

The use of Pedotransfer functions and the estimation of carbon stock in the Central Amazon region

Andréa da Silva Gomes¹, Ana Carolina de Souza Ferreira¹, Érika Flávia Machado Pinheiro¹, Michele Duarte de Menezes², Marcos Bacis Ceddia^{1*}

¹Federal Rural University of Rio de Janeiro/Institute of Agronomy – Dept. of Soil, BR-465, km 7 – 23897-000 – Seropédica, RJ – Brazil.

²Federal University of Lavras – Dept. of Soil Science, C.P. 3037 – 37200-000 – Lavras, MG – Brazil.

*Corresponding author <marcosceddia@gmail.com>

Edited by: Francesco Montemurro

Received August 04, 2016

Accepted November 18, 2016

ABSTRACT: Computer models have been used to assess soil organic carbon (SOC) stock change. Commonly, models require to determine soil bulk density (D_b), a variable that is often lacking in soil data bases. To partly overcome this problem, pedotransfer functions (PTFs) are developed to estimate D_b from other easily available soil properties. However, only a few studies have determined the accuracy of these functions and quantified their effects on the final quality of the spatial variability maps. In this context, the objectives of this study were: i) to develop one PTF to estimate D_b in soils of the Brazilian Central Amazon region; ii) to compare the performance of PTFs generated with three other models generally used to estimate D_b in soils of the Amazon region; and iii) to quantify the effect of applying these PTFs on the spatial variability maps of SOC stock. Using data from 96 soil profiles in the Urucu river basin in Brazil, a multiple linear regression model was generated to estimate D_b using SOC, pH, sum of basic cations, aluminum (Al^{+3}), and clay content. This model outperformed the three other PTFs published in the literature. The average estimation error of SOC stock using our model was 0.03 Mg C ha⁻¹, which is markedly lower than the other PTFs (1.06 and 1.23 Mg C ha⁻¹, or 15 % and 17 %, respectively). Thus, the application of a non-validated PTF to estimate D_b can introduce an error that is large enough to skew the significant difference in soil carbon stock change.

Keywords: Içá Formation, multiple linear regression, ordinary kriging

Introduction

Pedotransfer functions (PTFs) are predictive models of certain soil properties using data from soil surveys (Bouma, 1989). These functions fill the gap between the available soil data and the properties that are more useful or required for a particular model or quality assessment. In this work, PTFs are used as physical-mathematical models that allow the estimation of D_b from soil data, which are needed to convert carbon contents from percentage of dry weight to carbon mass per unit of area (Howard et al., 1995; Benites et al., 2007).

The lack of D_b data is a limiting factor in regions such as the Brazilian Amazon. With a few exceptions, the lack of detailed maps along the region is partly due to: a) accessibility (often only possible by boat or airplane), b) a vast territory extension is covered by the Tropical Amazon Forest, and c) lack of more detailed ancillary maps, including topographic, geologic, and geomorphologic maps. Bernoux et al. (1998) and Tomasella and Hodnett (1998) provided the first baseline to predict D_b from databases. These authors used data from the RADAMBRASIL project (RADAMBRASIL, 1978). More recently, Benites et al. (2007) used data from the Soil Archives of Embrapa (Brazilian Corporation of Agriculture Research) to develop a PTF to estimate D_b for most Brazilian biomes. These three PTFs are frequently applied to predict D_b of soils in Brazil (Bernoux et al., 2002).

Typically, PTF-estimated attributes are used directly in numerical modeling, although the accuracy of the attribute estimates and their effect on modeling

results are not often investigated or they are simply ignored (Deng et al., 2009). Despite the importance of these SOC stock estimations for Brazilian soils in the Amazon region, it is important to highlight that most studies that have applied the PTFs above to estimate D_b do not present quantitative information about the error caused by the application of these PTFs. Bernoux et al. (1998), reported that D_b estimation using their PTF could lead to an error of ≤ 10 % in the final calculation of carbon stock (CS) in 323 horizons used in the study.

In this context, the purpose of this study was to predict D_b from readily available soil properties of Içá Formation in the Brazilian Amazon region, considering the minor effort and uncertainty principles. We have also compared the performance of the models generated in this study with those in the literature (Bernoux et al., 1998; Tomasella and Hodnett, 1998; Benites et al., 2007). Finally, we quantified the error due the application of different PTFs to estimate D_b in the spatial variability of soil carbon stock.

Materials and Methods

Study site, soils, and the database

The study site is located in the central region of the Amazon State near the Urucu River in the municipality of Coari, Brazil, more specifically between the geographic coordinates 4° 45'S and 65° 25'W and the average elevation is 60 m above the sea level (Figure 1). The region is located about 640 km from the state capital Manaus, and it can only be accessed by boat

or airplane. The climate is equatorial (Af-Köppen Climate Classification) where the temperature in the coldest month is higher than 20 °C, with no pronounced dry period and a mean annual precipitation of 2500 mm.

The sampled area covers about 80 km², spanning longitudinally across the Içá Geologic Formation, which is part of the *Solimões* Geologic Domain. The soils of the study site were formed from sediments of Içá Geologic Formation (RADAMBRASIL, 1978). The Içá Formation consists of fine to medium sandstone and siltstone with clay conglomerates and yellow-red. The Holocene alluvium of the Quaternary Period deposits is related to the current Amazonian drainage networks. The sediments of Içá Formation cover an area of 563,264 km² (36 % of the Amazon State) and were deposited in the Tertiary-Quaternary Period.

A soil survey was conducted in the Oil Province of Urucu River (named Geólogo Pedro de Moura) between the years of 2008 and 2009. This work resulted in the generation of a soil map along with its respective report (Ceddia et al., 2015), which covers an area of 79.665 km² (Figure 1). Throughout the soil survey, 96 soil profiles

were described and sampled by horizon, totaling 483 horizons/samples. Due to the limitations imposed by the native vegetation, the 315 field observations were restricted to the vicinity of access roads and only data from 96 soil profiles were used for this study.

The soils were classified based on the Brazilian Soil Classification System (Embrapa, 1999). The soil-mapping units of the study site as well as the number of profiles and area of occurrence are shown in Table 1.

In each horizon of the 96 soil profiles, disturbed soil samples were taken for the following soil chemical attributes: pH (water), Ca²⁺, Mg²⁺, K⁺, Na⁺, Al³⁺, H⁺, P, CEC, SB (sum of basic cations), V value (base saturation) and Al³⁺ saturation. The soil chemical analysis was performed according to Embrapa (1997). SOC was measured by wet combustion (Walkley and Black, 1932). Soil physical data consisted of particle size measurements, comprising sand (2.00-0.05 mm), silt (0.05-0.002 mm) and clay (< 0.002 mm) by the Pipette method.

Undisturbed soil sampling for D_b was done by the core method using standard steel cylinders of 53 cm³ volume ($h = 42$ mm, $d = 40$ mm). In each of the 96 soil profiles, the steel cylinders were inserted into the center of each soil horizon (perpendicular to the surface). The soil-filled cylinder was carefully removed from the ring holder and the soil extending beyond both cylinder ends was trimmed flush using a sharp knife. Protective plastic covers were used to prevent samples from drying out. The samples were transported to the laboratory and were oven dried (105 °C) until constant weight (24-48 h).

