Soil Organic Carbon Stock at 0-30 cm Map for Brazil – Technical Report

Authors: Gustavo M. Vasques[1], Ricardo O. Dart[1], Jesus F. M. Baca[1], Marcos B. Ceddia[2], Maria de Lourdes Mendonça Santos[3]

Contributors: Daniel V. Perez[1], Guillermo F. Olmedo[4], Mario A. Guevara[5], Ronald Vargas[6], Andréa S. Gomes[2], Mauricio R. Coelho[1], Humberto G. Santos[1], Wenceslau G. Teixeira[1]

[1]Embrapa Solos, [2]UFRRJ and [3]Embrapa Cocais, Brazil; [4]INTA, Argentina; [5]University of Delaware, USA; [6]FAO/GSP, Italy

Soil data: The soil profile dataset was composed by merging the *Sistema de Informação de Solos Brasileiros* (https://www.bdsolos.cnptia.embrapa.br/consulta_publica.html), *Escola Superior de Agricultura Luiz de Queiroz* (http://www.esalq.usp.br/gerd) and *Sistema de Proteção da Amazônia* (not online) data sources. After partial error checking and removal of duplicate soil profiles, 10,029 (8439 georeferenced) profiles remained.

Covariates: Raster layers representing soil formation factors, including 7 MODIS mosaics (https://modis.gsfc.nasa.gov) from 11 dates (77 bands), 19 WorldClim mosaics (www.worldclim.org), and a 3-arc second SRTM DEM (https://lta.cr.usgs.gov/SRTM) were assembled, resampled to 1-km spatial resolution, and processed in SAGA GIS (Conrad et al., 2015) to produce NDVI (11 dates) and 37 terrain attribute layers, totaling 145 covariate layers.

Data preparation: Soil organic carbon stock at 0-30 cm (OCS) was calculated as the sum of stocks from horizons/layers within 0-30 cm. Missing BD values were predicted from carbon, clay and silt contents, and depth, with R^2 of 0.60 (for samples with carbon contents >= 60 g kg^{-1}), or from carbon and sand contents, and depth, with R^2 of 0.36 (for samples with carbon contents < 60 g kg^{-1}). The environmental covariates values were extracted to the soil sampling points by spatial overlay. Samples without OCS data were removed, and principal components analyses were done on soil variables, and then on environmental covariates to identify and remove 18 outliers. The remaining 7015 samples were split randomly into training (5575 samples ~80%) and validation (1439 samples ~20%) sets, and OCS was transformed to logarithm for modeling.

Prediction: A set of 45 covariates were selected based on the number of missing values, correlation with OCS, covariate redundancy/collinearity, variable importance in preliminary random forest model, and presence of undesirable spatial artifacts. A preliminary full multiple linear regression model was derived to identify and remove five influential outliers. An ensemble prediction model combining nine methods (stepwise multiple linear regression, elastic net, principal components regression, partial least squares regression, multivariate adaptive regression splines, cubist, regression tree, random forest, and extreme gradient boosting) using the caret (Kunh, 2017), caretEnsemble (Deane-Mayer & Knowles, 2016), and method-specific R (R Core Team, 2017) packages. Ten-fold cross-validation was used to optimize the methods parameters and train the nine OCS model, and generalized least squares was used to combine the predictions from the nine methods in the ensemble model. Predictions were made across Brazil at 1-km spatial resolution, and then back-transformed to OCS original units.

Uncertainty assessment: Prediction errors were calculated on the training and validation sets, respectively, using the back-transformed predictions. Prediction uncertainty measures for each pixel were calculated using the argument `se=TRUE` in the `predict` function, which calculates the lower and upper prediction bounds at p=0.05.

Results: The final training and validation sets have 5570 and 1439 samples, respectively, after removal of incomplete samples (without OCS or any of the 45 selected covariates). The observed OCS varies from 2.6 to 417.3 t ha^{-1}, with a mean of 4.8, median of 4.1 and standard deviation of 3.3 t ha^{-1}. The 45 covariates selected to train the models include 23 MODIS (19 band and four NDVI), seven WorldClim (two precipitation, three temperature, water vapor pressure and solar radiation), and 15 SRTM-derived relief layers (DEM, LS factor, plan curvature, general curvature, CTI, MRRTF, MRVBF, vertical distance to channel network, valley depth, mass balance index, catchment slope, terrain surface convexity, terrain ruggedness index, vector ruggedness measure and total insolation). The mean OCS predictions across Brazil vary from 4.9 to 238.3 t ha^{-1}, with a mean of 42.0 and median of 41.1 t ha^{-1}. The training and validation R^2 are 0.77 and 0.24, mean errors are 4.9 and 5.2 t ha^{-1}, and root mean square errors are 19.5 and 28.3 t ha^{-1}, respectively. The estimated total soil organic carbon stock at 0-30 cm for Brazil is 36.3 ± 0.2 Pg (1 Pg = 10^9 t) (mean ± 1.96 * standard deviation of prediction), which compares to previous estimates of 36.4 Pg by Bernoux et al. (2002) and 36.6 Pg by Fidalgo et al. (2007).

Final remarks: The OCS and uncertainty maps were produced using legacy soil data, freely available environmental covariates, and free and open source software to be part of the Global Soil Organic Carbon Map, one of the proposed aims of FAO's Global Soil Partnership, in support of the Sustainable Development Goal Indicator 15.3.1. The soil data, covariate data and software providers are acknowledged. Questions and suggestions to improve these maps are welcome.

References:

Bernoux, M. et al. (2002) Brazil's soil carbon stocks. Soil Sci. Soc. Am. J. 66: 888–896.

Conrad, O. et al. (2015) System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geosci. Model Dev. 8: 1991–2007.

Fidalgo, E.C.C. et al. (2007) Estoque de carbono nos solos do Brasil. Bol. Pesq. Des. 8. Embrapa Solos, Rio de Janeiro, Brazil.

Kuhn, M. et al. (2017). caret: Classification and regression training. Version 6.0-77.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Version 3.4.2.

Deane-Mayer, Z.A. & Knowles, J.E. (2016). caretEnsemble: Ensembles of caret models. Version 2.0.0.

Contact information: G.M. Vasques, gustavo.vasques@embrapa.br; R.O. Dart, ricardo.dart@embrapa.br; J.F.M. Baca, jesus.baca@embrapa.br; M.B. Ceddia, ceddia@ufrrj.br; M.L. Mendonça Santos, lourdes.mendonca@embrapa.br