https://doi.org/10.1590/1678-992X-2022-0255

ISSN 1678-992X

SCIENTIA AGRICOLA

Research article

Hierarchical pedotransfer functions for predicting bulk density in Brazilian soils

Aline Mari Huf dos Reis[1]*, Wenceslau Geraldes Teixeira[1], Ademir Fontana[1], Alexandre Hugo Cezar Barros[1], Daniel de Castro Victoria[2], Gustavo Mattos Vasques[1], Alessandro Samuel-Rosa[3], Marta Vasconcelos Ottoni[4], José Eduardo Boffino de Almeida Monteiro[2]

[1]Embrapa Solos, R. Jardim Botânico, 1024 – 22460-000 – Rio de Janeiro, RJ – Brasil.
[2]Embrapa Agricultura Digital, Av. André Tosello, 209 – 13083-886 – Campinas, SP – Brasil.
[3]Universidade Tecnológica Federal do Paraná, Prolongamento da Rua Cerejeira, s/n – 85892-000 – Santa Helena, PR – Brasil.
[4]Serviço Geológico do Brasil – SGB/CPRM, Av. Pasteur, 404 – 22290-255 – Rio de Janeiro, RJ – Brasil.
*Corresponding author <huf.aline@gmail.com>

Edited by: Gabriel Ramatis Pugliese Andrade

Received November 22, 2022
Accepted May 21, 2023

ABSTRACT: Bulk density (BD) is a soil physical property used as a soil quality indicator and variations in this measurement influence soil water content and carbon stock estimates. This study aims to compile a database of samples of bulk density, textural fractions, and organic carbon values, as well as evaluate the accuracy of published pedotransfer functions (PTF) that predict bulk density, and propose a hierarchical PTF to predict the bulk density of Brazilian Soils. The performance of eleven PTFs and the newly proposed PTFs were evaluated and compared using the root mean square error (RMSE) and coefficient of determination (R^2) based on a testing soil database collected from the literature. We noticed a slight improvement in accuracy when organic carbon and coarse and fine sand fractions were included as predictors alongside silt and clay. The best results with existing PTFs were obtained by PTF-A in Tomasella and Hodnett (1998) (RMSE = 0.20 g cm^{-3}) and PTF-F in Benites et al. (2007) (RMSE = 0.17 g cm^{-3}). Our proposed PTFs use textural fractions and organic carbon as predictors in a hierarchical form. The proposed PTF-4, which uses fine sand, coarse sand, clay, and organic carbon, presented the lowest value for RMSE (0.14 g cm^{-3}) for BD prediction.

Keywords: transfer functions, data compilation, textural fractions, organic carbon content

Introduction

Soil bulk density (BD), which is the ratio of the mass of an oven-dry soil sample to its volume (solids plus pores) (Hillel, 1998), is an essential property for assessing the sustainability of soil management practices in any region (Botula et al., 2015; Palladino et al., 2022). It is an efficient indicator of soil structure, reflecting compaction and its effects on soil-water-plant-atmosphere relationships (De Vos et al., 2005; Assouline, 2006). It influences soil solution fluxes, root growth and density, and seed germination. Soil bulk density is also a key parameter in the determination of soil organic carbon (SOC) and the stocks of other elements (Lettens et al., 2005; Don et al., 2011; Inagaki et al., 2017; Chen et al., 2018), and is commonly used as a predictor variable for certain physical hydraulic PTFs (Karup et al., 2017; Gunarathna et al., 2019; Silva et al., 2020).

Field sampling and direct measurements of BD can be expensive, labor-intensive, and time-consuming (Kaur et al., 2002; Abdelbaki, 2016; Nasta et al., 2020). As a result, BD is only sometimes determined in soil surveys and routine soil laboratories (De Vos et al., 2005; Don et al., 2011). Corroborating Minasny and Hartemink (2011), it is not feasible to measure all soil physical and chemical properties continuously, particularly in areas with rock fragments and/or woody debris (Nanko et al., 2014; Sevastas et al., 2018). Therefore, it is necessary to use robust systems to estimate the soil properties of a given location.

Pedotransfer functions (PTFs) offer a viable alternative to obtaining certain soil properties from previously known information (Minasny et al., 1999; Saxton and Rawls, 2006; Barros et al., 2013; Ottoni et al., 2019). Several PTFs have been developed to predict BD from texture, organic carbon (OC), pH, sum of exchangeable cations (Alexander, 1980; Manrique and Jones, 1991; Kaur et al., 2002; Souza et al., 2016) or even other variables such as slope, depth, soil type, and land use (Palladino et al., 2022). However, only some PTFs have been developed using Brazilian soil data (Tomasella and Hodnett, 1998; Bernoux et al., 1998; Benites et al., 2007).

We aimed to propose a hierarchical system of PTFs to predict BD based on a large database of Brazilian soils and to compare its performance with existing BD-PTFs from the literature using a testing database of Brazilian soil BD, regardless of genesis, type of land use and cover, and management.

Materials and Methods

Data selection and description

Different Brazilian soil databases were consulted to extract information on Bulk Density Data (BD), such as the Hydrophysical Database for Brazilian Soils (HYBRAS) (Ottoni et al., 2018), the Brazilian Soil Data Repository (SoilData/FEBR) (Samuel-Rosa et al., 2020), and other private Brazilian soil datasets provided by a number of researchers. The 3,050 observed BD data, including all information on sand, silt, and clay

contents (Table 1), were selected. This dataset covers all the 12 textural classes according to the United States Department of Agriculture (USDA) classification (Soil Survey Staff, 2014) (Figure 1). This soil database also contains information on coarse and fine sand data for several soils samples (1,081 samples) in addition to OC (2,827). Table 1 shows the summary statistics of the soil variables from the soil database of this study. The 990 samples have information on BD, clay, silt, coarse sand, fine sand, and OC data available.

The method used for quantifying organic carbon was oxidation with 0.0667 mol L^{-1} $K_2Cr_2O_7$ and titration with 0.1 mol L^{-1} $Fe(NH_4)_2(SO_4)_2.6H_2O$, as described by Fontana and Campos (2017). Particle-size analysis was performed by dispersing the soil with NaOH solution. The total of sand fractions was obtained by sieving, and clay was determined by pipette and hydrometer. The difference between sand and clay determined silt. The total sand fraction was further divided into coarse (2.00-0.210 mm) and fine (0.210-0.053 mm) fractions using the method described by Donagemma et al. (2017).

Table 1 – Ranges of bulk density (BD), soil texture (according to the USDA classification), and organic carbon (OC) for the assessed Brazilian soil databases.