Development of the pedotransfer function

The first step in the PTFs generation was the selection of the model development (training) and validation of the dataset. The 96 soil profiles were randomly split (80/20) into training and validation sets. Thus, the dataset used for training consisted of 77 soil profiles (adding up to 378 horizons), while the dataset used for validation consisted of 19 soil profiles (adding up to 105 horizons). The spatial distribution of the training and validation soil profiles along the study area are presented in Figure 1.

The stepwise multiple regression routine was used for explanatory analysis relating D_b to soils attributes. The Akaike Information Criterion (AIC) with a p -value of 0.05 was used to include or exclude variables. All linear model assumptions were checked (multicollinear-

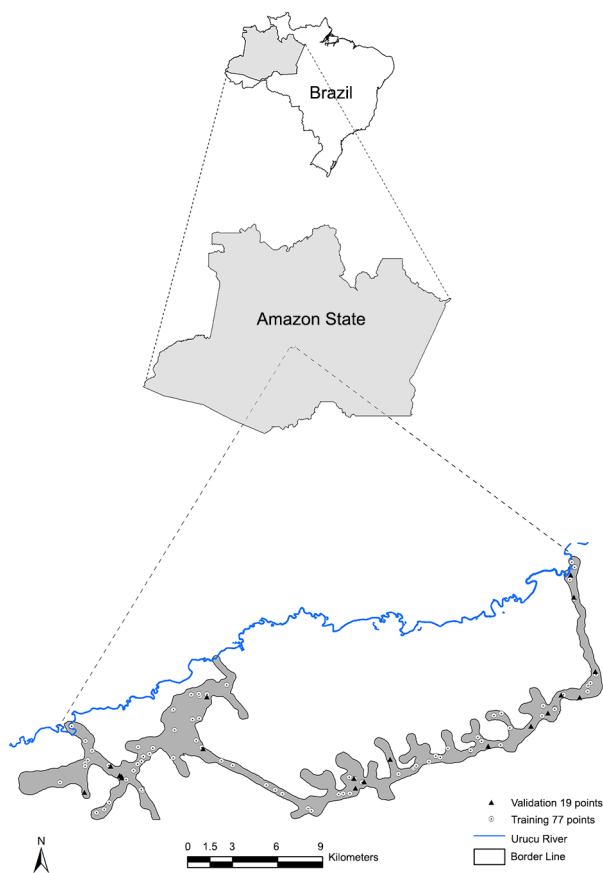


Figure 1 – The location of the study site in the Central Amazon (Brazil) and the spatial distribution of the soil profiles used for training and validation of the pedotransfer function.

Table 1 – Distribution of soil class in the data set.

Order (Embrapa, 1999)	FAO-WRB (FAO, 1998)	Number of soil profiles	Total %
Argissolos	Acrisols, Lixisols	53	55
Cambissolos	Cambisols	39	41
Espodosolo	Spodosol	1	1
Neossolos	Arenosols, Fluvisols	2	2
Planossolo	Planosol	1	1
Total		96	100

ity, homoscedasticity and the normality of regression residues). Multicollinearity was minimized by removing variables with variance inflation factors > 4. To verify the premise of homoscedasticity was performed the Breusch-Pagan test and regression normality residues was performed by the Kolmogorov-Smirnov (KS). The stepwise regression was performed by the R software (R Development Core Team, version 3.1.1).

The explanatory analysis was conducted considering two possibilities: a) construction of a unique regression model to estimate soil bulk density for all soil depths, and b) construction of two regression models, one for surface horizons (A and AB) and another for subsurface horizons (BA, B, BC, C). The final choice of the model to be used was based on the evaluation of the indices AIC, R^2 and standard error (SE) of the stepwise regression. Therefore, the best model presented the highest R^2 and the lowest SE and AIC.

The evaluation of the pedotransfer function performance

This study compares the applicability of the proposed multiple regression model with three existing models in the literature (Bernoux et al., 1998; Tomasella and Hodnett, 1998; Benites et al., 2007), which are presented in Table 2.

We simulated a situation where the database of the study area lacked D_b data. Thus, to estimate the carbon stock, it would be necessary to use a pedotransfer function available in the literature. Commonly, when D_b data are not available in soils of the Amazon region, researchers choose one of these three PTFs, although the criteria to choose a specific PTF is subjective and not clearly explained. Bernoux et al. (1998) and Tomasella and Hodnett (1998) generated the first two PTFs to predict D_b data from properties for soil across the Amazon basin and both used the soil data set generated by the RADAMBRASIL project (RADAMBRASIL, 1978). More recently, Benites et al. (2007) generated a model to predict D_b from readily available soil properties of Brazilian soils found in most biomes. The latter study constructed a database from the Soil Archives of Embrapa Solos in Rio de Janeiro, Brazil.

The predicted values (\hat{y}_i) of D_b , using different PTFs were compared with the 104 observed values (y_i) in the validation dataset. The differences between the 104 predicted and observed values ($\hat{y}_i - y_i$) were used to

calculate the following error measurements: the mean prediction error (MPE), Eq. (1); the root mean squared prediction error (RMSPE), Eq. (2); the mean absolute error (MAE), Eq. (3); and the prediction coefficient of determination (R^2), Eq. (4). The results were also evaluated graphically by the ratio 1:1 of the observed vs predicted values.

$$\text{MPE} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i) \quad (1)$$

$$\text{RMSPE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (2)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (3)$$

$$R^2 = \frac{[\text{COV}(y_i, \hat{y}_i)]}{\text{Var}(y_i) \cdot \text{Var}(\hat{y}_i)} \quad (4)$$

where: \hat{y}_i is observed D_b of i th soil sample; y_i is predicted D_b of i th soil sample, and n is total number of observations.

The MPE (accuracy error) enables the evaluation of an average tendency for overestimation (positive values) or underestimation (negative values). The closer to zero MPE is, the greater the model accuracy. MAE is an accuracy indicator, but it does not reveal the trend to over- or underestimation. This is because it uses an absolute value because of the difference between the observed and predicted data. The RMSPE should be zero, when a perfect fit between the observed and predicted data is achieved.

Spatial variability map of soil organic carbon stock using measured and estimated values of D_b by PTFs

For each of the 96 soil profiles, the calculation of the SOC stock was performed at 0-100 cm depth. The classical way of calculating SOC stock (C mass per area) for a given depth consists of summing C stocks by horizon, determined as a product of D_b , SOC content, and horizon thickness (Eq.5), according to Bernoux et al. (2002):

$$\text{SOC stock} = (\text{SOC} \times D_b \times T) \quad (5)$$

where: SOC stock is soil organic carbon stock (kg C m^{-2}); SOC is soil organic carbon (g kg^{-1}); D_b is the soil bulk density (Mg m^{-3}) and, T is the horizon thickness (m).

Table 2 – Soil bulk density estimate models from the literature (Benites et al., 2007).

