. Ecol. Manage. 75, 175–189.

Brown, S., Lugo, A., 1992. Aboveground biomass estimates fro tropical forests of the Brazilian Amazon. Interciência 17, 8–18.

Caruso, C., Quarta, F., 1998. Interpolation methods comparison. Comput. Math. Appl. 35 (12), 109–126.

Chilès, J.-P., Delfiner, P., 1999. Geostatistics: Modeling Spatial Uncertainty. Wiley, New York, 720 pp.

Cummings, D., Kauffman, J., Perry, D., Hughes, R., 2002. Aboveground biomass and structure of rainforests in the southwestern Brazilian Amazon. For. Ecol. Manage. 163 (1–3), 293–307.

Departamento Nacional de Produção Mineral-DNPM, 1973–83. Projeto RADAMBRASIL: Levantamento de Recursos Naturais, vols. 1–23, Ministério das Minas e Energia, Departamento Nacional de Producão Mineral, Rio de Janeiro, Brazil.

Draper, N.R., Smith, H., 1998. Applied Regression Analysis, 3rd ed. Wiley, New York, 736 pp.

Fearnside, P., 1997. Greenhouse gases from deforestation in Brazilian Amazonia: net committed emissions. Clim. Change 35, 321–360.

Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation, Oxford University Press, 483 pp.

Gringarten, E., Deutsch, C., 2001. Variogram interpretation and modeling. Math. Geol. 33 (4), 507–534.

Holmes, K.W., 2003. Regional effects of deforestation on soil biochemistry in the southwest Amazon. Ph.D. Dissertation. University of California in Santa Barbara, 186 pp.

Houghton, R.A., Skole, D.L., Nobre, C.A., Hackler, J.L., Lawrence, K.T., Chomentowski, W.H., 2000. Annual fluxes of carbon from deforestation and regrowth in the Brazilian Amazon. Nature 403, 301–304.

Houghton, R.A., Lawrence, K.T., Hackler, J.L., Brown, S., 2001. The spatial distribution of forest biomass in the Brazilian Amazon: a comparison of estimates. Global Change Biol. 7, 731–746.

Houghton, R.A., 2005. Aboveground forest biomass and the global carbon balance. Global Change Biol. 11, 945–958.

Instituto Nacional de Pesquisas Espaciais (INPE), 2003. Projeto de estimativa de desflorestamento da Amazônia—PRODES. http://www.obt.inpe.br/prodes/.

Isaaks, E.H., Srivastava, R.M., 1989. Applied Geostatistics. Oxford University Press, 561 pp.

Laurance, W., Fearnside, P., Laurance, S., Delamonica, P., Lovejoy, T., Merona, J., Chambers, J., Gascon, C., 1999. Relationship between soils and Amazon forest biomass: a landscape study. For. Ecol. Manage. 118, 127–138.

Nanos, N., Montero, G., 2002. Spatial prediction of diameter distribution models. For. Ecol. Manage. 161, 147–158.

Pedlowski, M., Dale, V., Matricardi, E., Silva Filho, E., 1997. Patterns and impacts of deforestation in Rondônia, Brazil. Landscape Urban Plan. 38, 149–157.

Potter, C.S., 1999. Terrestrial biomass and the effects of deforestation on the global carbon cycle. Bioscience 49, 769–778.

Roberts, D.A., Numata, I., Holmes, K., Batista, G., Krug, T., Monteiro, A., Powell, B., Chadwick, O.A., 2002. Large area mapping of land-cover change in Rondônia using multitemporal spectral mixture analysis and decision tree classifiers. J. Geophys. Res. 107 (20), 40-1–40-18.