Summary Statistics	BD	Clay	Silt	Sand	Fine sand	Coarse sand	OC
	g cm^{-3}	%					
Minimum	0.63	0.0	0.0	0.0	0.0	0.0	0.0
Median	1.39	25.0	17.0	45.6	19.0	14.9	0.6
Mean	1.38	28.8	24.3	46.9	23.0	21.4	1.0
Maximum	1.96	96.0	88.2	98.8	97.2	97	9.8
Standard deviation	0.2	18.8	20.6	29.2	20.3	22.9	1.0
Coefficient of variation (%)	14.5	65.3	84.7	62.3	88.3	107.0	100
Number of samples	3,050	3,050	3,050	3,050	1,081	1,081	2,827

BD = bulk density; OC = organic carbon.

Evaluating PTFs

We tested twenty PTFs from the literature that use different variables to estimate BD. Some of these PTFs use variables such as the sum of bases, pH value, and cation values, in addition to the more commonly used ones such as texture and OC data. However, we found these PTFs ineffective for all samples as some need more detailed information. As a result, we selected only eleven already-published PTFs for tropical and temperate soils (Tomasella and Hodnett, 1998; Bernoux et al., 1998; Kaur et al., 2002; Benites et al., 2007; Ruehlmann and Körschens, 2009; Hollis et al., 2012; Al-Qinna and Jaber, 2013; Botula et al., 2015; Abdelbaki, 2016) with the best performance (lowest RMSEs) among the twenty models tested. These selected PTFs are presented in Table 2.

The PTFs in the literature provide ranges for each input variable used in the model. For instance, the PTF of Tomasella and Hodnett (1998) ranges from 0-100 % for clay and 0-71 % for silt (PTF-A). The PTFs of Bernoux et al. (1998) have ranges of 3.9-90.75 % for clay and 0.04-12.16 % for OC (PTF-B and C). The PTF of Kaur et al. (2002) presents ranges of 0-56 % for silt, 0-48.1 % for clay, and 0.07-2.32 % for OC (PTF-D). The PTFs of Benites et al. (2007) show ranges of 0-960 g kg^{-1} for clay, 0.3-206 g kg^{-1} for OC (PTF-E and F). The PTF of Ruehlmann and Körschens (2009) presents ranges of 2.7-574.2 g kg^{-1} of OC (PTF-G). The PTF of Hollis et al. (2012) has ranges of 0.5-17.4 % for OC, 0-88 % for clay, and 0-100 % for sand (PTF-H). The PTF of Al-Qinna and Jaber (2013) ranges from 0.1-4.3 % for OC, 20-86.2 % for sand (PTF-I). The PTF of Botula et al. (2015) presents ranges of 1-72 % for clay, 4-90 % for sand, and 0.09-5.36 % for OC (PTF-J). The PTF of Abdelbaki (2016) uses only OC with a range of 0-58 % (PTF-K). Therefore, for each BD-PTF estimate, the range limits of the variable predictors were considered.

Table 2 – Pedotransfer functions (PTF) to prediction of bulk density (BD).

PTF name	Author(s)	PTF	Local
PTF-A	Tomasella and Hodnett (1998)	$BD = 1.578 - 0.054(OC) - 0.006(Si) - 0.004(Cl)$	Brazil
PTF-B	Bernoux et al. (1998)	$BD = 1.352 - 0.0045(Cl)$	Brazil
PTF-C	Bernoux et al. (1998)	$BD = 1.398 - 0.0047(Cl) - 0.042(OC)$	Brazil
PTF-D	Kaur et al. (2002)	$BD = exp\{0.313 - 0.191(OC) + 0.02102(Cl) - 0.000476(Cl^2) - 0.00432(Si)\}$	India
PTF-E	Benites et al. (2007)	$BD = 1.5224 - 0.0005(Cl)$	Brazil
PTF-F	Benites et al. (2007)	$BD = 1.5688 - 0.0005(Cl) - 0.009(OC)$	Brazil
PTF-G	Ruehmann and Körschens (2009)	$BD = (2.684 - 140.943 \times 0.006) \times exp(-0.006OC)$	Germany
PTF-H	Hollis et al. (2012)	$BD = 0.69794 + 0.750636exp(-0.230550OC) + 0.0008687(Sa) - 0.0005164(Cl)$	Europe (various)
PTF-I	Al-Qinna and Jaber (2013)	$BD = 1.228 - 0.155 \times log(OC) + 0.008(Sa)$	Jordan
PTF-J	Botula et al. (2015)	$BD = 1.64581 - 0.00362(Cl) - 0.0016(Sa) - 0.0158(OC)$	Congo
PTF-K	Abdelbaki (2016)	$BD = 1.448exp(-0.03(OC))$	United States of America (various)

BD = bulk density (g cm^{-3}); *Cl* = clay content; *Si* = silt content; *Sa* = sand content; OC = organic carbon. The PTFs of Tomasella and Hodnett (1998), Bernoux et al. (1998), Kaur et al. (2002), Hollis et al. (2012), Al-Qinna and Jaber (2013), Botula et al. (2015), and Abdelbaki (2016) used percentual information about texture and OC. The PTFs of Benites et al. (2006) and Ruehlmann and Körschens (2009) used mass value of texture and OC (g kg^{-1}).

A hierarchical PTFs system for BD predictions

A hierarchical system was tested using textural properties (fine sand, coarse sand, total sand, silt, and clay) and OC as predictors. Four models were tested to estimate soil bulk density, as shown in Table 3. The simplest model, Function 1, includes only the three textural fractions (sand, silt, and clay) as predictors. Function 4 contains all the soil variables as inputs, excluding total sand. Function 2 includes OC in addition to the three textural fractions, and Function 3 includes coarse sand (0.2-2 mm), fine sand (0.05-0.2 mm), silt and clay.

The *Caret* package (Kuhn, 2008) in R software (R Core Team, 2022) was used to fit the PTFs, evaluate goodness of fit, and validate the selected functions. Linear models (lmStepAIC) were used to fit the equations, with significant predictors selected automatically using the smallest Akaike Interaction Criteria (AIC). We also evaluated RMSE and R^2 for the selected functions. To assess the prediction accuracy of the chosen model, we used 10-fold cross-validation (Souza et al., 2016; Haddad et al., 2018; Palladino et al., 2022), reporting the average values of RMSE and R^2 over ten runs. The dataset was split into ten folds in each run, with 90 % of the samples used for calibration and the remaining 10 % for validation.