Regression models (Region)	N_{sp}	N_h	Data density sp/km^2	Equations
Bernoux et al. (1998) (B.L.A.- 5,020,000 km^2)	690	323	0.00014	$D_b = 1.524 - 0.0038 (\% \text{ clay}) - 0.050 (\% \text{ SOC}) - 0.045 (\text{pH}_{\text{H}_2\text{O}}) + 0,001 (\% \text{ sand})$
Tomasella and Hodnett (1998) (B.L.A.- 5,020,000 km^2)	1162	396	0.00023	$D_b = 1.578 - 0.054 (\% \text{ SOC}) - 0.006 (\% \text{ silt}) - 0.004 (\% \text{ clay})$
Benites et al. (2007) (M.B.B - 8,516,000 km^2)	363	1542	0.00007	$D_b = 1.560 - 0.0005 (\text{g kg}^{-1} \text{ clay}) - 0.010 (\text{g kg}^{-1} \text{ SOC}) + 0.0075 (\text{SB})$

B.L.A. = Brazilian Legal Amazon Region; M.B.B. = Most Brazilian Biomes (Brazilian Territory); N_{sp} = number of soil profiles; N_h = number of soil horizons/layers with D_b data; Data Density (sp/km^2) = Number of soil profiles per square kilometers; D_b = soil bulk density; SOC = soil organic carbon; SB = sum of basic cations.

In the soil survey, the soil profiles were divided into horizons (A, B, and C). In most cases, the calculations concerned two horizons where the first horizon was typically entirely above 100 cm, and the second one crossed this 100 cm depth. When a horizon crossed the 100 cm boundary, only the horizon portion above that depth was used to calculate its SOC stock.

Considering the measured data as well as the four estimates of D_b (model generated in this work and three other published in the literature (Bernoux et al., 1998; Tomasella and Hodnett, 1998; Benites et al., 2007), a total of five dataset of SOC stocks at the layers 0-100 cm depth were generated. For each of the five dataset of SOC stocks, experimental semivariograms were calculated for spatial dependence evaluation and a theoretical model that best represented the data variability was set. The experimental semivariogram, $\gamma(h)$, of n spatial observations $Z(x_i)$, $i = 1, \dots, n$, was calculated using equation 6:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z_{x_i} - Z_{x_i+h}]^2 \quad (6)$$

where: $N(h)$ = the number of observations separated by distance h ; Z_{x_i} = the soil attribute value measured at a specific point (x_i) of the grid; Z_{x_i+h} = the soil attribute value measured at a specific neighbor point apart by distance h ($x_i + h$).

The theoretical model was validated by the Jack-knife tool (self-validation or cross-validation). Thereafter, five spatial variability maps of SOC stocks were generated by the ordinary kriging (OK) method. OK only uses primary data such as SOC stocks measured at sampled locations u to estimate SOC stocks at unsampled locations (Wackernagel, 2003). For the study site, SOC stock is the primary variable $Z_i(u)$, measured at sampled locations u to estimate SOC stocks at unsampled locations ($Z_{OK}^*(u)$). The stationarity of the mean is assumed only within a local neighborhood $W(u)$, centered at the location u being estimated. Here, the mean is deemed to be a constant but unknown value, i.e., $m(u^*) = \text{constant but unknown}$, $\forall u^* \in W(u)$. The OK estimator (Eq. 7) is written as a linear combination of the $n(u)$ data $Z_i(u)$ with a single unbiasedness constraint (Eq. 8), as below:

$$Z_{OK}^*(u) = \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}(u) [Z_i(u)] \quad (7)$$

$$\text{with } \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}^{OK} = 1 \quad (8)$$

The unknown local mean $m(u)$ is filtered from the linear estimator by forcing the kriging weights (λ) to sum to 1 (Eq. 8). The weights λ are chosen so that the estimate $Z_{OK}^*(u)$ is unbiased, and that the estimation variance $\sigma_{OK(u_0)}^2$ (Eq. 9) is less than any other linear combination of the observed values. The minimum variance of $Z_{OK}^*(u)$ is given by:

$$\sigma_{OK(u_0)}^2 = \sum_{i=1}^N \lambda_{\alpha} \gamma(u_i, u_0) + \mu \quad (9)$$

and is obtained when

$$\sum_{i=1}^N \lambda_{\alpha} \gamma(u_i, u_j) + \mu = \gamma(u_i, u_0) \quad (10)$$

where: $\gamma(u_i, u_j)$ = the semivariance of z between the sampling points u_i and u_j ; $\gamma(u_i, u_0)$ = the semivariance of z between the sampling point u_i and the unvisited point u_0 ; Both quantities $\gamma(u_i, u_j)$ and $\gamma(u_i, u_0)$ are obtained from the theoretical model fitted to the experimental semivariogram; μ = the Lagrange multiplier required for the minimization.

The semivariogram calculation, Jack-knife and ordinary kriging procedures were conducted using the software Geoestat (Vieira et al., 1983). The kriged files were exported to software ArcGIS (ESRI, version 9.3), where the spatial variability maps of SOC were made.

The evaluation of the spatial variability error of SOC stock from measured and estimated values of D_b by PTFs

In order to compare the maps generated from different data of D_b , spatial analysis through map algebra (ArcGIS Raster Calculator Function) was performed. The spatial variability map of carbon stocks generated from the measured data (96 SOC stock values using measured values of carbon content and D_b) was considered the most appropriate map (reference map). Then, we calculated how much of each PTF used to estimate the D_b would over- or underestimate the soil carbon stock. This procedure allowed determining the carbon stock residual (up to 1 m) from the use of each PTF. The residuals were calculated subtracting (pixel by pixel) the carbon stock values in the reference map, in relation to the others generated from the application of PTFs. For this study, each pixel has an area of 1 ha (resolution of 100 by 100 m).

Results and Discussion

Descriptive statistics and the correlation between D_b and other soil attributes

Descriptive statistics of soil properties are shown in Table 3. The 483 soil horizons covered a wide range of soil textural classes, although sandy and franco silty were the most predominant classes. We highlight the relatively high silt content (mean value of 339, reaching up to 721 g kg⁻¹), since this granulometric fraction is not commonly higher than 200 g kg⁻¹ in the main classes of Brazilian soils. However, along the Central Amazon Region, the soils are formed over sediments of Içá and Solimões formations (consisted of fine to medium sandstone and siltstone). High values of silt were already reported by Tomasella and Hodnett (1998). The authors found silt values in their dataset that reached up to 800 g kg⁻¹. Considering the chemical attributes, the soils

presented high acidity and aluminum toxicity and low sum of basis (SB). As characteristics of the study site, the mean Al^{3+} content was $4.03 \text{ cmol}_c \text{ dm}^{-3}$ (ranging from 0.25 to $12.00 \text{ cmol}_c \text{ dm}^{-3}$). All chemical properties, except for pH measurements, had a CV > 45 %. The SOC contents ranged from 0.10 to 60.30 g kg^{-1} and a CV of 74 %.

The skewness and kurtosis coefficient could be used to infer about the normal data distribution (symmetric histogram). A zero value for both coefficients means that the attribute presents a normal distribution. Although there is no clear-cut guidelines, most studies consider data to be approximately normal in shape if the skewness and the kurtosis values range from -1.0 to +1.0 (Huck, 2012). Considering this range for skewness and kurtosis, with the exception of H^+ and SOC content, all soil properties follow within a normal distribution.

