MODELLLING AND DIGITAL SOIL MAPPING OF THE ORGANIC CARBON STOCK IN THE TOPSOIL (0-10 cm) OF RIO DE JANEIRO STATE, BRAZIL

M.L. Mendonça-Santos[1], H.G. Santos[1] and M.R.Coelho[1]

[1]Researcher at EMBRAPA Solos –Brazilian Agricultural Research Corporation / The National Centre of Soil Research. Rua Jardim Botânico 1024, 22.460-000, Rio de Janeiro, RJ, Brazil at EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária, Centro Nacional de Pesquisa de Solos. Rua Jardim Botânico 1024, 22.460-000, Rio de Janeiro, RJ, Brasil.
E-mail: loumendonca@cnps.embrapa.br

Abstract

A soil database with 431 soil profiles of Rio de Janeiro State was used in the scope of a research project entitled “Quantifying the magnitude, spatial distribution and organic carbon in soils of Rio de Janeiro State, using quantitative modeling, GIS and database technologies” (Projeto Carbono_RJ, funded by FAPERJ - Carlos Chagas Filho Foundation for Research Support in Rio de Janeiro State). Considering that these soil data were collected to other purpose, there was only a few sparse data to soil bulk density, which is essential to estimate of soil organic carbon (SOC) stock. To face this problem, pedotransfer functions (PTFs) were estimated to be used in the modeling of organic soil carbon of topsoil (0-10 cm), using *s.c.o.r.p.a.n* model. The following environmental correlates were used as predictor variables: satellite data, lithology and soil maps, DEM (Digital Elevation Model) and its derivatives as source of information for these variables. This dataset, that represents the best organized soil dataset in Brazil, is working as a trial for learning/teaching of Digital Soil Mapping (DSM) using a variety of methods for predicting soil classes and their properties. The "f" of the equation was modeled by means of multilinear analysis and regression-kriging. Seven different models were built and compared through statistical methods. In a general way, all models performed well to predict the SOC stock. Nevertheless, model 6 (M6) was an exceptional model, presenting the smallest AIC e RMSE, due to the use of existing soil information (polygon soil map) as predictor variable, in addition to the variables used in the other models. The result obtained in M6 was used for mapping topsoil carbon stock at spatial resolution of 90 m.

1 Introduction

Estimates of organic carbon stock in soils is an important aspect to be considered in relation to climatic changes and in relation to the last results of IPCC (Intergovernmental Panel on Climatic Change) about the worrying global climatic changes sceneries

According to Batjes & Sombroek (1997), the soils of the world constitute one of the five main reservoirs of carbon, together with the oceans, the lithosphere, the atmosphere and the terrestrial biomass. Therefore, the soils are essential for carbon sequestration representing approximately 75% of the carbon accumulation in the terrestrial ecosystem. The dynamics of the carbon sequestration by soils, depend on countless variables based on thermodynamic elements (nature and magnitude of carbon reservoir), and in the characteristics of biomes and on its responses to the different land uses and management systems (Batjes, 1998). The soil thus works s such, as source and reservoir (or sink) of carbon, depending on the relative rates of incorporation and decomposition of carbon by the action of soil organisms. In order to estimate the net flow of carbon in the terrestrial ecosystems is firstly necessary an understanding of the processes of soils formation and the spatial variability of organic carbon in the landscape. The spatial variability data is important to estimate stock of soil carbon and also understanding the biophysical processes that can affect the flow of organic carbon in soils. Besides, the patterns and processes vary considerably in the landscape, what limits extrapolations. Therefore, specific regionalized studies are important to assure a proper scale of study, as well as the rules for extrapolation of results and detailing of carbon dynamics in soils.

For prediction and digital mapping of the carbon stock in a landscape scale was used the digital quantitative techniques, named as Digital Soil Mapping – DSM, defined by Lagacherie and McBratney (2007) as "*the creation and population of spatial soil information systems by numerical models inferring the spatial and temporal variations of soil types and soil properties from soil observation and knowledge and from related environmental variables*". The main use of the this approach is to replace the polygon-based soil maps of the past with digital maps of soil classes and properties and their associated uncertainties for areas previously mapped, or for new areas. These maps are stored and manipulated in digital form in a GIS environment, creating the possibility of vast arrays of data for analysis and interpretation at any time.

