

**UNIVERSIDADE DE SÃO PAULO  
CENTRO DE ENERGIA NUCLEAR NA AGRICULTURA**

**OSVALDO JOSÉ RIBEIRO PEREIRA**

**Mapping soil organic carbon storage in deep soil horizons of  
Amazonian Podzols**

**Piracicaba**

**2015**



**OSVALDO JOSÉ RIBEIRO PEREIRA**

**Mapping soil organic carbon storage in deep soil horizons of  
Amazonian Podzols**

**Versão revisada de acordo com a Resolução CoPGr 6018 de 2011**

**Tese apresentada ao Centro de Energia  
Nuclear na Agricultura da Universidade de  
São Paulo para a obtenção do título de Doutor  
em Ciências**

**Área de Concentração: Química na  
Agricultura e no Ambiente**

**Orientador: Profa. Dra. Célia Regina Montes**

**Piracicaba**

**2015**

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Dados Internacionais de Catalogação na Publicação (CIP)  
**Seção Técnica de Biblioteca - CENA/USP**

Pereira, Osvaldo José Ribeiro.

Mapeamento do estoque de carbono orgânico em horizontes profundos de Espodossolos da Amazônia; Mapping soil organic carbon storage in deep soil horizons of Amazonian Podzols; orientadora Célia Regina Montes. - - versão revisada de acordo com a Resolução CoPGr 6018 de 2011. - - Piracicaba, 2015.

129 p. : il.

Tese (Doutorado – Programa de Pós-Graduação em Ciências. Área de Concentração: Química na Agricultura e no Ambiente) – Centro de Energia Nuclear na Agricultura da Universidade de São Paulo.

1. Carbono 2. Espodossolos – Rio Negro 3. Matéria orgânica do solo  
4. Monitoramento ambiental 5. Mudança climática 6. Sensoriamento remoto  
I. Título

CDU 631.417.1 : 528.855 (811.3)

## **DEDICATION**

I dedicate my thesis to my loving parents, Rosa Maria Ribeiro Pereira and Oliveira José Pereira, whose words of encouragement have helped on the conclusion of this work.

To my sisters Elaine and Gisele.



## ACKNOWLEDGEMENTS

First, I would like to thank the São Paulo Research Foundation (FAPESP) for the financial support of the present work (Process number 2012/12882-5). I could not conclude this research without the grants provided by FAPESP.

I thank the Brazilian Coordination for the Improvement of Higher Level Personnel (CAPES) for the financial support of part of this research developed at the University of Toulon (France).

I would like to express my sincere gratitude to my formal advisor Prof. Célia Regina Montes (CENA/NUPEGEL - USP), for her continuous support of my Ph.D study and the sharing of important knowledge concerning the Amazonian region.

My sincere thanks also goes to Prof. Yves Lucas (PROTEE - Université de Toulon), who provided me the opportunity to establish some of the methods presented in this work and for sharing his vast knowledge about the Podzols of Rio Negro basin.

I thank the Prof. Adolpho José Melfi (IEE/ESALQ/NUPEGEL - USP) for his constant collaboration, his advices and insights about mineralogy and soil geochemistry.

My thanks also goes to the Professors Nádia Regina do Nascimento (IGCE - UNESP, Rio Claro) and Guilherme Taitson Bueno (UFG/GO) whose have shared a series of concepts about Amazonian Podzols and the Ferralsol/Podzol soil system.

I am also grateful to my dear friend Débora Ishida, who always supported me on my research and has contributed on the laboratorial analysis of soil samples.

I thank my dear friend Prof. Teresa Cristina Tarlé Pissarra (Department of Rural Engineering - UNESP), who was supportive and helped on enriching my knowledge about cartography and remote sensing of natural environments.

I take this opportunity to express gratitude to all of the NUPEGEL laboratory, for their help and support.

I place on record, my sincere thanks to all of the PROTEE (Toulon/France) laboratory, whose have supported the development of the first part of this research.

I am also grateful to the CENA's employees, whose were essential for the conclusion of this research.

Finally, I also place on record, my sense of gratitude to one and all, who directly or indirectly, have contributed to the conclusion of my PhD thesis.



## ABSTRACT

OSVALDO, J. R. P. **Mapping soil organic carbon storage in deep soil horizons of Amazonian Podzols**. 2015. 129 p. Tese (Doutorado) - Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Piracicaba, 2011.