The mean D_b measured was 1.25 Mg m^{-3} , with minimum and maximum values of 0.49 and 1.67 Mg m^{-3} , respectively. The lowest D_b value is due to the highest amount of soil organic matter (SOM) in well drained soils (complex CXa1-typic *dystrudepets* and PVAa-typic *chapludults*) as opposed to poorly drained soils (consociation PACd-typic *endoaquults*) in the Amazon Forest (Ceddia et al., 2015). The relief is considered the main factor influencing the variability of soil types and their attributes in

Içá Formation, as well as the type and diversity of vegetation and consequently the carbon supply to the soil (RADAMBRASIL, 1978; Ceddia et al., 2015). The coefficient of variation for D_b was 16 %, slightly higher than that previously reported for the Amazon Basin (Moraes et al., 1995; Bernoux et al., 1998). Moraes et al. (1995) observed CV values of 7 % for Alfisols and Ultisols and 13 % for Oxisols, whereas others (Bernoux et al., 1998) have reported a CV of 10 % and 11 % for Oxisols and Alfisols and Ultisols, respectively.

The regression model generated and the comparison with previously published models

The results of the stepwise analysis to select the regression models and respective predictors for D_b for all soil depths, as well as for the surface and sub surface horizons, are presented in Table 4. All models generated showed a normal distribution of the regression residuals (Kolmogorov-Smirnov test). With the exception of the regression model for soil surface layer, the models (all horizons and subsurface) fulfilled the homoscedasticity premise as confirmed by the Breusch-Pagan test (Table 4).

The D_b varies according to soil depth (Harrison and Bock, 1981; Leonaviciute, 2000) however, our results show no improvement in the prediction capacity of the models after doing a separate calibration of superficial horizons from the sub-superficial horizons. Similarly, De Vos et al. (2005) did not find any significant enhancement in the predictive capacity when datasets from top and subsoil layers in forest soils of the Flanders region (Belgium) were separated. Other authors working with different regions in the world, such as Han et al. (2012) in China, Sequeira et al. (2014) and Heuscher et al. (2005) in the United States, also did not find any improvement in the prediction of D_b generating separated regression model for surface and subsurface soil layers. The PTFs published in the literature, which are recommended to be applied to the Amazon region (Bernoux et al., 1998; Tomasella and Hodnett, 1998) and to most Brazilian biomes (Benites et al., 2007), are used for all soil depths. However, more recently the performance of models was evaluated in soils of most Brazilian biomes

Table 3 – Descriptive statistics of the soil properties.

Soil property	Mean	Min.	Max.	SD	CV	Skewness	Kurtosis
D_b (Mg m^{-3})	1.25	0.49	1.67	0.20	16	-0.76	0.44
Sand (g kg^{-1})	393	22	918	173.70	44	0.52	0.002
Silt (g kg^{-1})	339	26	721	122.58	36	0.11	0.04
Clay (g kg^{-1})	266	13	640	131.06	49	0.21	-0.65
pH_{H_2O}	4.58	3.50	5.70	0.39	8	-0.07	-0.29
SB ($\text{cmol}_c \text{ dm}^{-3}$)	1.65	0.30	4.52	0.87	52	0.70	-0.45
Al^{3+} ($\text{cmol}_c \text{ dm}^{-3}$)	4.03	0.25	9.50	1.84	45	0.31	-0.25
H^+ ($\text{cmol}_c \text{ dm}^{-3}$)	3.64	0.05	21.70	3.41	94	2.15	5.35
SOC (g kg^{-1})	7.29	0.10	31.68	5.46	74	1.74	3.59

Min. = minimum; Max. = maximum; SD = standard deviation; CV = coefficient of variation (%); D_b = soil bulk density; SOC = soil organic carbon; SB = sum of basic cations.

Table 4 – Candidate models to predict soil bulk density for all soil depth and dividing the dataset into surface and subsurface horizons.

Models (All soil depth)	Int.	SOC	pH_{H_2O}	SB	Al^{3+}	Clay	Silt	Sand	AIC	SE	R^2	KS (p value)	BP (p value)
1	1.171	-0.0237	0.0622	-0.0230	-0.0124	0.0002			-1593.53	0.1205	0.6672	0.6303	0.2566
VIF	-	1.75	1.50	1.57	2.57	2.87							
Models (Surface)	Int.	SOC	pH_{H_2O}	SB	Silt	Sand	Clay	Al^{3+}	AIC	SE	R^2	KS (p value)	BP (p value)
2	0.609	-0.015	0.145						-546.08	0.1427	0.4745	0.7751	0.0238
VIF	-	1.05	1.05										
Models (Subsurface)	Int.	SOC	Silt	Clay	Al^{3+}	pH_{H_2O}	Sand	SB	AIC	SE	R^2	KS (p value)	BP (p value)
3	1.419	-0.029	0.0002						-1049.72	0.1167	0.2474	0.8385	0.0621
VIF	-	1.00	1.00										

Surface horizons refer to A and AB horizons; Sub-surface horizons refer to BA, B, BC and C horizons; Int. = Intercept; SOC = soil organic carbon; AIC = Akaike Information Criterion; SE = Standard error of the estimation; KS (p value) - values higher than 0.05 mean that the regression residues are normally distributed (KS = Kolmogorov-Smirnov test); BP (p value) - values higher than 0.05 means that the regression residues are homocedastic (BP = Breusch-Pagan test); SB = sum of basic cations; VIF = Variance Inflation Factor; As a reference, when the VIF value of the soil attribute is lower than 4, it means that the respective attribute does not present colinearity with any other predictor variable.

after separation for surface (to 30 cm depth) and sub-surface (below 30 cm) (Benites et al., 2007). The authors did not find any advantage of partitioning the dataset into groups of soil depth and soil orders (Benites et al., 2007). Here, we highlight the model number 1 in Table 4 (Eq. 11) that could be used for all soil depths. This model explained 67 % of the variance and presented the lowest value of AIC and the highest value of R^2 . The SOC was the main predicted variable followed by $\text{pH}_{\text{H}_2\text{O}}$, SB, Al^{+3} and Clay.

$$D_b = 1.171 - 0.0237(\text{SOC}) + 0.0622(\text{pH}) - 0.0230(\text{SB}) - 0.0124(\text{Al}^{+3}) + 0.0002(\text{Clay}) \quad (11)$$

Consistent with the findings of many authors, SOM clearly plays a dominant role (Adams, 1973; De Vos et al., 2005; Heuschek et al., 2005; Han et al., 2012) because of its much lower density than mineral soil particles and its aggregation effect on soil structure. The D_b strongly correlates with SOM content and soil texture (Adams, 1973; Manrique and Jones, 1991). Barros and Fearnside (2015), working with oxisols, also in the Central Amazon region, found that the clay content accounts for about 70 % of the variation in soil bulk density. However, the importance of each variable depends on the study site. For example, De Vos et al. (2005) found that the addition of texture as a predictor has a minor effect on D_b estimation of forest soils. In fact, the authors observed that SOM accounted for 55 to 57 % of the total variation in D_b , whereas soil texture explained only 20 to 26 %.