Predictions of soil classes and properties in the digital mapping are based on relationships among soils and the factors and processes of soil formation that enter in the equations as predictor variables. The logic of this reasoning is based on the equation of Jenny (1941) formulated from the recognition of the factors of soil formation, in a more quantitative formulation,

$$S = f(cl, o, r, p, t)$$

Where, S represents the soil, cl = climate, o = organisms, including anthropic activities, r = relief, p = parent material and t = time.

McBratney et al., (2003) generalized and formulated a similar equation, with the objective of explaining the responsible variables for the processes of soil formation, through an empiric quantitative description of the relationships among other factors spatially geo-referenced (environmental co-variables), used here as prediction spatial functions. Seven factors are considered: s = soil and other properties of the soil in a certain point; c = climate, climatic properties of the atmosphere in certain point; o = organisms, vegetation or fauna or anthropic activities; r = topography, attributes of the landscape; p = parent material, lithology; a = age, time factor; n = space, spatial location.

Each factor will be represented by a group of one or more continuous or categorical variables, for example, **r** for elevation, slope or other derived attribute of a DEM. The sources of data, the methods to estimate **f**, as well as the steps to execute the scorpan are presented and discussed in McBratney et al. (2003).

In this work, the procedures of digital mapping were used to predict the stock of organic carbon of the topsoil (0-10cm) in the State of Rio de Janeiro. For that purpose, a multilinear analysis was used as predictive model and some environmental variables as predictors. Seven different models were built and statistically compared. The best model was applied to the digital mapping of the soil carbon stock

2 Materials and Methods

2.1 Study Area

The study area is the whole State of Rio de Janeiro, located between the geographical coordinates 41^o and 45^o W and 20°30´ and 23°30´S with about 44.000 km2 (Figure 1) comprehending 89 topographic sheets of IBGE in the scale 1:50.000. The area is characterized by eight large landscape types described as Coastal Plains, North-Northwest Fluminense, Rio Paraíba of South Middle Valley, Mountainous Área, Upper Itabapoana River Plateau, Serra dos Orgãos, Bocaina and Mantiqueira described in Mendonça-Santos et al., (2008), constituting the so called "*geoenvironments* " where soil profiles have been studied in order to characterizes the soil organic matter.

2.2 Digital and Field Data

The same soil database that have been used to estimate soil classes in Rio de Janeiro State by Mendonça-Santos et al., (2008), was used in this work, in which we have added 16 soil profiles described to cope the objectives of the RJ_Soil Carbon Project (Mendonça-Santos *et al*., 2005), summing with 431 soil profiles.

The spatial distribution of soil profiles is illustrated in Figure 1. Considering that these soil data were collected to other purpose, there was only a few sparse data to soil bulk density, which is essential to estimate of soil carbon stock. To cope this problem, pedotransfer functions (PTFs) for the mineral upland and lowland soils were performed as auxiliary information for soil organic carbon estimation.

In this application the following covariates were used as predictor variables to build the spatial soil organic carbon models: Geocover™ mosaic (bands 7, 4 and 2 in RGB), freely distributed by NASA (https://zulu.ssc.nasa.gov/mrsid/); the NDVI index (using band 2 instead of 3); Land Use/Land Cover (LULC) map of Rio de Janeiro State, produced by Mendonça-Santos *et al.*, (2003); the Lithology class map (Rio de Janeiro, 2001) and SRTM DEM 90m, obtained from the CGIAR database at http://srtm.csi.cgiar.org (Jarvis *et al.*, 2006) and modified by Mendonça-Santos et al., (2008) and its derivates extracted using the LandMapR software (McMillan, 2003), were used in the building of the predictive models.

The soil dataset was complemented with the covariates of environmental factors for each soil data point. An ancillary dataset representing the whole study area was interpolated onto a 90-m grid corresponding to the SRTM DEM, and populated with environmental and soil variables. Exploratory statistical analysis was performed on soil data. The modelling and prediction of soil carbon was performed using multilinear regression and regression-kriging. The output results were imported and mapped in a GIS environment.