The Podzols of the world are divided into intra-zonal and zonal according to their location. Zonal Podzols are typical for boreal and taiga zone associated to climate conditions. Intra-zonal podzols are not necessarily limited by climate and are typical for mineral poor substrates. The Intra-zonal Podzols of the Brazilian Amazon cover important surfaces of the upper Amazon basin. Their formation is attributed to perched groundwater associated to organic matter and metals accumulations in reducing/acidic environments. Podzols have a great capacity of storing important amounts of soil organic carbon in deep thick spodic horizons (Bh), in soil depths ranging from 1.5 to 5m. Previous research concerning the soil carbon stock in Amazon soils have not taken into account the deep carbon stock (below 1 m soil depth) of Podzols. Given this, the main goal of this research was to quantify and to map the soil organic carbon stock in the region of Rio Negro basin, considering the carbon stored in the first soil meter as well as the carbon stored in deep soil horizons up to 3m. The amount of soil organic carbon stored in soils of Rio Negro basin was evaluated in different map scales, from local surveys, to the scale of the basin. High spatial and spectral resolution remote sensing images were necessary in order to map the soil types of the studied areas and to estimate the soil carbon stock in local and regional scale. Therefore, a multi-sensor analysis was applied with the aim of generating a series of biophysical attributes that can be indirectly related to lateral variation of soil types. The soil organic carbon stock was also estimated for the area of the Brazilian portion of the Rio Negro basin, based on geostatistical analysis (multiple regression kriging), remote sensing images and legacy data. We observed that Podzols store an average carbon stock of  $18 \text{ kg C m}^{-2}$  on the first soil meter. Similar amount was observed in adjacent soils (mainly Ferralsols and Acrisols) with an average carbon stock of  $15 \text{ kg C m}^{-2}$ . However if we take into account a 3 m soil depth, the amount of carbon stored in Podzols is significantly higher with values ranging from  $55 \text{ kg C m}^{-2}$  to  $82 \text{ kg C m}^{-2}$ , which is higher than the one stored in adjacent soils ( $18 \text{ kg C m}^{-2}$  to  $25 \text{ kg C m}^{-2}$ ). Given this, the amount of carbon stored in deep soil horizons of Podzols should be considered as an important carbon reservoir, face a scenario of global climate change.

**Keywords:** Rio Negro basin. Amazon. Soil carbon stock. Remote sensing.



## RESUMO

OSVALDO, J. R. P. **Mapeamento do estoque de carbono orgânico em horizontes profundos de Espodosolos da Amazônia**. 2015. 129 p. Tese (Doutorado) - Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Piracicaba, 2011.

Os Espodosolos podem ser divididos em zonais e intrazonais de acordo com área onde ocorrem. Os Espodosolos zonais são típicos de áreas boreais e taiga, delimitados por condições climáticas. Já os intrazonais não são condicionados pelo clima. Os Espodosolo intrazonais brasileiros ocupam uma grande extensão da alta bacia amazônica, tendo sua formação atribuída à ocorrência de lençóis freáticos suspensos associados à acumulação de complexos organometálicos em ambientes ácidos redutores. Esses solos tem a capacidade de estocar grandes quantidades de carbono orgânico em horizontes espódicos profundos (Bh), em profundidades que podem variar de 1,5m a 5m. Pesquisas atuais relacionadas ao estoque de carbono em solos amazônicos, não levam em consideração os estoques encontrados no horizonte Bh (abaixo de 1m de profundidade). Sendo assim, o principal objetivo da presente pesquisa foi quantificar e mapear o estoque de carbono nos solos da bacia do Rio Negro, tendo-se em vista aquele estocado no primeiro metro de solo, bem como o carbono armazenado em até 3m de profundidade. A quantidade de carbono orgânico estocado nos solos da bacia do Rio Negro foi estimada em diferentes escalas de mapeamento, desde mapas locais até a escala da bacia do Rio Negro. Imagens de sensoriamento remoto de alta resolução espacial e espectral foram essenciais para viabilizar o mapeamento dos solos nas áreas estudadas e permitir a estimativa do estoque de carbono. Uma análise multisensor foi adotada buscando-se gerar informações biofísicas indiretamente associadas à variação lateral dos tipos de solo. Após o mapeamento do estoque de carbono em escala regional, partiu-se para a estimativa na escala da bacia do Rio Negro, com base em análise geoestatística (krigagem por regressão linear), imagens de sensoriamento remoto e base de dados de domínio público. Após o mapeamento do estoque de carbono na escala da bacia, constatou-se que os Espodosolos têm um estoque médio de  $18 \text{ kg C m}^{-2}$ , para 1m de profundidade, valor similar ao observado em solos adjacentes (Latosolos e Argissolos) os quais tem um estoque de  $15 \text{ kg C m}^{-2}$ . Quando são considerados os estoques profundos, até 3m, a quantidade de carbono dos Espodosolos é superior com valores variando de  $55 \text{ kg C m}^{-2}$  a  $82 \text{ kg C m}^{-2}$ . Estoque relativamente maior que aquele observado em solos adjacentes para esta profundidade ( $18 \text{ kg C m}^{-2}$  a  $25 \text{ kg C m}^{-2}$ ). Portanto, o estoque de carbono profundo dos Espodosolos, não deve ser negligenciado levando-se em conta cenários futuros de mudanças climáticas.