In the literature, the relationship between D_b and chemical attributes (such as pH, SB and Al^{+3}) is scarcely reported and a direct physical link is not clearly presented. In most cases, these attributes are inserted into the model due to their availability in most datasets, because chemical attributes are determined at low costs (Bernoux et al., 1998). For instance, Bernoux et al. (1998) used pH in water and Benites et al. (2007) used SB to estimate D_b . However, these authors did not provide a physical explanation regarding the relationship between either pH or SB with D_b . In tropical soils, the main clay minerals are kaolinite, goethite, hematite and gibbsite, which are colloids with variable charges depending on the pH. The main clay minerals in the study site are kaolinite and goethite, consequently, as the pH is very low (4.58 on average) and organic colloids are mainly responsible for the CEC of the soil, cation bridges are formed between soil particles. Cation flocculants (Al^{+3} , Fe^{+3} and H^+) promote the approximation of the colloidal particles, which is the first step for the aggregation formation. The next step of aggregate formation is cementation, where SOC has an important role, acting as an agent that cements flocculated particles (Tisdall and Oades, 1982). Finally, soil macro-aggregates are formed, which improves the soil porosity and consequently reduces D_b . In uncultivated soils, like most forest soils, the OM has a dominant effect over D_b (Adams, 1973) and,

naturally, becomes the main predictor variable (De Vos et al., 2005). Considering that the main colloids of the study site in charges that depend on the pH, we hypothesize that Al^{+3} and H^+ availability increases as the pH is lowered. Therefore, colloids flocculate and increase soil porosity and decrease soil bulk density.

Compared to other PTFs, the proposed model (Eq.11) presented the best agreement (Figures 2A, B, C and D). Considering the MPE index, our proposed model presented the lowest bias, once it reached the ideal value (MPE = 0). The model proposed by Tomasella and Hodnett (1998) tended to underestimate D_b (MPE = -0.03 kg dm^{-3}), whereas the models by Benites et al. (2007) and by Bernoux et al. (1998) tended to overestimate D_b (0.11 and 0.15 kg dm^{-3} , respectively). The MAE value ranged between 0.09 and 0.19 kg dm^{-3} , and the highest value was observed in the Bernoux et al. (1998) model, followed by the Tomasella and Hodnett (1998) and the Benites et al. (2007) models. The Bernoux et al. (1998) model presented the highest RMSPE value and our model the lowest (Table 5). Barros and Fearnside (2015) also developed PTFs for Oxisols in the Central Amazon region and compared the performance of their model with those presented by authors reported in this work (Benites et al., 2007; Bernoux et al., 1998; Tomasella and Hodnett, 1998). The authors found that the application of these three PTFs overestimate soil bulk density, which is in agreement with our findings. However, we highlight that the soil bulk density determined by Barros and Fearnside (2015) is significantly lower than what we found in the region of the Urucu River (average and median values of 0.66 and 0.62, respectively). This can explain why even the model developed by Tomasella and Hodnett (1998) also overestimated soil bulk density in the Oxisols evaluated by Barros and Fearnside (2015).

Therefore, the model developed using local data of the study site to predict D_b outperformed the three others, both in terms of accuracy and precision. These results confirm that it is very difficult for one particular PTF to be precise and accurate to predict a soil attribute in a vast territory. This is particularly true in the Amazon region that encompasses different ecosystems within an area of $5,020,000 \text{ km}^2$. The low performance of the other three models is probably a consequence of the low-density data used to generate the regression models (Table 2). These models did not capture adequately the high variability of the Amazon soils. The better performance of the model generated in this study was expected, since local models can capture more effectively the specificities of the relationship among soil attributes. This is also why PTFs should not be generalized to wider regions that do not follow the assumptions underlying the relationship between the target and predicting variables. Considering these aspects, it is important to highlight that the PTF generated in this work should only be used in regions belonging to the Içá Formation and that are covered with forest.

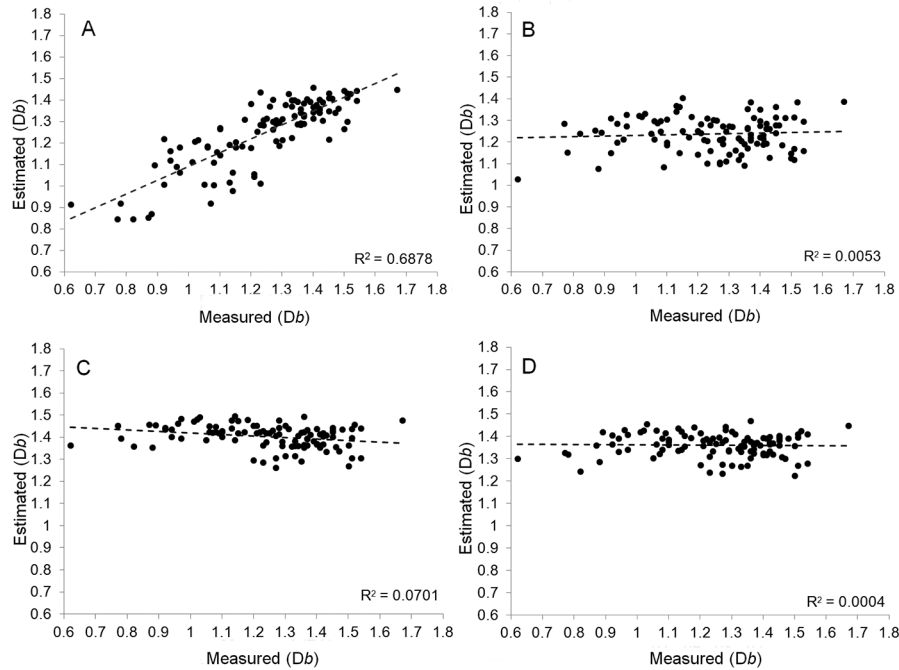


Figure 2 – Plot of predicted vs. observed bulk density (Mg m^{-3}) considering the four pedotransfer functions (PTFs). A) Generated Model for all depth; B) Tomasella and Hodnett (1998); C) Bernoux et al. (1998); D) Benites et al. (2007).

Table 5 – Evaluation indices of proposed and existing regression models for D_b estimation.

Models	MPE	MAE	RMSPE	R^2
Generated Model (All depth)	0.00	0.09	0.11	0.6878
Benites et al. (2007)	0.11	0.17	0.24	0.0004
Tomasella and Hodnett (1998)	-0.03	0.19	0.22	0.0053
Bernoux et al. (1998)	0.15	0.19	0.27	0.070
Reference values	0	0	0	1

MPE = Mean Prediction Error; MAE = Mean Absolute Error; RMSPE = Root Mean Squared Prediction Error.

Table 6 – Models and their parameters fitted for SOC stock's semivariograms.

Models	C_0	C_1	Sill (m)	Range (m)
Measured data of D_b and SOC	2.87	1.87	4.74	3981
Generated Model (All depth)	2.95	1.10	4.05	2800
Benites et al. (2007)	2.38	3.01	5.39	2180
Tomasella and Hodnett (1998)	3.30	1.58	4.88	8000
Bernoux et al. (1998)	3.09	2.98	6.07	2400

C_0 = nugget effect; C_1 = contribution; Sill = $C_0 + C_1$; SOC = soil organic carbon.

Applying the measured and predicted values of D_b to model the spatial variability of SOC stock

As the measurement of D_b is essential to predict SOC stock, the impact of using PTFs in the spatial variability map of SOC stock was evaluated. Thus, at this stage, we compared the semivariogram parameters and maps of SOC stock performed from the measured data with those generated when applying the four PTFs. The application of geostatistical techniques requires that the spatial dependence between the observations of SOC stock be proved. For this, the experimental semivariogram should be calculated in order to set a theoretical model that best represents data variability (Figures 3A, B, C and D).