Model assessment

Evaluation of PTF predictive performances was based on the root mean square error (RMSE) Eq. (1) and adjusted coefficient of determination (R^2_{adj}) Eq. (2) of the function according to the following equations:

$$RMSE = \sqrt{\left[\left(1/N\right)\sum_{i=1}^{N}\left(BD_{pred,i} - BD_{obs,i}\right)^2\right]} \quad (1)$$

$$R^2_{adj} = 1 - \left(1 - \frac{\sum_{i=1}^{N}\left(BD_{pred,i} - BD_{obs,i}\right)^2}{\sum_{i=1}^{N}\left(BD_{obs,i} - \overline{BD}\right)^2}\right)\frac{(N-1)}{(N-p)} \quad (2)$$

in which $BD_{obs,i}$ is the i-*th* observed value, $BD_{pred,I}$ the i-*th* predicted value, N the number of observations, and p the number of predictors. The closer to zero the RMSE, the more accurate the predictions. The R^2 identifies values between 0.0 and 1.0, where a value of 1.0 indicates a perfect fit.

Table 3 – Hierarchical structure to define predicted pedotransfer functions (PTFs).

Functions	Number of predictors	Variables used in the hierarchical models	Samples for each set
PTF-1	3	Total sand + silt + clay	3,050
PTF-2	4	Total sand + silt + clay + organic carbon	2,827
PTF-3	4	Fine sand + coarse sand + silt + clay	1,081
PTF-4	5	Fine sand + coarse sand + silt + clay + organic carbon	990

The results obtained by the available PTFs and the proposed hierarchical functions are plotted against the observed values in a 1:1 line (identity line) for comparison. The accuracy of the linear models was assessed using the scatter plots of residuals (observed minus predicted) versus the observed values of BD.

Results

Our database contains soil samples from all textural classes, as shown in Figure 1. The clayey texture is the most represented, with 643 samples, while silty texture has the smallest number of samples, only six. The lowest BD was observed in clay and silt clay loam classes (0.63 g cm^{-3}), whereas the highest BD value was observed in a sample from the sandy clay class (1.96 g cm^{-3}) (Figure 2).

The soil classes were classified according to the World Reference Base for Soil Resources (WRB) classification (FAO, 2015), which includes twelve classes. The Acrisol class (796 samples) and Ferralsol class (671 samples) were the most predominant in our dataset, reflecting their common occurrence in Brazil. On the other hand, the Luvisol class was the least represented, with only four samples in our database (Figure 3).

Not all 3,050 samples provided fine sand, coarse sand, and OC information. Therefore, the number of observed data varied among the PTFs fitted. PTF-1 used 3,050 samples, PTF-2 2,827 samples, PTF-3 1,081 samples and PTF-4 (which included coarse and fine sand, silt, clay, and OC) was available for 990 observed BD data. This subset of data was used to evaluate the hierarchical system, where BD was predicted using PTF-1, 2, and 3, which had fewer variable predictors.

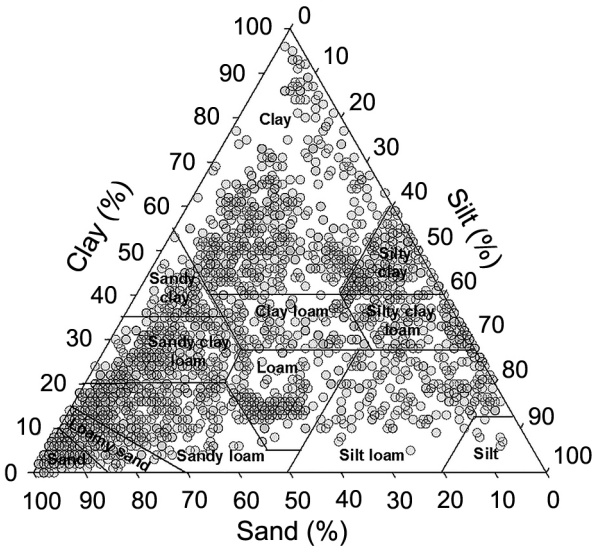

Figure 1 – Distribution of the 3,050 points in the textural triangle according to the United States Department of Agriculture (USDA) classification.

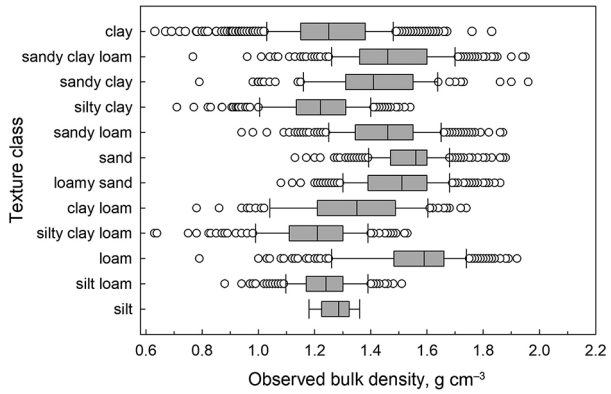

Figure 2 – Box-plots of observed data in all textural soil classes. The circles represent the outliers.

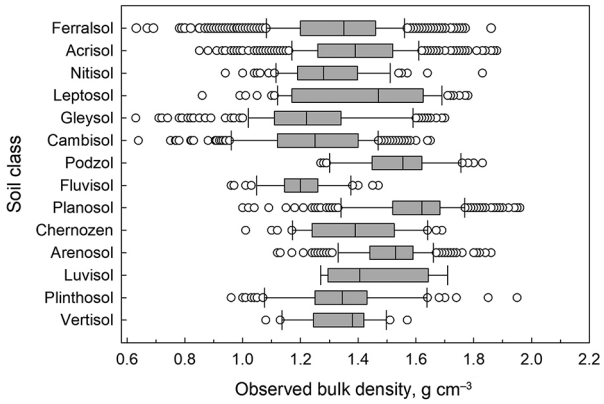

Figure 3 – Box-plots of observed data in all World Reference Base for Soil Classes. The circles represent the outliers.

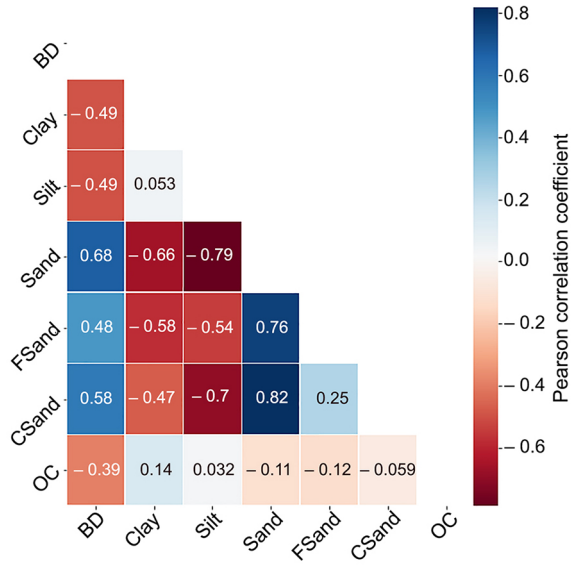

Figure 4 – Correlation heatmap of variables from 990 soil samples. BD = bulk density; OC = organic carbon; FSand and CSand = fine and coarse sand respectively.