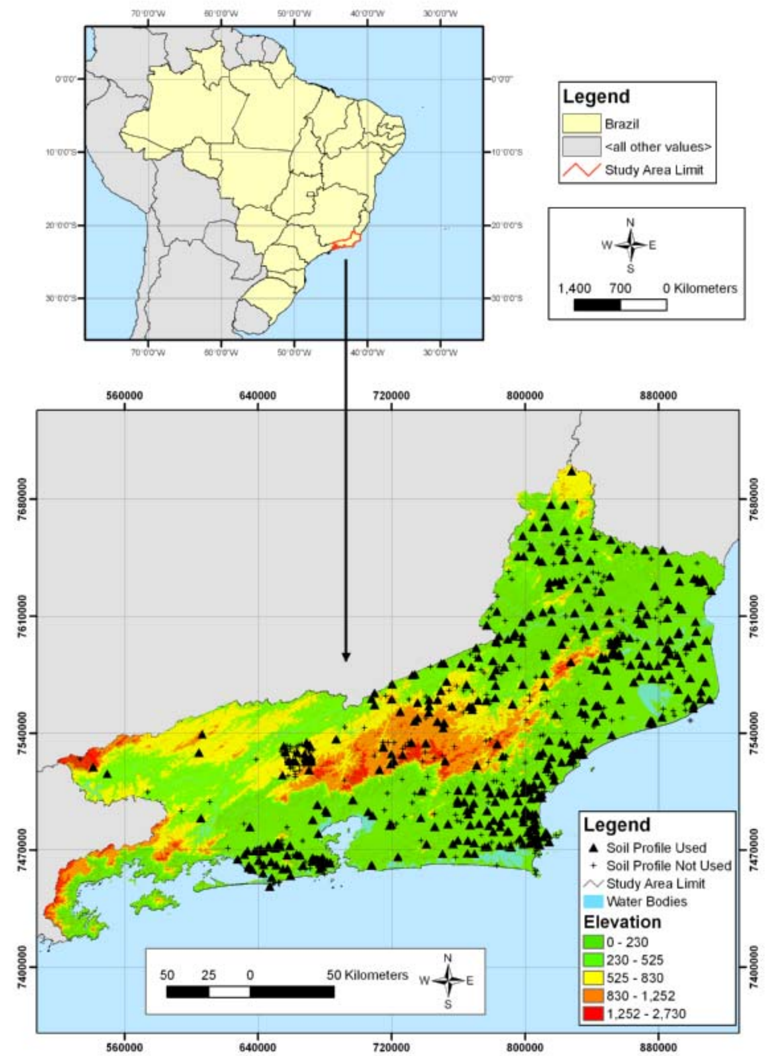

Figure 1. The study area location and the soil profile distribution on the elevation map, extracted from the SRTM DEM (Jarvis et al., 2006) at 90 m pixel resolution.

2.3 Inference Models

The soil organic carbon stock was calculated in mass per unit of volume, through the equation: $StockC = C \times d \times p$, where:

C is the content of carbon (g/kg), d is the soil bulk density (g/cm3) and p is the depth (cm).

Seven models of analysis multilinear were elaborated. The models were differentiated by the number of predictor variables used, the use or not of stepwise and number of observations (profiles) used in the adjustment of each model. The performance of the models was estimated statistically, using besides RMSE (estimate of the standard deviation of the residual mistake), the AIC (Akaike's Information Criterion), that is an index that considers the number of parameters used in the model representing a commitment between the adjustment and the parsimony of the model. The model that presents the smallest AIC will be the best. AIC is calculated in agreement with (Akaike, 1973):

$$AIC = -2\log like + 2m,$$

where loglike is the logarithm of the prediction, and **m** it is the number of parameters used in the model.

For continuous variables, as in the case of soil carbon stock is given that:

$$\mathrm{AIC} = N\ln\left(\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2\right) + 2m$$

Where **N** is the total number of soil profiles that were used in the model.

The model that presented smaller value of AIC was used for the final prediction and of the organic carbon stock mapping (0-10 cm) of the soil, being added to the predicted values, the krigged residues of the model adjustment (method regression-krigging).

