Geoderma 232–234 (2014) 479–486

Contents lists available at ScienceDirect

Geoderma

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

A regional-scale assessment of digital mapping of soil attributes in a tropical hillslope environment

Waldir de Carvalho Junior [a,*], Philippe Lagacherie [b], Cesar da Silva Chagas [a], Braz Calderano Filho [a], Silvio Barge Bhering [a]

[a] Embrapa Solos, Rua Jardim Botânico, 1024, Rio de Janeiro, RJ, Brazil
[b] INRA, LISAH, Montpellier, France

ARTICLE INFO

Article history:
Received 30 December 2013
Received in revised form 3 June 2014
Accepted 6 June 2014
Available online 17 June 2014

Keywords:
R
GlobalSoilMap.Net
Kriging
Linear models
Regression tree

ABSTRACT

The purpose of this study was to analyze the relationships between soil attributes and environmental covariates in a tropical hillslope environment on a regional scale to estimate spatial distribution of soil attributes and identify statistical and geostatistical techniques that could represent the variation of the soil attributes. The study was performed in Bom Jardim County, Brazil, and covered an area of 390 km^2 with a soil database of 208 sample points distributed in six depth layers (0.53 pts/km^2). The study used 18 environmental covariates derived from DEM and satellite imagery. The models evaluated were linear regression, regression trees and ordinary and regression kriging. An exploratory analysis showed that DEM, NDVI, MRVBF, MSP, b3/b2, b5/b7, SPI, SWI, SLOPE and ASPECT were correlated with soil properties. The models performance had a mean crossvalidation r^2 of 0.13. The best results were achieved with kriging models, with a crossvalidation r^2 of 0.19. A comparison between multiple linear regression and regression trees showed that the tree model yielded the best results. The sample density alone could not explain the results, but an interaction between DEM accuracy, sample density, covariates and geological conditions was suitable as an explanatory factor. Studies of tropical hillslope digital soil mapping on regional scales need to be more exhaustively focused to develop this research area.

1. Introduction

It is recognized that Earth's natural resources are not inexhaustible. However, the demand imposed upon these resources is immense and is increasing beyond the ability of natural cycles to replenish the resource supply. Thus, the sustainability and the management of these resources are vital for the survival of all life on the planet, including humans. The management of these resources requires knowledge and information about them (Odeh et al., 2007). Land is an important component of the planet's natural resources and needs to be conserved for future generations. In this context, soils play an important role. To benefit from the ecological and economic functions of the soil in a sustainable manner, landholders, corporate stakeholders and government departments need access to quantitative information on soils (Malone et al., 2009). There is a clear need for information on precise quantitative relationships between soils and key environmental factors to facilitate soil data collection and soil modeling worldwide. Such relationships form the basis of digital soil mapping (DSM) techniques, which are widely considered to represent the future of soil surveys (Lagacherie and McBratney, 2007).

Digital maps of soil attributes can meet this need through the use of interrelationships between soil attributes and landscape covariates. Numerous approaches to the topic of digital maps of soil classes and attributes have been developed, and various techniques have been applied in this area of study. Grunwald (2009) summarized the DSM and modeling papers (2007–2008) published in Geoderma and Soil Science Society of America journals and showed that of the 90 papers surveyed, 37 used regression and variants of regression, 29 used classification/discrimination, 17 used kriging and 12 used tree-based models. The principal approaches associated with DSM involved data-driven knowledge tools such as geostatistics, data mining, neural networks and expert-knowledge modeling. For example, Carvalho Junior et al. (2011) and Chagas et al., 2011 used a neural network approach to map soil classes and attributes; Ciampalini et al. (2012) and Odeh et al. (2007) used regression and kriging to estimate soil attributes; and Malone et al. (2009) used statistics and neural networks to map carbon in the soil.

These DSM techniques apply analyses of the spatial relationships of soil, terrain attributes and remote sensing images to model the environment. Bui (2010) has demonstrated the importance of geographical soil databases for mapping ecosystem functions. Focusing on this goal, GlobalSoilMap.Net will draw on emerging technologies to provide a

* Corresponding author.
E-mail addresses: waldir.carvalho@embrapa.br (W. de Carvalho Junior), lagache@supagro.inra.fr (P. Lagacherie), cesar.chagas@embrapa.br (C. da Silva Chagas), braz.calderano@embrapa.br (B. Calderano Filho), silvio.bhering@embrapa.br (S.B. Bhering).