Table 4 – Results obtained by evaluating the BD-PTFs for the presented database.

PTF	RMSE	R^2	Number of samples
	g cm^{-3}		
PTF-A	0.20	0.37	2827
PTF-B	0.24	0.23	2955
PTF-C	0.22	0.34	2711
PTF-D	0.27	0.09	1855
PTF-E	0.18	0.25	3050
PTF-F	0.17	0.34	2804
PTF-G	0.42	0.12	3050
PTF-H	0.19	0.22	3039
PTF-I	0.34	0.13	1695
PTF-J	0.21	0.06	2330
PTF-K	0.20	0.12	3050

RMSE = root mean square error; R^2 = coefficient of determination; Number of samples evaluated by the eleven BD-PTFs respecting the ranges of the input variables.

A correlation analysis was performed on a set of 990 samples (Figure 4), which included all the mandatory information for all the PTF that was tested. The visual results of BD-PTFs can be seen in Figure 5A-K, which shows BD's observed and predicted values.

Evaluated PTF-A and -F had the best indices (RMSE, R^2). However, PTF-F exhibited more significant variation, particularly for high BD values (RMSE = 0.18 g cm^{-3}) (Table 4). Among the BD-PTFs analyzed in Table 4, the PTF-D, -G, -I had the highest RMSE values (0.27, 0.42, and 0.34 g cm^{-3}), indicating poor performance. Additionally, the BD estimates obtained using the PTF-J had a low correlation coefficient (R^2 = 0.06), indicating poor goodness of fit. The highest residues were observed for BD values below 1.0 and above 1.5 g cm^{-3}.

The coefficients and predictors of our proposed hierarchical functions, selected using step AIC, are presented in Table 5. The calibration statistics of the four developed PTFs are provided in Table 6, while the 10-fold cross-validation statistics are shown in Table 7. Despite differences in sample size, the sample standard deviation (SD) of all four training datasets was 0.21 g cm^{-3}. The value of $RMSE_{calib}$ decreased with an increasing number of predictors in PTF-4. To correct the $RMSE_{calib}$ to the sample standard deviation, we divided RMSE by SD. This yielded a corrected $RMSE_{calib}$ of 0.81 and 0.66 for PTF-1 and PTF-4, respectively. The presentation of the visual outcomes for our proposed hierarchical PTFs can be observed in Figure 6A-D. Additionally, Figure 7A-D illustrates the results of the residual analysis. Furthermore, Figure 8 provides a comprehensive evaluation of the PTFs applied to the same set of samples (990 samples).

Table 5 – Coefficients of developed pedotransfer functions (PTFs) with different predictors.

Functions	Predictors	PTF
PTF-1	*Sa, Cl*	$BD = 1.286 + 3.208 \times 10^{-3} (Sa) - 2.013 \times 10^{-3} (Cl)$
PTF-2	*Sa, Cl, OC*	$BD = 1.358 + 2.79 \times 10^{-3} (Sa) - 2.328 \times 10^{-3} (Cl) + 0.052 (OC)$
PTF-3	Sa_f, Sa_c, Cl	$BD = 1.198 + 2.971 \times 10^{-3} (Sa_f) + 4.472 \times 10^{-3} (Sa_c) - 8.706 \times 10^{-4} (Cl)$
PTF-4	Sa_f, Sa_c, *Cl, OC*	$BD = 1.243 + 2.983 \times 10^{-3} (Sa_f) + 4.187 \times 10^{-3} (Sa_c) - 6.208 \times 10^{-2} (OC) - 5.793 \times 10^{-4} (Cl)$

Sa = total sand content; Sa_f = fine sand content; Sa_c = coarse sand content; *Cl* = clay content; OC = organic carbon.