3 Results and Discussion

3.1 Results from the Models

Developed PTFs, together with the predictor variables, are shown in Table 1. Given the outstanding difference between carbon contents and soil texture of lowladsoils and the others soils (here denominated mineral soils), it was necessary to build 2 PTFs equations, separately. Those PTFs was applied for the estimate of the soil density and later on, for the calculation of the stock of organic carbon.

Table 1 - Pedotransfer Functions (PTFs) for soil bulk density, estimated from soil organic carbon content (%) and soil particles sizes (sand, silt and clay - %).

Layers	Predictors	Number of Parameters	R^2	N	Average	Standard deviation	Max	Min	PTF_bulk density
0-10 Lowland soils	C content, Fine sand, Thick sand, Silt, Clay	5	0.94671	131	1.09851	0.2023249	1.5373	0.6569	0 + -0.00620188898666584 * ("C cont_0-10") + 0.00168408543309972 * ("Thick sand 0-10") + 0.00112348306614799 * ("Fine sand 0-10") + 0.00155733827838204 * ("Silt 0-10") + 0.00105136516518877 * ("Clay 0-10")
0-10 Upland soils (minerals)	C content, Fine sand, Thick sand, Silt, Clay	5	0.3712	594	1.32962	0.1124466	1.71	0.7759	0 + -0.0104432003873496 * ("C content_0-10") + 0.00148435216740785 * ("Thick sand 0-10") + 0.00155020523466563 * ("Fine sand 0-10") + 0.00165886998238256 * ("Silt 0-10") + 0.00133985448863544 * ("Clay 0-10")

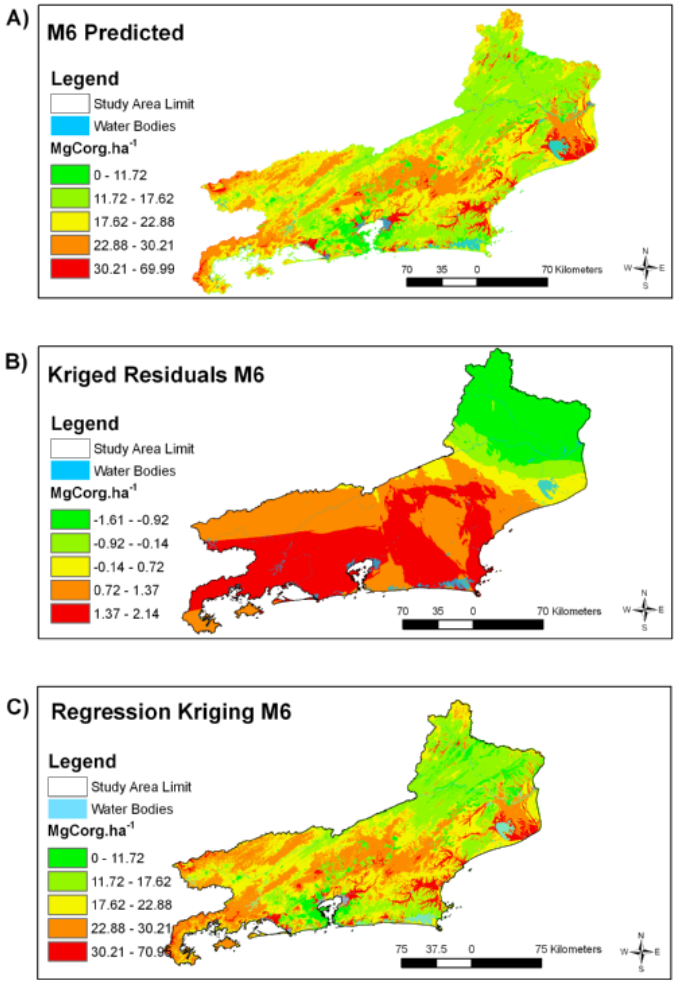

Figure 2. Digital soil map of organic carbon (0-10 cm) of Rio de Janeiro State. A) result of the predictive *s.c.o.r.p.a.n.* soil-landscape modelling (multilinear regression); B) kriging of the modelling residues; C) final result obtained by the sum of the krigged residues with the values predicted by the multilinear regression (regression krigging).