The SOC stock at 0-100 cm exhibited spatial dependence. In fact, for all experimental semivariograms, both the lag distance (h) and semivariance $\gamma(h)$ increased until they reached an approximately constant value called sill variance (known as the priori variance of the random variable (Table 6)). The same theoretical model

(spherical) was fitted for all the experimental semivariograms, which differ from each other in the nugget parameters (C_0), structural variance (C_1) and range (a). The semivariogram of reference (Figure 3A) is the one calculated from the measured data of D_b and soil organic carbon, as these are the best data we have to evaluate the SOC stock variability. In general, the parameters of the semivariogram using the pedotransfer generated in this study were close to those of the reference semivariogram (Figure 3B). The experimental semivariogram calculated using D_b estimated from the PTF proposed by Tomasella and Hodnett (1998) presented the highest range (Figure 3D). The range is the lag distance at which the semivariogram reach its sill. This is the spatial dependence and beyond it, the variance bears no relation to the separation distance (Webster and Oliver, 1990). On the other hand, the experimental semivariogram calculated using D_b estimated using the PTF proposed by Benites et al. (2007) and Bernoux et al. (1998), presented

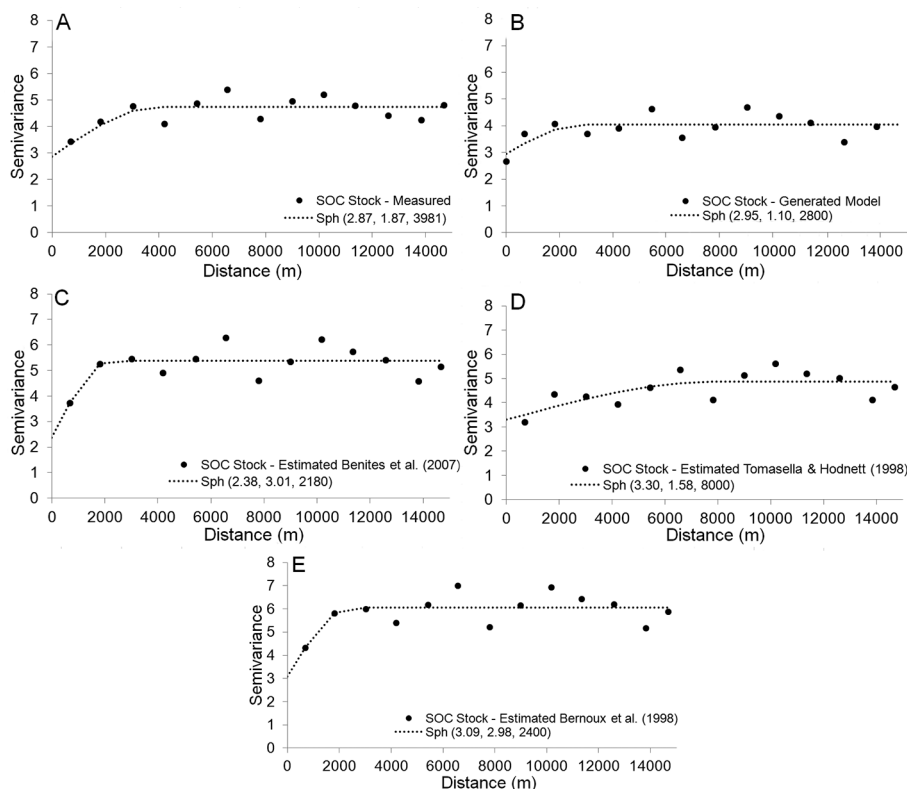


Figure 3 – Carbon stock semivariograms using measured and estimated data of D_b . A) Semivariogram generated with the measured data of D_b ; B) Semivariogram generated applying the D_b estimated using the PTF proposed in this work; C) Semivariogram generated applying the D_b estimated using the PTF proposed by Benites et al. (2007); D) Semivariogram generated applying the D_b estimated using the PTF proposed by Tomasella and Hodnett (1998); E) Semivariogram generated applying the D_b estimated using the PTF proposed by Bernoux et al. (1998). Sph = Spherical model (nugget effect, contribution and range values); PTFs = Pedotransfer functions; SOC = soil organic carbon.

the lowest ranges (spatial dependence of 2180 and 2400 m, respectively) (Figures 3C and E).

The semivariogram with the highest nugget effect value (3.30) was the one generated from the application of the PTF proposed by Tomasella and Hodnett (1998), while the lowest nugget effect value (2.38) was the one developed by Benites et al. (2007). This implies that the SOC stock has a higher initial variability when using the model proposed by Tomasella and Hodnett (1998).

The structured variability using ordinary kriging allowed the generation of the spatial variability map of SOC stocks up to 100 cm of soil depth (Figures 4A, B, C, D and E). The SOC stock map generated with measured values of D_b and SOC (reference map) ranged from 4.89 up to 9.93 kg C m⁻² (Figure 4A). These values are similar to those found in the literature for the Amazon region. Ceddia et al. (2015), using different results published in the literature, compared the average values of SOC stock up to the 100 cm soil depth for the same region and found that the estimative of SOC stock ranged from 7.32 to up to 9.01 kg cm⁻².

The SOC stock variability maps performed with D_b estimated by our PTF and the PTFs presented by Tomasella and Hodnett (1998) (Figures 4B and D, respec-

tively) had amplitude values closer to those in the reference map. The SOC stock variability maps using the D_b estimated by PTFs proposed by Benites et al. (2007) and Bernoux et al. (1998) tended to overestimate these values (Figures 4C and E). These maps showed the highest upper limit values of SOC stock (13.96 kg cm⁻² and 12.81 kg cm⁻², respectively).

Bernoux et al. (2002) argue about the uncertainty sources in estimating SOC stock. The first source related to different database information used in the estimates. The second source referred to the error associated with the estimation of D_b by PTFs. However, according to the authors, the most important source of uncertainty originates from the SOC analytical methods. Therefore, they reaffirm the need for a more complete documentation (metadata) about the database used to generate PTFs. According to the authors, in addition to the statistical data information, the database must contain which methods were used to obtain the attributes used as predictors. The authors refer to (Garten and Wulfschleger, 1999) as traditional method in soil science and claim that this method is not completely accurate and is considerate the main source of uncertainty in estimating SOC. We also observed in this study that variations in D_b through ap-