**Palavras-chave:** Bacia do Rio Negro. Amazonas. Carbono orgânico do solo. Sensoriamento remoto.



## LIST OF FIGURES

Figure 2.1 - Studied areas. (a): Amazon Forest site; (b) Zambebian Flooded Grasslands (Zambia site). Multispectral composition (reference image) for both regions (Synthetic OLI bands: Blue/Green; Red/Near Infrared and SWIR 1/SWIR 2 - RGB).....	32
Figure 2.2 - Flowchart of the methodology applied to evaluate the fusion algorithms by a zonal approach and by unsupervised classification of fused compositions. ....	35
Figure 2.3 - Flowchart illustrating the Ehlers pansharpener procedure (EHLERS et al., 2006).....	37
Figure 2.4 - Box-plots with jitters. Dark gray jitters (Landsat/HRC); Light gray jitters (Landsat/OLI). The “out of range” boxes represent Br. algorithm with high ERGAS values. ....	45
Figure 2.5 - Overall quality of the fusion methods ATWT, Eh., DWT, GS and PC grouped by clusters. The MS compositions illustrated the ATWT fused images: (a) Amazon – Landsat/HRC; (b) Amazon – Landsat/OLI; (c) Zambia – Landsat/HRC and (d) Zambia – Landsat/OLI. A representation of the zonal sample windows is shown in Figure 2.5a, as light gray lines.....	47
Figure 3.1 - Situation of the study area. The sampled area represents the region where the soil samples were collected. The map illustrates the extrapolation area (multi-sensor composition: Land Surface Temperature, SAVI, and NDMI – R, G, B, respectively).....	58
Figure 3.2 - Flow chart showing the methodology employed in this work for generating the SOC regional map. SAVI, soil adjusted vegetation index; NDMI, normalized difference moisture index. .	59
Figure 3.3 - Scatterplot showing the relation between soil adjusted vegetation index (SAVI) and the following variables: (a) land surface temperature (LST); (b) altitude; and (c) normalized difference moisture index (NDMI). (d) Projection of the normalized factor coordinates of variables (biophysical variables) in the $1 \times 2$ factor plane obtained by principal component analysis. Group 1: seasonally flooded and overflowed Podzols. Group 2: poorly drained Podzols. Group 3: well-drained Podzols and Ferralsols. ....	65
Figure 3.4 - Producer’s and user’s accuracies for the SVM classification of multi-sensor and OLI Landsat 8 compositions. The designation of each class is shown in Table 3.1.....	68
Figure 3.5 - Average carbon stock for the three main Podzol groups described in the study area.....	70
Figure 3.6 - Soil map illustrating the spatial distribution of the soil types. The numbers for the soil units represent the cluster groups described in Figure 3.3 .....	73
Figure 4.1 - Situation of the studied area highlighting the location of the soil sample profile.....	80
Figure 4.2 - Plot of the predicted data against the observed data. (a): Model 1; (b): Model 2. Dashed lines are the 1:1 lines.....	87
Figure 4.3 - Scatterplots and goodness of fit indexes of proposed and previous soil bulk density PTF functions. (a): Proposed Model 2; (b): Benites et al. (2007); (c) Bernoux et al. (1998) and (d): Tomasella; Hodnett (1998). Dashed lines are the 1:1 lines.....	88
Figure 4.4 - (a) Observed and fitted exponential depth function SOC; (b) Observed and predicted exponential depth function, based on the validation dataset. Dashed line is the 1:1 line.....	89
Figure 4.5 - Measured SOC stock. (a): Dataset A (Ferralsols and Acrisols); (b): Dataset B (Podzols). <sup>a</sup> Typical Ferralsol horizons. <sup>b</sup> Typical Podzol horizons with average thickness for evaluated Podzol profiles.....	90
Figure 4.6 - Example of fitting models to a typical Podzol profile. (a): Equal-area Spline; (b) Sum of Sines; (c): Fourier.....	91