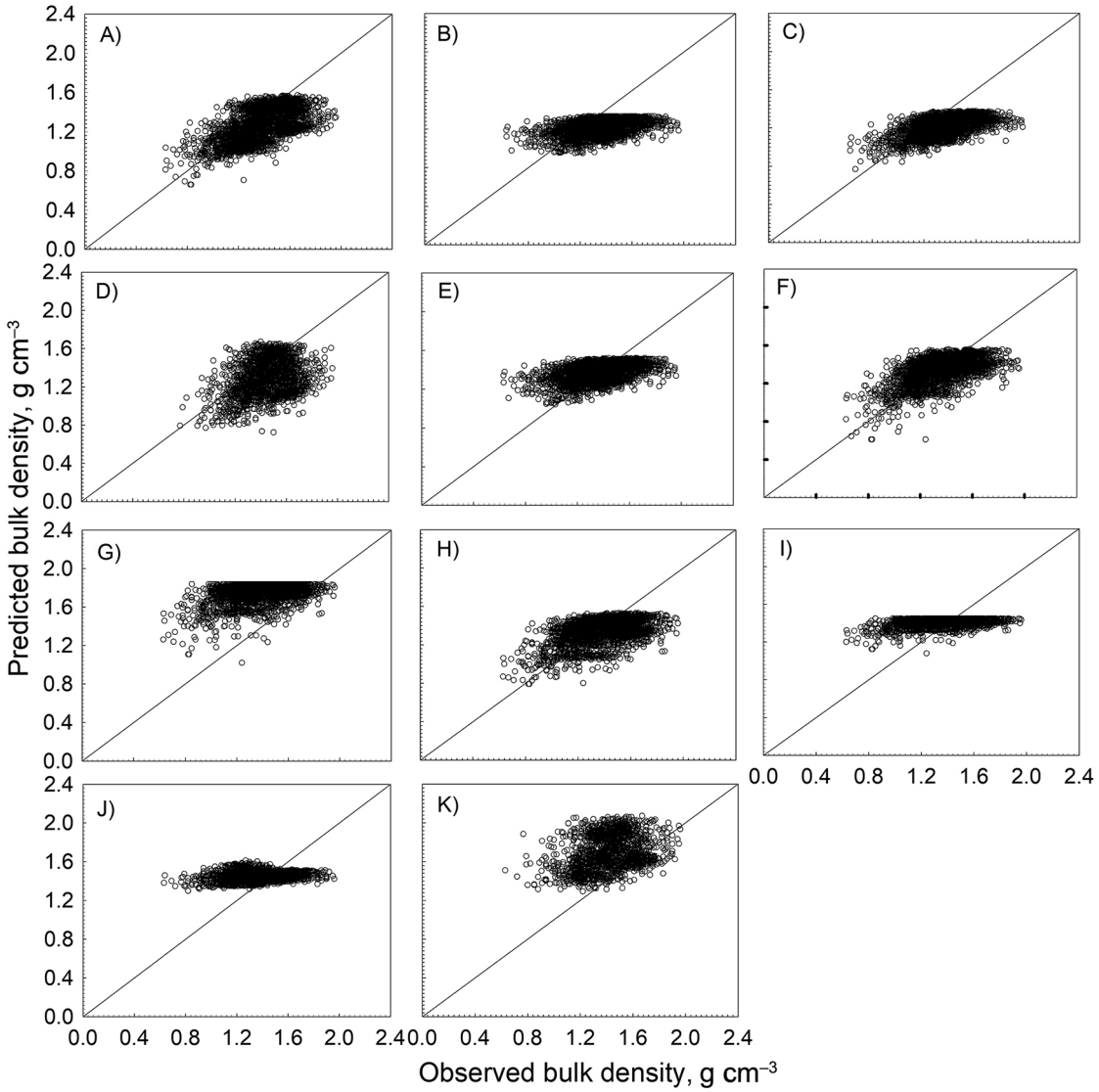

Figure 5 – Observed and predicted bulk density values by eleven BD-PTFs to Brazilian Soils. (A) PTF-A by Tomasella and Hodnett (1998), (B) PTF-B by Bernoux et al. (1998), (C) PTF-C by Bernoux et al. (1998), (D) PTF-D by Kaur et al. (2002), (E) PTF-E by Benites et al. (2007), (F) PTF-F by Benites et al. (2007), (G) PTF-G by Ruehmann and Körschens (2009), (H) PTF-H by Hollis et al. (2012), (I) PTF-I by Al-Qinna and Jaber (2013), (J) PTF-J by Botula et al. (2015), and (K) PTF-K by Abdelbaki (2016).

Discussion

Soils with a predominance of sand fractions (i.e., sand, sand clay, sand, loam) tend to show high values for BD. In contrast, clayey soils are quite unpredictable as they can have low (e.g., 0.63 g cm^{-3}) and high BD values (e.g., above 1.80 g cm^{-3}). We omitted Organosols, which typically show very low BD values, in our analysis.

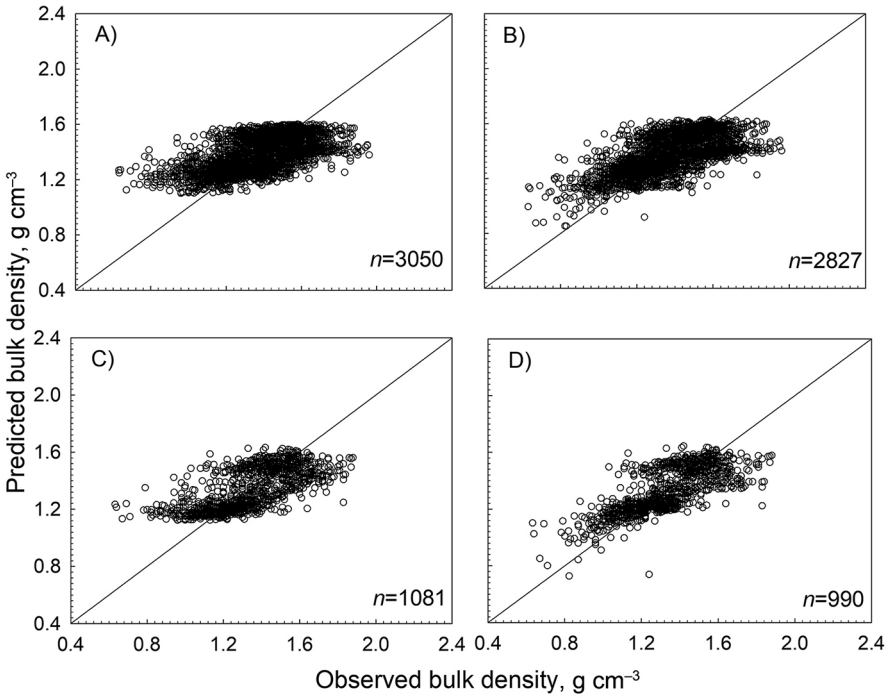

Figure 6 – Comparison between observed and predicted bulk density values by our proposed pedotransfer functions (PTFs). (A) PTF-1, (B) PTF-2, (C) PTF-3, (D) PTF-4.

Table 6 – Goodness of fit criteria (RMSE and R^2) for the calibration functions (PTF) to predict bulk density.

Summary	PTF-1	PTF-2	PTF-3	PTF-4
Number of samples	3,050	2,827	1,081	990
$RMSE_{calib}$	0.17	0.16	0.16	0.14
R^2_{calib}	0.35	0.43	0.44	0.56

$RMSE_{calib}$ = root mean square error (g cm^{-3}) of calibration data; R^2_{calib} = coefficient of determination of calibration data.

Table 7 – Statistics (RMSE and R^2) of the k-fold cross validation.

Summary	PTF-1	PTF-2	PTF-3	PTF-4
Number of samples	275 × 10	254 × 10	97 × 10	88 × 10
$RMSE_{validation}$	0.17	0.16	0.16	0.14
$R^2_{validation}$	0.35	0.43	0.45	0.56

$RMSE_{validation}$ = root mean square error (g cm^{-3}) of validation; R^2 = coefficient of determination.

A correlation analysis was performed on 990 samples, which included all the mandatory information for all tested PTF. A correlation heatmap is a type of plot that visualizes the strength of the relationship between two variables. The correlation analysis shows that BD is negatively correlated with silt and clay and positively correlated with total sand, fine and coarse sand (Figure 4). This is one reason why certain published BD-PTFs use only one or two particle size fractions as predictors (Bernoux et al., 1998; Benites et al., 2007). Despite the moderately negative correlation (–0.39) between BD and OC, we kept OC in some of our functions as it is usually available in soil surveys. Bulk density is negatively correlated with clay content and positively correlated with sand content, which reflects sandy soils (e.g., Arenosols, Spodosols) typically having high values of BD, and clayey soils, especially the well-structured Ferralsols and Acrisols in the tropics, having low BD values (Ottoni et al., 2018; Batjes et al., 2020). On the other hand, positive correlations between BD and total sand, fine and coarse sand are demonstrated in PTFs 1, 2, 3, and 4.

The PTFs of Tomasella and Hodnett (1998) (PTF-A), and Bernoux et al. (1998) (PTF-B, C), were developed using Amazonian soil samples. Bernoux et al. (1998) found an R^2 of approximately 0.50 for estimating BD with clay and OC content. Tomasella and Hodnett (1998) obtained an R^2 value almost equal to 0.60 when estimating BD with silt, clay, and OC content. Benites et al. (2007) also developed PTFs for Brazilian Soils across most biomes and found an R^2 value of 0.63 when predicting BD with clay and OC content (PTF-E, F). Boschi et al. (2018) evaluated a set of 222 soil profile data from all Brazilian biomes, totaling 884 samples, and found that the PTF proposed by Benites et al. (2007) (PTF-F) had good performance (RMSE = 0.19 g cm^{-3}).