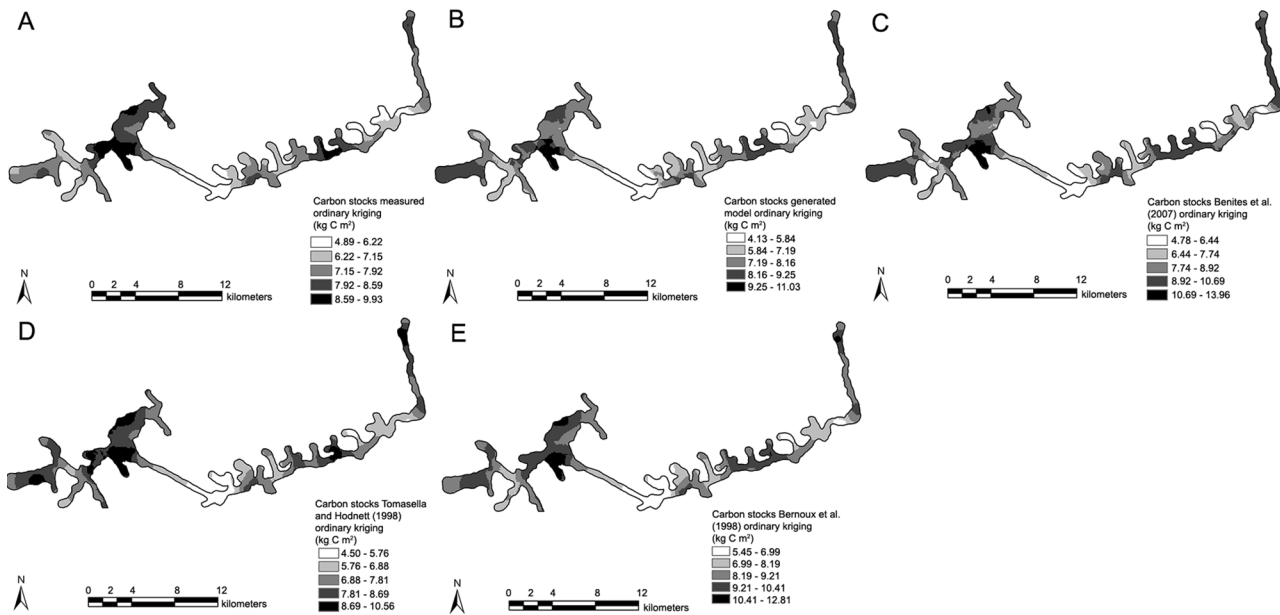


Figure 4 – Spatial variability maps of SOC stock using ordinary kriging. A) Map generated with the measured data of D_b ; B) Map based on the D_b estimated using the PTF generated in this work; C) Map based on the D_b estimated using the PTF generated by Benites et al. (2007); D) Map based on the D_b estimated using the PTF generated by Tomasella and Hodnett (1998); E) Map based on the D_b estimated using the PTF generated by Bernoux et al. (1998). SOC = Soil Organic Carbon; PTFs = Pedotransfer functions.

plications of PTFs not suitable to an environment, other than the one in which they were developed, could cause significant errors in SOC stock measurements.

The residuals caused by applying different PTFs on the spatial variability maps of SOC stock

The spatial variability maps of the residuals as well as the statistics of the residuals are shown in Figures 5A, B, C and D and Table 7, respectively. The residual classes shown in Figures 5A and C confirmed that the PTFs generated in this study and those developed by Tomasella and Hodnett (1998) presented the lowest residual and tend to underestimate even more the SOC stock, mainly when compared to the maps generated by Benites et al. (2007) and Bernoux et al. (1998). These results are in accordance with the values of the MPE index previously discussed in item 3.2. We highlight the overestimation of SOC stock when applying the PTFs generated by Benites et al. (2007) and Bernoux et al. (1998). Using these PTFs, the overestimation of SOC stock could reach up to 4.20 and 2.95 kg cm⁻², respectively (Figures 5B and D). The mean value of the SOC stock applying the PTF function generated in this study is practically the same of that in the reference map (7.49 Mg C ha⁻¹). On the other hand, applying the PTFs developed by Benites et al. (2007) and Bernoux et al. (1998), resulted in mean values of SOC stock of 8.56 and 8.73 Mg C ha⁻¹, respectively. When applying these PTFs, the mean residual value of SOC stock in relation to the reference map is 1.06 and 1.23 Mg C ha⁻¹, a mean overestimation of 15 and 17 %, re-

Table 7 – Statistics of the SOC stock maps and the residual caused by the application of different PTFs.

Models	Minimum	Maximum	Mean	SD
Kriged maps of SOC stocks (Mg C ha ⁻¹)				
Measured data (RM)	4.89	9.93	7.49	0.95
Generated Model (All depth)	4.13	11.03	7.53	1.18
Benites et al. (2007)	4.78	13.96	8.56	1.46
Tomasella and Hodnett (1998)	4.50	10.56	7.56	1.11
Bernoux et al. (1998)	5.45	12.81	8.73	1.17
Residual maps of SOC stocks (Mg C ha ⁻¹)				
MG (All depth)-RM	-1.78	1.50	0.03	0.58
Benites et al. (2007)-RM	-1.13	4.20	1.06	0.83
Tomasella and Hodnett (1998)-RM	-1.12	1.84	0.06	0.49
Bernoux et al. (1998)-RM	-0.09	2.95	1.23	0.52

RM = Reference map; SD = Standard Deviation; SOC = soil organic carbon.

spectively (Table 7). Considering the residuals found in this work, the error in the estimation of SOC stocks, using different PTFs to estimate D_b could be higher than 10 %, as previously reported by Bernoux et al. (1998).

The results here presented can also give support to answer a significant and generic problem with the estimation of changes in terrestrial biospheric carbon, which is the smallest detectable change. The smallest difference in SOC stock that could be detected after 5 years under an herbaceous bioenergy crop was about 1 Mg C ha⁻¹ in the southeastern United States (Garten and Wullschlegler, 1999). Considering the average residual

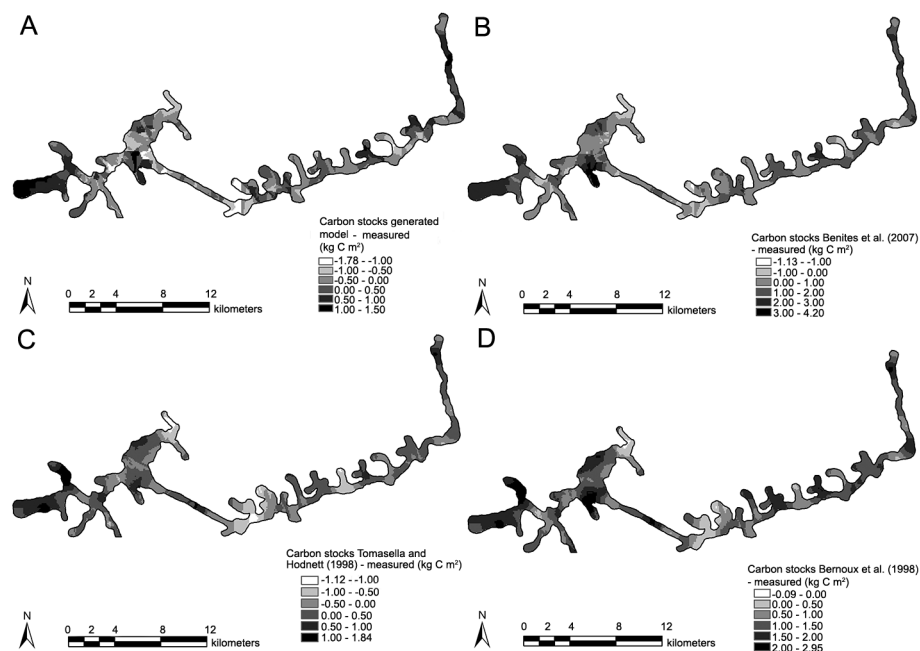


Figure 5 – Spatial variability maps of the SOC stock residual. A) Residual based on the D_b estimated by the PTF generated in this work; B) Residual based on the D_b estimated using the PTF generated by Benites et al. (2007); C) Residual based on the D_b estimated of the PTF generated using Tomasella and Hodnett (1998); D) Residual based on the D_b estimated using the PTF generated by Bernoux et al. (1998). SOC = Soil Organic Carbon; PTFs = Pedotransfer functions.

of SOC stock caused by applying the PTFs developed by Benites et al. (2007) and Bernoux et al. (1998) in the study site, it is possible to conclude that the simple application of a PTF to estimate D_b could introduce an error enough to offset the smallest detectable change of the SOC stock estimate.