Figure 5.1 - Map of the studied area, showing the major soil orders of Rio Negro basin in the original map scale of 1:250,000 (IBGE, 2008). The legacy data (Dataset 1) was provided by IBGE (2008) and Embrapa (2014). Field sample data (Dataset 2) represents the samples obtained in the frame of this research in Podzol regions.....	100
Figure 5.2 - Flowchart of the overall SOC stock prediction method.....	105
Figure 5.3 - Histograms of soil organic carbon (SOC) stock for Rio Negro basin: (a) measured SOC data at 1m soil depth; (b) measured SOC data at 3m soil depth; (c) logarithmically transformed (LnSOC) SOC data at 1m soil depth; (d) logarithmically transformed (LnSOC) SOC data at 3m soil depth.....	107
Figure 5.4 - Normal QQPlot of soil organic carbon (SOC) stock for Rio Negro basin: (a) measured SOC data at 1m soil depth; (b) measured SOC data at 3m soil depth; (c) logarithmically transformed (LnSOC) SOC data at 1m soil depth; (d) logarithmically transformed (LnSOC) SOC data at 3m soil depth.....	108
Figure 5.5 - Experimental and modelled variograms of soil carbon stock at 1m (a) and 3m (b) soil depth for OK.....	109
Figure 5.6 - Standardized coefficients chart to 1m soil depth, highlighting the most significant classes.....	112
Figure 5.7 - Standardized coefficients chart to 3m soil depth, highlighting the most significant classes.....	112
Figure 5.8 - Experimental and modelled variograms of soil carbon stock at 1m (a) and 3m (b) soil depth for RK.....	112
Figure 5.9 - Measured against predicted values of SOC stock: (a) Regressed surface to 1m soil depth; (b) regressed surface at 3m soil depth; (c) RK at 1m soil depth; (d) RK at 3m soil depth. The values were converted back to SOC stock in $\text{kg C m}^{-2}$ (exponential of Log-SOC).....	115
Figure 5.10 - Predicted SOC stock map according to RK procedure at 1m soil depth.....	116
Figure 5.11 - Predicted SOC stock map according to RK procedure at 3m soil depth.....	117
Figure 5.12 - SOC Stock map obtained by subtracting the 3m soil depth map (Figure 5.11) from the 1m soil depth map (Figure 5.10).....	117

## LIST OF TABLES

Table 2.1 - Summary of remotely sensed data.....	33
Table 2.2 -Band by band correlation values between the original multispectral image and the degraded fused images.....	43
Table 2.3 - Descriptive statistics (arithmetic mean, variance and standard deviation) for the fused images according to the four quantitative evaluation algorithms.....	44
Table 2.4 - Overall accuracy (O.A. in %) and kappa coefficient for the classified images according to different fusion methods.....	48
Table 3.1 - Producer and user's accuracy (PA and UA, respectively) for ISODATA clustering according to the field-truth (ROI). The classes of water bodies and bare soils are not shown. ....	67
Table 3.2 - Confusion matrix of the multisensor classified image, representing the classes' similarity.....	69
Table 3.3 - Average Carbon Stock for Podzols.....	71
Table 3.4 - Total carbon stock according to each soil unit. The stock is represented in Teragram ( $10^{12}$ grams) and the area in hectares. ....	71
Table 4.1 - Descriptive statistics of the soil attributes of the training and validation data (Datasets 1 and 2).....	82
Table 4.2 - Evaluation indices for the three fitting models considering Dataset B (Podzols).....	91
Table 5.1 - List of ancillary data used to predict the distribution of soil organic carbon stock.....	102
Table 5.2 - Prediction error parameters for Ordinary Kriging. The number of samples refers to the training dataset (85%). MSE and RMSE values are expressed in $\text{kg C m}^{-2}$ .....	110
Table 5.3 - Summary of the SMRL variables selection at 1m soil depth (RK).....	111
Table 5.4 - Summary of the SMRL variables selection at 3m soil depth (RK).....	111
Table 5.5 - Model performance to predict soil carbon stock ( $\text{kg C m}^{-2}$ ) based on validation dataset.....	117