The soil-landscape model s.c.o.r.p.a.n. was then accomplished, using the soil and landscape information. In Table 2 the built models are given with details. The model M1 encompasses the extracted relief variables in LandMapR. In model M2 the same relief derivatives were used, but in this case, a stepwise procedure was undertaken in order to find which variables have larger correlation with the soil organic carbon stock. In the models M3 and M4 the relief variables and the Geocover were used (Landsat 7 ETM+ with the bands 7-4-2, NDVI), with the difference that in model M4 was undertaken a stepwise procedure. The stepwise in the model M4 did not allow the entrance of NDVI in the model. The variable NDVI excluded two profiles out of the model (in these two profiles the reflectance in the bands 4 and 2 was 0). In the model M5, besides the terrain variables and the Geocover, the lithology map was also used. The model M6 encompasses all the variables of the model M5 and also, an existing polygon soil map. In the last model, M7 uses the variables of the model M5, in addition to the LULC map.

Table 2. Predictive models s.c.o.r.p.a.n. built to estimate the soil carbon stock in topsoil (0-10cm).

Models	Predictors Variables – *SCORPAN model*	*Stepwise*	Number of Soil Profiles
M1	***R*** (ELEV, ASPECT, PLAN, PROF, QWETI, SLOPE)	____________	429
M2	***R*** (ELEV, ASPECT, PLAN, PROF, QWETI, SLOPE)	ELEV, ASPECT, PLAN, QWETI, SLOPE	429
M3	***O*** (Landsat ETM$^+$ -B7, B4, B2 e NDVI), ***R*** (ELEV, ASPECT, PLAN, PROF, QWETI, SLOPE)	all	427
M4	***O*** (Landsat ETM$^+$-B7, B4, B2 e NDVI), ***R*** (ELEV, ASPECT, PLAN, PROF, QWETI, SLOPE)	B7, B4, ELEV, ASPECT, PLAN, QWETI, SLOPE	429
M5	***O*** (Landsat ETM$^+$-B7, B4, B2 e NDVI), ***R*** (ELEV, ASPECT, PLAN, PROF, QWETI, SLOPE), ***P*** (Litology Map – vector format)	all	427
M6	***S*** (Soil Map - polygon), ***O*** (Landsat ETM$^+$-B7, B4, B2 e NDVI), ***R*** (ELEV, ASPECT, PLAN, PROF, QWETI, SLOPE), ***P*** (Litology Map – vector format)	all	427
M7	***O*** (Landsat ETM$^+$-B7, B4, B2, NDVI and LULC Map), ***R*** (ELEV, ASPECT, PLAN, PROF, QWETI, SLOPE), ***P*** (Litology Map – vector format)	all	427

3.2 Accuracy Assessment of Models

The result of the carbon stock prediction, the performance of the indexes AIC and RMSE (estimate of the standard deviation of the residual error) and the number of parameters (variables) used in each tested model are illustrated can in the Table 3.

Generally, all the seven models presented good predictions for the carbon stock. Considering that the difference among the indexes of the seven models did not present a significant variation.
The best result found for modeling of carbon stock was the model 6 (M6), by the smaller indexes AIC and RMSE (Table 3).

Table 3. Comparison of the performance of the models.

Models	RMSE	Number of arameters	AIC
M1	14,26907	6	2286,603
M2	14,25226	5	2284,604
M3	14,15396	10	2279,976
M4	14,08333	7	2267,386
M5	13,09959	19	2215,548
M6	**11,95091**	**29**	**2146,578**
M7	12,68869	28	2196,807

The Figure 3 illustrates the result of the best model (M6). That map, obtained by modeling and digital mapping, allows the modeled property be spatially viewed in a continuous way in the grid determined by the availability of data and objectives of the work, facilitating to observe the variation and distribution of the stock of carbon in the landscape.