Conclusions

A linear regression model was generated to estimate D_b for soils of the Central Amazon region, Brazil. The developed PTF used soil attributes easily found in soil survey reports such as SOC, pH in water, sum of basic cations, Al^{+3} , and clay content as predictor of D_b . The PTF generated should only be used in regions belonging to the Içá Geological Formation and covered with forest.

Compared to the three most used PTFs to estimate D_b in Amazon soils (Bernoux et al., 1998; Tomasella and Hodnett, 1998; Benites et al., 2007), our PTF outperformed the other models, showing not only better accuracy, but also the lowest bias.

We also quantified the effect of using PTFs to estimate the D_b on spatial variability of SOC stock. The results showed that the uncertainty caused by the estimative of D_b using PTFs could be much higher than the values reported in the literature. The PTFs of Benites et al. (2007) and Bernoux et al. (1998) caused an overestimation of 1.06 and 1.23 Mg C ha⁻¹ in the SOC stock, which represented 15 % and 17 %, respectively. This

means that the simple application of a PTF to estimate D_b could introduce an error large enough to skew the significant difference in soil carbon stock change.

Acknowledgements

The authors acknowledge Petrobras for the financial support to perform this study (Contract Petrobras/UFRRJ (Federal Rural University of Rio de Janeiro) / FAPUR (Foundation for Scientific and Technological Research Support of UFRRJ), N° 0050.0036944.07.2). The authors thank the Rio de Janeiro State Foundation for Research Support (project number E-26/100.401/2013).

References

- Adams, W.A. 1973. The effect of organic matter on the bulk and true densities of some uncultivated podzolic soils. *Journal of Soil Science* 24: 10-17.
- Barros, H.S.; Fearnside, P.M. 2015. Pedo-transfer functions for estimating soil bulk density in Central Amazonia. *Revista Brasileira de Ciência do Solo* 39: 397-407.
- Benites, V.M.; Machado, P.L.O.A.; Fidalgo, E.C.C.; Coelho, M.R.; Madari, B.E. 2007. Pedotransfer functions for estimating soil bulk density from existing soil survey reports in Brazil. *Geoderma* 139: 90-97.
- Bernoux, M.; Arrouays, D.; Cerri, C.; Volkoff, B.; Jolivet, C. 1998. Bulk densities of Brazilian Amazon soils related to other soil properties. *Soil Science Society of America Journal* 62: 743-749.

- Bernoux, M.; Carvalho, M.C.S.; Volkoff, B.; Cerri C.C. 2002. Brazil's soil carbon stocks. *Soil Science Society of America Journal* 66: 888-896.
- Bouma J. 1989. Using soil survey data for quantitative land evaluation. *Advances in Soil Science* 9: 177-213.
- Ceddia, M.B.; Vilella A.L.O.; Pinheiro, E.F.M.; Wendroth, O. 2015. Spatial variability of soil carbon stock in the Urucu river basin, Central Amazon-Brazil. *Science of the Total Environment* 526: 58-69.
- Deng, H.; Ye, M.; Schaap, M.G.; Khaleel, R. 2009. Quantification of uncertainty in pedotransfer function-based parameter estimation for unsaturated flow modeling. *Water Resources Research* 45: W04409. DOI:10.1029/2008WR007477.
- De Vos, B.; Meirvenn, M.V.; Quatae, P.; Decker, J.; Muys, B. 2005. Predictive quality of pedotransfer functions for estimating bulk density of forest soils. *Soil Science Society of America Journal* 69: 500-510.
- Empresa Brasileira de Pesquisa Agropecuária [EMBRAPA]. 1997. *Manual of Methods of Soil Analysis = Manual de Métodos de Análises de Solo*. 2ed. Embrapa-CNPS, Rio de Janeiro, RJ, Brazil (in Portuguese).
- Empresa Brasileira de Pesquisa Agropecuária [EMBRAPA]. 1999. *Brazilian System of Soil Classification = Sistema Brasileiro de Classificação de Solos*. 2ed. Embrapa, Brasília, DF, Brazil (in Portuguese).
- Garten, C.T.; Wullschleger, S.D. 1999. Soil carbon inventories under a bioenergy crop (Switchgrass): measurement limitations. *Journal of Environmental Quality* 28: 1359-1365.
- Han, G.Z.; Zhang, G.L.; Gong, Z.T.; Wang, G.F. 2012. Pedotransfer functions for estimating soil bulk density in China. *Soil Science* 177: 158-164.
- Harrison, A.F.; Bockock, K.L. 1981. Estimation of soil bulk-density from loss-on-ignition values. *Journal of Applied Ecology* 8: 919-927.
- Heuscher, S.A.; Brandt, C.C.; Jardine, M.P. 2005. Using soil physical and chemical properties to estimate bulk density. *Soil Science Society of America Journal* 106: 52-62.
- Howard, P.J.A.; Lovel, P.J.; Bradley, R.I.; Dry, F.T.; Howard, D.M. 1995. The carbon content of soil and its geographical distribution in Great Britain. *Soil Use Management* 11: 9-15.
- Huck, S.W. 2012. *Reading Statistics and Research*. 6ed. Pearson, Boston, MA, EUA.
- Leonaviciute, N. 2000. Predicting soil bulk and particle densities by pedotransfer functions from existing soil data in Lithuania. *Geografijos Metraštis* 33: 317-330.
- Manrique, L.A.; Jones, C.A. 1991. Bulk density of soils in relation to soil physical and chemical properties. *Soil Science Society of America Journal* 55: 476-481.
- Moraes, J.L.; Cerri, C.C.; Melillo, J.M.; Kicklighter, D.; Neill, C.; Skole, D.L.; Steudler, P.A. 1995. Soil carbon stocks of the Brazilian Amazon basin. *Soil Science Society of America Journal* 59: 244-247.
- RADAMBRASIL Project. 1978. SB.20 Purus: Geology, Geomorphology, Pedology, Vegetation and Potential Land Use = SB.20 Purus: Geologia, Geomorfologia, Pedologia, Vegetação e Uso Potencial da Terra. Ministério das Minas e Energia, Rio de Janeiro, RJ, Brazil. p. 566. (in Portuguese).
- Sequeira, C.H.; Wills, S.A.; Seybold, C.A.; West, L.T. 2014. Predicting soil bulk density for incomplete databases. *Geoderma* 213: 64-73.
- Tomasella, J.; Hodnett, M.G. 1998. Estimating soil water retention characteristics from limited data in Brazilian Amazonia. *Soil Science* 163: 190-202.
- Vieira, S.R.; Hatfield, J.L.; Nielsen, D.R.; Biggar, J.M. 1983. Geostatistical theory and application to variability of some agronomical properties. *Hilgardia* 51: 1-75.
- Wackernagel, H. 2003. *Multivariate Geostatistics: An Introduction with Applications*. 3ed. Springer, Berlin, Germany.
- Walkley, A.; Black, I.A. 1932. An examination of the Degtjareff method for determining soil organic matter and proposed modification of the chromic acid titulation method. *Soil Science* 37: 29-38.
- Webster, R.; Oliver, M.A. 1990. *Statistical Methods in Soil and Land Resource Survey*. Oxford University Press, Oxford, UK.