## LIST OF ABBREVIATIONS

AGB	Above Ground Biomass
AIC	Akaike's Information Criterion
ATWT	À Trous Wavelet Transform
Br.	Brovey
CAST	China Academy of Space Technology
CBERS	China–Brazil Earth Resources Satellite
CC	Correlation Coefficient
CREN	Natural Resources and Environmental Studies Division (IBGE)
DEM	Digital Elevation Model
DN	Digital Number
DOM	Dissolved Organic Matter
DSM	Digital Soil Mapping
DWT	Discrete Wavelet Transform
Eh	Ehlers
EMBRAPA	Brazilian Company of Farming Research
ERGAS	Erreur Relative Globale Adimensionnelle de Synthèse
ETM+	Enhanced Thematic Mapper plus
FFT	Fast Fourier Transformation
FLAASH	Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes
G.S.	Gram Schmidt
GIS	Geographic Information System
GLCF	Global Land Cover Facility
HPF	High Pass Filter
HRF	High, Dense Rainforest
IBGE	Brazilian Institute of Geography and Statistics
IHS	Intensity Hue and Saturation
INPE	Brazilian Institute for Space Research
IPCC	Intergovernmental Panel on Climate Change
Kc	Kappa Coefficient
LPF	Low Pass Filter
LSE	Land Surface Emissivity
MODTRAN	5S Physically-based Calibration Model of Atmosphere Transference
MS	Multispectral
MSE	Mean Squared Error
NASA	National Space Agency
NDMI	Normalized Difference Moisture Index.
NIR	Near Infrared
O.A.	Overall Accuracy
OK	Ordinary Kriging
OLI	Operational Land Image
PA	Producer's Accuracy
PC	Principal Components
PTF	Pedotransfer Functions
QQ	Quantile-Quantile Plot
RADAM	Radar of Amazon (RADAM-Brasil)

RCE	Reference Channel Emissivity
RGB	Red, Green and Blue image composition
RK	Regression Kriging
RMSE	Root Mean Square Error
ROI	Regions of Interest
SAR	Synthetic Aperture Radar
SAVI	Soil Adjusted Vegetation Index
SLMR	Stepwise Linear Multiple Regression
SM	Spatial Quality Metric Index
SOC	Soil Organic Carbon
SOM	Soil Organic Matter
SOTER-LAC	Soils and Terrain Database - Latin America
SR	Symbolic Regression
SRTM	Shuttle Radar Topographic Mission
SSIM	Structural Similarity Index
SVM	Support Vector Machine
SWIR	Shortwave Infrared
TIR	Thermal Infrared
TIRS	Thermal Infrared Sensor
TOA	Top of Atmosphere Radiance
TOC	Total Organic Carbon
UA	User's Accuracy
USGS	United States Geological Survey
UTM	Universal Transverse Mercator
WDC	Water Dispersible Clay