Al-Qinna and Jaber (2013) proposed several PTFs that were evaluated using various methodologies to estimate the BD of Jordanian soils, located in an arid region. We tested one of these PTFs (PTF-I) that uses

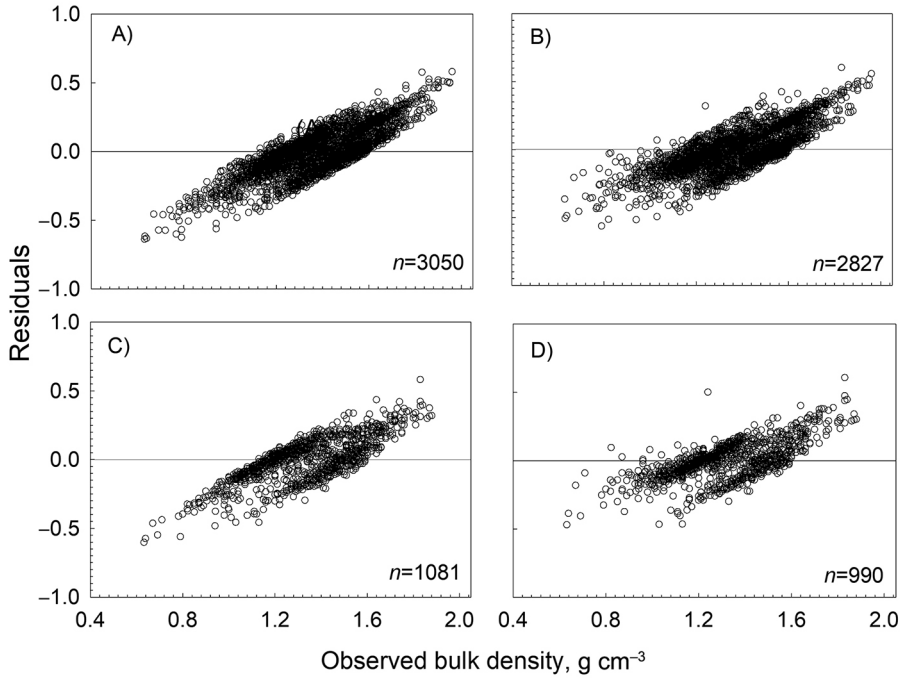

Figure 7 – Observed vs residuals of proposed pedotransfer functions (PTFs). (A) PTF-1, (B) PTF-2, (C) PTF-3, and (D) PTF-4.

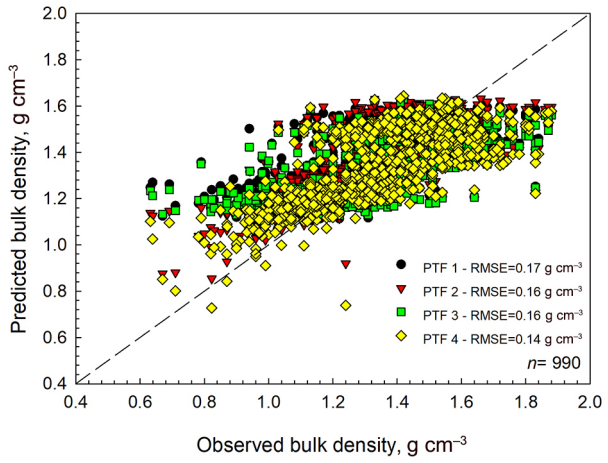

Figure 8 – Comparison of bulk density estimates by the hierarchical system using four pedotransfer functions (PTFs) for the same data set.

sand and log (OC) content, which was developed using stepwise regression. They found an RMSE equal to 0.126 g cm^{-3} for their analysis whereas in our study this PTF performed poorly (Table 4).

We also tested the PTF proposed by Abdelbaki (2016), which uses only the OC content as a predictor of BD (PTF-K). The author utilized an extensive soil database from the US (SSURGO) comprising approximately 174,339 samples, with OC content ranging from 0- to 58 %, and BD values varying from 0.30 to 2.30 g cm^{-3}. In his study, the author found an RMSE of 0.13 g cm^{-3}, superior to other PTFs tested. However, our study obtained an RMSE value of 0.20 g cm^{-3} when using this same PTF. This same author tested several PTFs, and the ones proposed by Ruehlmann and Körschens (2009) and Hollis et al. (2012) had the best performances. The PTF by Ruehlmann and Körschens (2009) uses only the OC content and was developed using a global dataset (PTF-G). The PTF proposed by Hollis et al. (2012) was developed using a European dataset, which includes OC, sand, and clay content (PTF-H). In our study, we obtained RMSE values of 0.42 and 0.19 g cm^{-3} for the PTFs by Ruehlmann and Körschens (2009) and Hollis et al. (2012), respectively (PTF-G, H) (see Table 4).

The results of the statistical indices (RMSE, R^2) for the calibration and the validation of the models were very similar. Pedotransfer function number 1 estimated the BD for all samples in the database (n = 3,050). Since the other PTFs use predictor variables that are not present in all samples, the number of samples used to calibrate PTF 2, 3 and 4 were, 2,827, 1,081 and 990, respectively.

The proposed PTF-1 can be compared with the PTFs B and E, developed by Bernoux et al. (1998) and Benites et al. (2007), in which predictor parameters also correspond to granulometric fractions. The PTFs B and E obtained RMSE values equal to 0.22 and 0.18 g cm^{-3}, respectively, while PTF-1 had the lowest RMSE (0.17 g cm^{-3}, as shown in Table 7).

The proposed PTF-2, which uses textural fractions and OC to estimate BD, can be compared to other PTFs as PTF-A, -C, -D, -F, -H, -I, -J that presented RMSE values equal to 0.22, 0.20, 0.27, 0.17, 0.19, 0.34, and 0.21 g cm^{-3} (Table 4), respectively, for the BD estimate. The BD prediction by PTF-2 showed a lower RMSE (0.16 g cm^{-3}), presenting the same RMSE value in the cross-validation assessment (Table 7).

The proposed PTF-3 was fitted using 1,081 samples, equiring fine and coarse sand as predictors instead of total sand. However, PTF-3, like PTF-1, uses only particle size fractions as predictors. When comparing predictions for these 1,081 samples, the RMSE values were 0.159 and 0.168 g cm^{-3}, respectively, showing a slight improvement in the accuracy when total sand is fractionated.