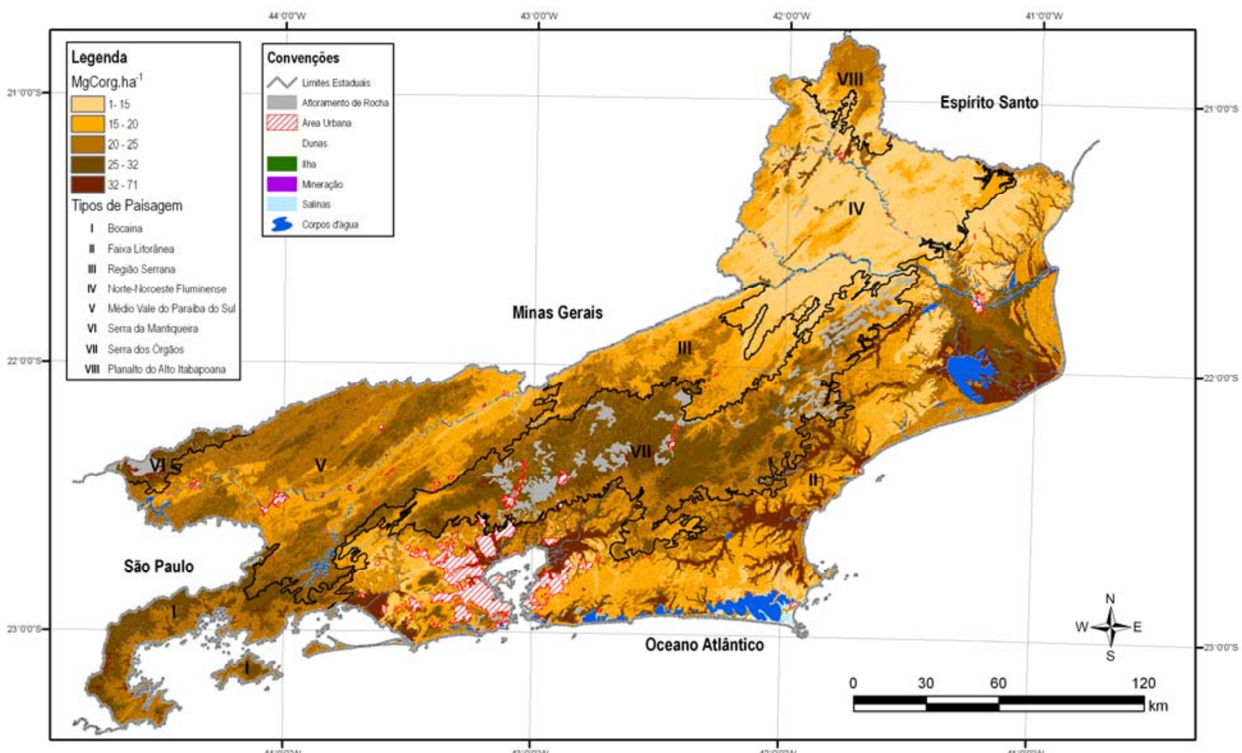

Figure 3. The final map of the soil organic carbon of Rio de Janeiro State (0-10 cm) with environmental units of the state of Rio de Janeiro (Rio de Janeiro, 2001).

It is observed different values for the organic carbon stock, varying from less than 3 to 70 Mg of organic carbon per hectare, for the surface layer of the soil (0-10 cm). That variation has a strong correlation with the soil type and with its position in the landscape. That correlation with the landscape was clearer when analyzed the in relation to the geo-environments defined for Rio de Janeiro (Lumbreras et al. 2003) (Figue 4). In that map is possible to observe that the unit II (Faixa Litorânea) presents larger stocks of organic carbon, due to the lowland environments, such as, mangroves (surroudigs of Guanabara Bay, Sepetiba Bay, Guaratiba, among others), rivers, lakes (Lagoa Feia, Lagoa de Maricá among other) and close to the coast and its corresponding soils. In opposition, the unit IV (North-northwest Fluminense) it presents the smallest stocks of organic carbon for the studied depth.

4 Conclusions

Through this specific application of prediction soil organic carbon stock in soils, the whole process of DSM is demonstrated, using the soil formation factors as predictor variables for construction of the models. The work was adequate to test the methodology of carbon stock prediction in the soil at the depth 0 to 10 cm of the surface. Seven predictive models were tested. In general, the seven tested models have shown efficient and did not present great variations. However, the best result for carbon stock was obtained with application of the model 6 (M6), that presented the smallest indexes AIC and RMSE (Table 3). This model encompasses information on existing soil map, satellite images DEM and its derivates and lithology map.