## TABLE OF CONTENTS

<b>1.INTRODUCTION .....</b>	<b>19</b>
1.1. INTRODUÇÃO .....	24
<b>2. New Approaches to Evaluate Fusion Algorithms Using Landsat 8 and CBERS 2B Images in Natural Regions of Amazon Forest and Zambeian Flooded Grasslands .....</b>	<b>30</b>
2.1. Introduction .....	30
2.2. Methodology .....	32
2.2.1. Remote Sensing Data Acquisition and Preprocessing .....	33
2.2.2. Description of the Applied Fusion Algorithms .....	34
2.2.3. Qualitative Assessment .....	37
2.2.4. Quantitative Assessment .....	38
2.2.5. Quality Evaluation by Unsupervised Classification .....	40
2.3. Results .....	41
2.3.1. Qualitative Assessment .....	41
2.3.2. Quantitative Assessment .....	42
2.3.3. Indirect Quantitative Assessment .....	48
2.4. Conclusions .....	50
<b>3. A multi-sensor approach for mapping plant-derived carbon storage in Amazonian Podzols.....</b>	<b>55</b>
3.1. Introduction .....	55
3.2. Methodology .....	57
3.2.1. Study Area .....	57
3.2.2. Field data .....	58
3.2.3. Image data and processing methods .....	58
3.2.4. Soil map and correlation with field sample data .....	63
3.3. Results and Discussion .....	64
3.3.1. Vegetation and topographic features related to lateral variation in podzols .....	64
3.3.2. Classification of soil cover and generation of regional soil map .....	66
3.3.3. Mapping the deep-SOC stock in Podzol regions .....	69
3.4. Conclusions .....	74
References .....	74
<b>4. Evaluation of pedotransfer equations to predict deep soil carbon stock in tropical Podzols compared to other soils of Brazilian Amazon forest .....</b>	<b>77</b>

4.1. Introduction .....	78
4.2. Methodology.....	79
4.2.1. Field Sample Data .....	79
4.2.2. Estimation of Soil Bulk Density .....	81
4.2.3. Modeling the vertical distribution of SOC .....	83
4.3. Results... ..	86
4.3.1. Predicting Soil Bulk Density in Amazon Soils.....	86
4.3.2. Modeling the vertical distribution of SOC stock in Amazon soils .....	88
4.3.3. The SOC stock in Dataset A and B .....	92
4.4. Conclusions .....	92
<b>5. Mapping deep plant-derived soil carbon storage in soils of the Rio Negro basin.....</b>	<b>97</b>
5.1. Introduction .....	97
5.2. Methodology.....	99
5.2.1. Study Area .....	99
5.2.2. Field Sample Data .....	100
5.2.3. Ancillary data .....	102
5.2.4. Mapping the SOC stock.....	103
5.2.5. Evaluation of Predicted SOC stock maps .....	106
5.3. Results.....	106
5.3.1. Descriptive Statistics .....	106
5.3.2. Mapping the SOC stock in Rio Negro basin .....	109
5.4. Conclusions .....	119
References .....	120
<b>6. GENERAL REMARKS.....</b>	<b>125</b>
<b>REFERENCES .....</b>	<b>127</b>

## 1. INTRODUCTION

According to several researches developed in the last decades (POST et al., 1982; BURINGH, 1984; KIMBLEET et al., 1990; SOMBROEKET et al., 1993; ESWARANET et al., 1993; BATJES, 1996), the soils of the world have the capacity of storing about 2.2 Gt (Gigatons) of carbon, which make them one of the most important global carbon sink. The amount of carbon stored in soils represents more than twice of the carbon found in atmosphere and is three times bigger than the one stored as above ground biomass (AGB) on vegetation. In Amazon region, most of the soil organic carbon (SOC) is stored as soil organic matter (SOM) on the first 0.3m soil depth (BATJES; SOMBROEK, 1997). Given this, the precise estimation of SOC stock in soils, is an urgent matter considering future models of global climate change.

Most of the studies regarding global and regional estimates of SOC stock have taken into account the soil carbon stored in soils of temperate regions of Europa and North America. Few surveys were developed in natural areas of tropical and equatorial soils (BATJES, 1996). In 1996 Batjes (1996) carried out a study in tropical soils of America, Africa and Asia, which allowed for the update of the values presented in previous studies (POST et al., 1982; BURINGH, 1984; KIMBLE et al., 1990; SOMBROEK et al., 1993; ESWARAN et al., 1993). Therefore, Batjes (1996) concluded that the SOC stored in soils of tropical regions varies from 1.46 to 1.54 Pg (Petagrams), considering a soil depth of 0 to 0.3 m and 0 to 0.5 m, respectively. Surface and sub-superficial soil horizons were taken into account for SOC stock estimates because these depths are directly involved in interactions with the atmosphere and are sensitive to land use and environmental changes. However, great amounts of SOC occur in soil depths up to 2m in deep soil carbon pools of Acrisols, Ferralsols and Nitisols located in tropical regions, as well as in intrazonal Podzols (SOMBROEK et al., 1993). A few researches have considered the rule of deep SOC stock of Podzols on the global carbon soil reservoir (MONTES et al., 2011; PEREIRA et al., 2015). Therefore, a precise estimation of SOC stock in deep soil horizons is pointed out as a critical factor in implementing C trading programs, which depends on the understanding of the spatial distribution of SOC in order to quantify the capacity of soils in storing carbon.