http://dx.doi.org/10.1016/j.geoderma.2014.06.007
0016-7061/

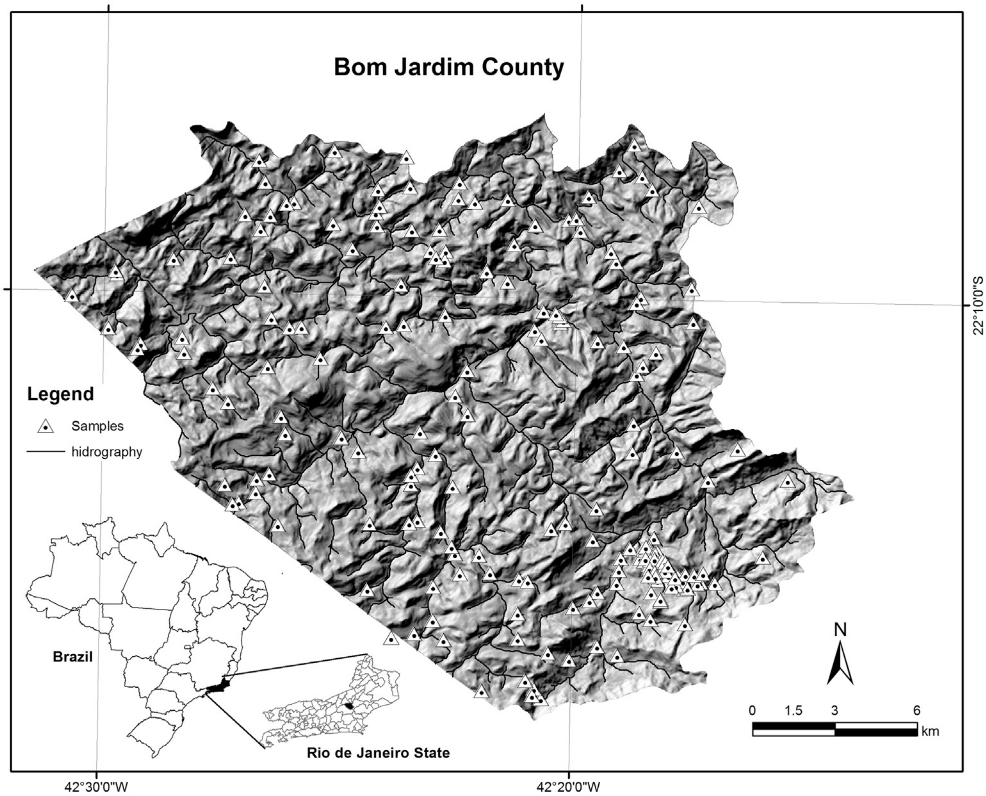

Fig. 1. Study area and locations of the soil points for the dataset used in this study.

more detailed view of the quality of the world's soils and will help scientists, governments and farmers make better decisions about producing food, eradicating hunger, managing climate change and reducing environmental degradation (Media Release, 2010). As a first step, the digital maps of soil properties specified in GlobalSoilMap should be built from the legacy of soil data that has resulted from the efforts of past generations of soil surveyors (Mayr et al., 2008). The GlobalSoilMap.Net project recommends that pedologists seek to rescue these data, which are currently not fully used (Rossiter, 2008), and use the legacy data to provide precise quantitative soil information.

This study aims i) to assess the relationships between the soil attributes derived from a sparse legacy soil dataset and continuous environmental covariates, ii) to apply appropriate DSM functions to map soil attributes and iii) to compare the results with those obtained in other pedological contexts. The target properties and depth intervals were those specified in GlobalSoilMap.Net. The study examined pedological aspects of a tropical hillslope. Few previous studies have assessed the efficiency of digital soil mapping in this context.

2. Material and methods

2.1. Study area

The study area, in Bom Jardim County (Brazil), is located in the Tropical Atlantic Forest region of Rio de Janeiro State (Fig. 1). This 390 km^2 area includes a mountainous landscape and represents a vulnerable and fragile ecosystem that is subject to high annual rainfall, greater than 1200 mm/year, and has a very rugged topography associated with low-fertility soils occupied by small family farms. The regional soil pattern is complex, due primarily to a history of geological uplift that has mixed the area's lithologic features. In the legacy soil maps of the region (Calderano Filho et al., 2009), three principal soil classes (Oxisols, Inceptisols and Ultisols) are shown to be spatially distributed over complex soil map units.

2.2. Soil dataset

The soil dataset used in this study consists of 208 soil site samples collected between 2009 and 2011. This soil dataset consists of 74 soil profiles, 44 extra soil profiles and 90 A horizon samples. The extra soil profiles were sampled at discontinuous depths, and the A horizons were sampled only at the depth of the surface layer. The soil samples were described and analyzed following a set of specifications defined by Embrapa (1997) and Embrapa (2006). The results of analyses of several soil properties, namely, clay content (g/kg), silt (g/kg), sand (g/kg), soil organic carbon (SOC) (g/kg) and pH H_2O, were used in this study. The locations of the sampling points were specified in the UTM zone 23 south cartographic system.

Interpolation was necessary to obtain values for the soil properties at the fixed depth intervals required by the GlobalSoilMap.Net

Table 1
Decision rules to select the appropriate DSM function.
Adapted from Ciampalini et al. (2012).

		Correlation between paired samples?	
		NO (>0.02)	YES (≤0.02)
Correlation between distance matrices?	No (>0.10)	Means	Linear regression or RT[a]
	Yes (≤0.10)	Ordinary kriging	Regression kriging

[a] RT = regression tree.