The last proposed PTF is PTF-4, which uses textural fractions and OC content as predictors, showed a lower RMSE (0.14 g cm^{-3}) in validation and, therefore, is considered a more accurate PTF when fine and coarse sand, clay and OC are available. On the other hand, PTF-1 showed higher residuals, underestimating (positive values) the BD prediction for samples with values above 1.60 g cm^{-3}, and overestimating (negative values) the BD for samples with BD less than 1.0 g cm^{-3}. The PTF with the smallest error was PTF-4 with an RMSE of 0.14 g cm^{-3}, while the worst (RMSE = 0.17 g cm^{-3}) was observed when using only simple textural fractions (sand and clay – PTF-1).

Satisfactory performance is indicated when the RMSE for BD predictions falls within the range of 0.12 to 0.25 g cm^{-3}, as defined by De Vos et al. (2005). Using this criterion, some evaluated models show good performance for Brazilian soils. As suggested by Al-Qinna and Jaber (2013), it is essential to balance the simplicity, applicability, accuracy, precision, and reliability of models generated in specific circumstances. Therefore, we believe that our four proposed PTFs are suitable for use in various scenarios within Brazilian territory where there is a lack of available BD data.

The script and data bank will be available on the GitHub of the first author of this manuscript.

We proposed four new simple functions to predict bulk density (BD) values for most Brazilian Soils. These functions use granulometric fractions and organic carbon as predictor variables. A hierarchical system of BD predictions showed a slight improvement in accuracy when organic carbon and coarse and fine sand fractions were included as predictors. However, the BD predictions still showed high residuals for soils with low BD values (i.e., below 1.00 g cm^{-3}) and high BD values (i.e., above 1.50 g cm^{-3}). Incorporating a structural parameter representing the arrangement of soil particles is crucial to improving the accuracy of BD predictions. The best results with existing PTFs were obtained for PTF-A and -F, proposed by Tomasella and Hodnett (1998) and Benites et al. (2007), respectively.

Acknowledgments

The first author would like to acknowledge the Fundação de Apoio à Pesquisa e ao Desenvolvimento (FAPED) - and the ZARC Project. The second author would like to acknowledge the financial support provided by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). The authors also acknowledge the two anonymous reviewers' helpful comments to improve this manuscript.

Authors' Contributions

Conceptualization: Teixeira WG. **Data curation:** Huf dos Reis AM, Teixeira WG, Fontana A, Barros AHC, Vasques GM, Samuel-Rosa A, Ottoni MV. **Formal analysis:** Huf dos Reis AM, Teixeira WG, Vasques GM, Samuel-Rosa A, Victoria DC. **Methodology:** Huf dos Reis AM, Teixeira WG, Fontana A. **Project administration:** Huf dos Reis AM, Teixeira WG. **Writing – original draft:** Huf dos Reis AM, Teixeira WG. **Writing – review & editing:** Huf dos Reis AM, Teixeira WG, Fontana A, Barros AHC, Vasques GM, Victoria DC, Samuel-Rosa A, Ottoni MV, Monteiro JEB. **Funding acquisition:** Monteiro JEB.

References

Abdelbaki AM. 2016. Evaluation of pedotransfer functions for predicting soil bulk density for U.S. soils. Ain Shams Engineering Journal 9: 1611-1619. https://doi.org/10.1016/j.asej.2016.12.002

Alexander EB. 1980. Bulk densities of California soils in relation to other soil properties. Soil Science Society of America Journal 44: 689-692. https://doi.org/10.2136/sssaj1980.03615995004400040005x

Al-Qinna MI, Jaber SM. 2013. Predicting soil bulk density using advanced pedotransfer functions in an arid environment. Transactions of the ASABE 56: 963-976. https://doi.org/10.13031/trans.56.9922

Assouline S. 2006. Modeling the relationship between soil bulk density and the water retention curve. Vadose Zone Journal 5: 554-563. https://doi.org/10.2136/vzj2005.0083

Barros AHC, van Lier QJ, Maia AHN, Scarpare FV. 2013. Pedotransfer functions to estimate water retention parameters of soils in northeastern Brazil. Revista Brasileira de Ciência do Solo 37: 379-391. https://doi.org/10.1590/S0100-06832013000200009

Batjes NH, Ribeiro E, van Oostrum A. 2020. Standardised soil profile data to support global mapping and modelling (WoSIS snapshot 2019). Earth System Science Data 12: 299-320. https://doi.org/10.5194/essd-12-299-2020

Benites VM, Machado PLOA, Fidalgo ECC, Coelho MR, Madari BE. 2007. Pedotransfer functions for estimating soil bulk density from existing soil survey reports in Brazil. Geoderma 139: 90-97. https://doi.org/10.1016/j.geoderma.2007.01.005

Bernoux M, Cerri C, Arrouays D, Jolivet C, Volkoff B. 1998. Bulk densities of Brazilian Amazon soils related to other soil properties. Soil Science Society of America Journal 62: 743-749. https://doi.org/10.2136/sssaj1998.03615995006200030029x

Boschi RS, Bocca FF, Lopes-Assad MLRC, Assad ED. 2018. How accurate are pedotransfer functions for bulk density for Brazilian soils? Scientia Agricola 75: 70-78. https://doi.org/10.1590/1678-992x-2016-0357

Botula Y-D, Nemes A, Van Ranst E, Mafuka P, De Pue J, Cornelis WM. 2015. Hierarchical pedotransfer functions to predict bulk density of highly weathered soils in central Africa. Soil Science Society of America Journal 79: 476-486. https://doi.org/10.2136/sssaj2014.06.0238

Chen S, Richer-de-Forges AC, Saby NPA, Martin MP, Walter C, Arrouays D. 2018. Building a pedotransfer function for soil bulk density on regional dataset and testing its validity over a larger area. Geoderma 312: 52-63. https://doi.org/10.1016/j.geoderma.2017.10.009

De Vos B, Van Meirvenne M, Quataert P, Deckers 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. https://doi.org/10.2136/sssaj2005.0500

Donagemma GK, Viana JHM, Almeida BG, Ruiz HÁ, Klein VA, Dechen SCF, Fernandes RBA. Análise granulométtrica. p.95-116. In: Teixeira PC, Donagemma GK, Fontana A, Teixeira WG. (Org) 2017. Manual of Soil Analysis Methods = Manual de Métodos de Análise de Solo. 3ed. Embrapa, Brasília, DF, Brazil (in Portuguese).

Don A, Schumacher J, Freibauer A. 2011. Impact of tropical land-use change on soil organic carbon stocks - a meta-analysis: soil organic carbon and land-use change. Global Change Biology 17: 1658-1670. https://doi.org/10.1111/j.1365-2486.2010.02336.x

Fontana A, Campos DVB. Carbono orgânico. p.360-367. In: Teixeira PC, Donagemma GK, Fontana A, Teixeira WG. (Org) 2017. Manual of Soil Analysis Methods = Manual de Métodos de Análise de Solo. 3ed. Embrapa, Brasília, DF, Brazil (in Portuguese).