The Rio Negro Basin, located at the upper Amazon Basin, is marked by the occurrence of Ferralsol and Acrisols widely distributed and closely related to Podzols. Red clayey Ferralsols are commonly found at the margin of strongly dissected low elevation plateaus that belong to the pan American Ucayali peneplain (CAMPBELL et al., 2006). By contrast,

Podzols occur on poorly drained depressions of the central parts of the plateaus. Podzol formation is associated the downward and downslope migration of organic acids (PEDRO, 1987) in highly porous sandy materials during the lowering of perched groundwaters. Accordingly, the soil orders in Rio Negro basin comprise two major end-members on the low elevation plateaus of this region, notably, Ferralsols/Acrisols association and Podzols. Recent research (MONTES et al., 2011) have shown that the transition between these two soil groups is marked by a clear increasing in deep SOC stock from ferralitic soils (Acrisols and Ferralsols) to Podzols.

The occurrence of Amazonian Podzols was reported on both crystalline and sedimentary rocks. The Podzol formation is caused by the downward and downslope migration of organic acids in porous sandy materials during the lowering of the groundwater (BRAVARD; RIGHI, 1989; NASCIMENTO et al., 2004). The accumulation of organic acids causes the weathering of clay minerals and the formation of SOM-rich (Soil Organic Matter) organo-metallic complexes (NASCIMENTO et al., 2004). Subsequently, the Al and Fe previously incorporated in the mineral phases of the ferralitic environments (Ferralsols and Acrisols), become predominantly bound to organic matter in Podzols. The high porosity of the eluvial horizon (E) of Podzols explains the short residence time of water from the perched groundwater (NASCIMENTO et al., 2008) and its fast lateral fluxes (LUCAS et al., 1996) that enhances the leaching and acidification of the soils, driving the formation of thick deep Bh horizons, rich in organic matter. At the upper Rio Negro Basin there is a significant area covered by Podzols (Giant Podzols), where the time of evolution of these soils was sufficient to lead to the formation of large areas of Podzols with deep thick spodic horizons (DUBROEUCQ; VOLKOFF, 1998). In these regions, Montes et al. (2011) reported a SOC stock of about  $13.6 \pm 1.1$  PgC, which is at least 12.3 PgC higher than previous researches (BATJES; DIJKHOORN, 1999). However, the estimate carried out by Montes et al. (2011) was developed according to a set of sample data located at the upper Rio Negro Basin, without precise extrapolation to the entire region of this basin.

Nowadays, with a significant advance in remote sensing imagery and geoprocessing techniques, some researches (VAN-MEIRVENNE et al., 1996; POST et al., 2001) have shown the importance in developing digital SOC stock maps at different spatial scales such as plot, watersheds, regional, national and continental levels. With regards to Amazon forest, the Brazilian portion of the Amazon basin has an absolute area of  $3.84 \cdot 10^6$  km<sup>2</sup>, which makes it the world's largest continental basin. In such an extensive region, the lack of field sample data and the absence of systematic soil surveys in deep soil profiles are among the main reasons

for the absence of SOC stock maps in regional map scale. The proper estimation of the SOC stored in Amazon soils, depends on a range of factors likely to affect SOC content such as soil types, land use, annual input of C biomass, relief, natural vegetation cover, lithology and climate (POST et al., 2001).

The distribution of soil orders in Amazon basin is closely related to a series of environmental variables that are indirectly mapped through the interpretation of remote sensing imagery (eg.: relief, surface moisture, surface temperature and soil cover) provided by different sensor systems (PEREIRA et al., 2015). The northwest portion of the Amazon basin is characterized by the occurrence of a high dense evergreen rainforest, which makes it difficult the direct mapping of soil types by using remote sensing imagery (DUBROEUCQ; VOLKOFF, 1999; PEREIRA et al., 2015). Therefore, the lateral segmentation of soil orders depends on the association of remote sensing data (passive and active remote sensor systems) with field sample data and current soil maps available at regional map scale provided by legacy databases (IBGE, 2008; EMBRAPA, 2014). Given this, remote sensing data associated to field samples and legacy data, are essential in order to refine the current soil maps, which is critical to map and to quantify the SOC storage in Podzols of Rio Negro basin.