The spatial distribution of soil organic carbon has relationships with the different geo-environments of the study area, e.g. the highest stocks of organic carbon in the lowlands areas. A detailed study of the behavior and dynamic of soil carbon into the soil profiles studies in each one of these geo-environments is being undertaken and will be soon published.

Besides, the definition of the baseline of organic carbon stock in soils is very important for the definition of public policies of maintainable agricultural systems and environmental protection, working as scenery of the soil potential as a carbon sink.

Acknowledgements

To the Foundation Carlos Chagas for Research Support (FAPERJ) for the financial support to the Thematic Project E-26/171.360/2001 (EMBRAPA SEP 01.2002.202), entitled " Modelling of the magnitude and spatial distribution of the Organic Carbon in the State of Rio de Janeiro, using quantitative techniques, SIG and Base of Data." Our thanks also to EMBRAPA, for supporting the Research Project E-170.023/2001, registered as SEP/12.2002.001, entitled “Application of digital quantitative techniques to optimize soil mapping techniques for planning purposes and environmental administration”, in the scope that this work was developed.

5 References

Akaike, H. 1973. Information theory and an extension of maximum likelihood principle. *In*: B.N. Petrov and F. Csaki (Eds.), Second International Symposium on Information Theory. Akademia, Kiado, Budapest, pp. 267–281.

BajesS, N.H. and Sombroeck, W.G. Possibilities for carbon sequestration in tropical and subtropical soils. Global Change Biology, 3, 161-173. 1997.

Batjes, N.H. Mitigation of atmospheric CO2 concentrations by increased carbon sequestration in the soil. Biology and Fertility of Soils, 27, 230-235. 1998.

Jenny, H., 1941. Factors of soil formation. McGraw-Hill, New York.

Jarvis, A., Reuter, A., Nelson, E., Guevara, E. Hole-filled SRTM for the globe Version 3. Available from the CGIAR-CSI SRTM 90m. Homepage: http://srtm.csi.cgiar.org. 2006.

Lagacherie,P., McBratney, A.B. 2007. Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping. pp. 3-22. *In* (Lagacherie, P., McBratney, A.B. and Voltz, M. (Eds.) Digital soil mapping: an introductory perspective. Developments in soil science vol. 31. Elsevier, Amsterdam, 600 p.

Macmillan, R. A. LandMapR© Software Toolkit- C++ Version: User manual. LandMapper Environmental Solutions Inc., Edmonton, AB. 110 pp. 2003.

McBratney, A.B., Mendonça Santos, M.L., and Minasny, B. 2003. On digital soil mapping. Geoderma 117:3-52.

Mendonça-Santos, M. L.; Jacques, P.D.; Coelho, M.R.; Pimentel, J.; Santos, H.G. dos; Almeida, P.; Barbosa, E.L.de A.; Costa, T.C. e C. da; Ávila, R.M.; Shinzato, E.; Branco, P.C.M.P. de A. Mapeamento do Uso Atual e Cobertura Vegetal dos Solos do Estado do Rio de Janeiro. Rio de Janeiro: (Embrapa Solos. Boletim de Pesquisa e Desenvolvimento n. 22). 44 p. 2003.

Mendonça-Santos, M. L.; Santos, H. G. dos; Coelho, M. R.; Pares, J. G. Caracterização de paisagens e solos representativos do Estado do Rio de Janeiro para fins de estimativa de estoques de carbono no solo. Rio de Janeiro: (Embrapa Solos. Boletim de Pesquisa e Desenvolvimento n. 66). 80p. 2005.

Mendonça-Santos, M.L., H.G. Santos, R.O. Dart and J.G.Pares. 2008. Digital Mapping of Soil Classes in Rio de Janeiro State, Brazil: Data, Modelling and Prediction. P. 381-396. *In*: A.E. Hartemink, A. McBratney and M.L. Mendonça-Santos (eds.). Digital Soil Mapping with Limited Data. Springer, Amsterdam.

NASA (National Aeronautics and Space Administration). 2008. The Landsat program. http://landsat.gsfc.nasa.gov/about/ (last verified 25 Sep 2008).