Food and Agriculture Organization of the United Nations [FAO]. 2015. World Reference Base for Soil Resources 2015: International Soil Classification System for Naming Soils and Creating Legends for Soil Maps. FAO, Rome, Italy.

Gunarathna MHJP, Sakai K, Nakandakari T, Momii K, Kumari MKN, Amarasekara MGTS. 2019. Pedotransfer functions to estimate hydraulic properties of tropical Sri Lankan soils. Soil and Tillage Research 190: 109-119. https://doi.org/10.1016/j.still.2019.02.009

Haddad DB, Assis LS, Tarrataca L, Gomes AS, Ceddia MB, Oliveira RF, et al. 2018. Brazilian Soil Bulk Density Prediction Based on a Committee of Neural Regressors. IEEE, Rio de Janeiro, RJ, Brazil. https://doi.org/10.1109/IJCNN.2018.8489177

Hillel D. 1998. Environmental Soil Physics. Academic Press, San Diego, CA, USA.

Hollis JM, Hannam J, Bellamy PH. 2012. Empirically-derived pedotransfer functions for predicting bulk density in European soils. European Journal of Soil Science 63: 96-109. https://doi.org/10.1111/j.1365-2389.2011.01412.x

Inagaki TM, Moraes Sá JC, Caires EF, Gonçalves DRP. 2017. Why does carbon increase in highly weathered soil under no-till upon lime and gypsum use? Science of The Total Environment 599-600: 523-532. https://doi.org/10.1016/j.scitotenv.2017.04.234

Karup D, Moldrup P, Tuller M, Arthur E, de Jonge LW. 2017. Prediction of the soil water retention curve for structured soil from saturation to oven-dryness. European Journal of Soil Science 68: 57-65. https://doi.org/10.1111/ejss.12401

Kaur R, Kumar S, Gurung HP. 2002. A pedo-transfer function (PTF) for estimating soil bulk density from basic soil data and its comparison with existing PTFs. Soil Research 40: 847-858. https://doi.org/10.1071/SR01023

Kuhn M. 2008. Building predictive models in R using the caret package. Journal of Statistical Software 28: 1-26. https://doi.org/10.18637/jss.v028.i05

Lettens S, Van Orshoven J, van Wesemael B, De Vos B, Muys B. 2005. Stocks and fluxes of soil organic carbon for landscape units in Belgium derived from heterogeneous data sets for 1990 and 2000. Geoderma 127: 11-23. https://doi.org/10.1016/j.geoderma.2004.11.001

Manrique LA, Jones CA. 1991. Bulk density of soils in relation to soil physical and chemical properties. Soil Science Society of America Journal 55: 476-481. https://doi.org/10.2136/sssaj1991.03615995005500020030x

Minasny B, McBratney AB, Bristow KL. 1999. Comparison of different approaches to the development of pedotransfer functions for water-retention curves. Geoderma 93: 225-253. https://doi.org/10.1016/S0016-7061(99)00061-0

Minasny B, Hartemink AE. 2011. Predicting soil properties in the tropics. Earth-Science Reviews 106: 52-62. https://doi.org/10.1016/j.earscirev.2011.01.005

Nanko K, Ugawa S, Hashimoto S, Imaya A, Kobayashi M, Sakai H, et al. 2014. A pedotransfer function for estimating bulk density of forest soil in Japan affected by volcanic ash. Geoderma 213: 36-45. https://doi.org/10.1016/j.geoderma.2013.07.025

Nasta P, Palladino M, Sica B, Pizzolante A, Trifuoggi M, Toscanesi M, et al. 2020. Evaluating pedotransfer functions for predicting soil bulk density using hierarchical mapping information in Campania, Italy. Geoderma Regional 21: e00267. https://doi.org/10.1016/j.geodrs.2020.e00267

Ottoni MV, Ottoni Filho TB, Schaap MG, Lopes-Assad MLRC, Rotunno Filho OC. 2018. Hydrophysical database for Brazilian soils (HYBRAS) and pedotransfer functions for water retention. Vadose Zone Journal 17: 1-17. https://doi.org/10.2136/vzj2017.05.0095

Ottoni MV, Ottoni Filho TB, Lopes-Assad MLRC, Rotunno Filho OC. 2019. Pedotransfer functions for saturated hydraulic conductivity using a database with temperate and tropical climate soils. Journal of Hydrology 575: 1345-1358. https://doi.org/10.1016/j.jhydrol.2019.05.050

Palladino M, Romano N, Pasolli E, Nasta P. 2022. Developing pedotransfer functions for predicting soil bulk density in Campania. Geoderma 412: 115726. https://doi.org/10.1016/j.geoderma.2022.115726

Ruehlmann J, Körschens M. 2009. Calculating the effect of soil organic matter concentration on soil bulk density. Soil Science Society of America Journal 73: 876-885. https://doi.org/10.2136/sssaj2007.0149

Samuel-Rosa A, Dalmolin RSD, Moura-Bueno JM, Teixeira WG, Alba JMF. 2020. Open legacy soil survey data in Brazil: geospatial data quality and how to improve it. Scientia Agricola 77: e20170430. https://doi.org/10.1590/1678-992x-2017-0430

Saxton KE, Rawls WJ. 2006. Soil water characteristic estimates by texture and organic matter for hydrologic solutions. Soil Science Society of America Journal 70: 1569-1578. https://doi.org/10.2136/sssaj2005.0117

Sevastas S, Gasparatos D, Botsis D, Siarkos I, Diamantaras KI, Bilas G. 2018. Predicting bulk density using pedotransfer functions for soils in the upper Anthemountas basin, Greece. Geoderma Regional 14: e00169. https://doi.org/10.1016/j.GEODRS.2018.e00169

Silva AC, Armindo RA, Prevedello CL. 2020. Splintex 2.0: a physically-based model to estimate water retention and hydraulic conductivity parameters from soil physical data. Computers and Electronics in Agriculture 169: 105157. https://doi.org/10.1016/j.compag.2019.105157

Soil Survey Staff. 2014. Keys to Soil Taxonomy. USDA, Washington, DC, USA

Souza E, Fernandes Filho EI, Schaefer CEGR, Batjes NH, Santos GR, Pontes LM. 2016. Pedotransfer functions to estimate bulk density from soil properties and environmental covariates: Rio Doce basin. Scientia Agricola 73: 525-534. https://doi.org/10.1590/0103-9016-2015-0485

Tomasella J, Hodnett MG. 1998. Estimating soil water retention characteristics from limited data in Brazilian Amazonia. Soil Science 163: 190-202. https://doi.org/10.1097/00010694-199803000-00003