The availability of legacy data systematically collected in Amazon region in the last decades (IBGE, 2008), was important to improve the map scale of previous maps (1:1,000,000) regarding soils, geology, geomorphology and soil cover, originally provided on the frame of the RADAMBRASIL project in 70<sup>th</sup> decade (BRASIL, 1977). Nowadays, the most refined maps available for this region are provided at the map scale of 1:250,000, which represent an important increment regarding a better understanding of the Amazonian biome. Moreover, IBGE (2008) has delivered a range of soil samples related to more than 400 soil profiles within the region of the Rio Negro basin. These profiles account to more than 2,300 sampled horizons from soil depths ranging from 0.5 to 3m. The systematic database is compatible with GIS (Geographic Information Systems) allowing the spacialization of sampled profiles and the implementation of multivariate statistical analysis of environmental attributes with continuous coverage (eg.: relief and its derivatives, surface temperature, surface absolute reflectance in visible and infrared ranges of the electromagnetic spectrum) and categorical distribution (eg.: geology, soils, soil cover and geomorphology).

Legacy data might be the only systematic source of information in the Amazon region (EASTER et al., 2007; IBGE, 2008; EMRAPA, 2014) and its usage for spatializing SOC stock is essential in order to improve the current SOC stock maps in Amazon (MONTES et al., 2011), especially regarding the SOC stored in deep soil horizons. Some effort has been

made in previous research to map the SOC storage in Amazon soils, however the soil samples considered in these researches were provided by continental legacy data in the frame of the SOTER-LAC (Soils and Terrain Database/Latin America) initiative (BATJES, 2002; EASTER et al., 2007), which is compatible with the map scale of 1:1,000,000. Based on IBGE (2008) database and the implementation of multivariate statistics, it is possible to generate SOC stock maps with a map scale of 1:250,000. Moreover, the quantification of deep SOC stock in Amazon soils is essential, due to the lack of studies addressing these stocks.

Even the most refined systematic database available in Amazon basin (IBGE, 2008), has a few number of soil samples collected in Podzol areas. Moreover, the Podzol profiles have been sampled in soil depths ranging from 1 to 1.5m, disregarding deep spodic (Bh) horizons. Thus, detailed databases are necessary in area of Podzols in order to estimate the deep-SOC storage in Amazon soils. The estimation of deep-SOC storage is essential if we take into account a scenario of global climate change. The high rainfall around 2600 mm per year and the dense forest coverage are among the main factors that drive the podsolization process in hydromorphic conditions regarding Amazon Podzols. Climate models suggesting diminution in average annual rainfall usually take into account the mineralization of SOC pools in soil depths from 0-0.3 and 0-1m (CERRI et al., 2007).

The decreasing in annual rainfall in the region of Rio Negro basin could change drastically the average groundwater level in areas of Podzols, which could affect the soil hydrologic regime leading to the oxygenation of elluvial and subsequently spodic horizons (MONTES et al., 2011). The high availability of oxygen might increase the microbial activity in deep SOC pools of spodic horizons causing the mineralization of organic carbon and the releasing of significant amounts of CO<sub>2</sub> to atmosphere. Therefore, the proper quantification and mapping of deep SOC storage in Podzols is an urgent matter towards the proportion of SOC models considering a scenario of decreasing in annual rainfall in Amazon forest, for the next decades.

The central hypothesis is that the amount of carbon stored in Podzols is higher when compared to adjacent soils (Ferralsols and Acrisols) and its spatial distribution is related to environmental variables that can be inferred by using regional maps, as well as from remote sensing imagery. Therefore, the main goal of this research was to quantify and to map the SOC storage of Amazon soils in the region of Rio Negro basin, taking into account the carbon pool stored in deep soil horizons. This research is organized in four chapters evolving the use of remote sensing imagery, field sample data and legacy data, in order to estimate the amount of SOC stored in soils of the Rio Negro basin. In the first chapter we explored the use of

multi-sensor imagery on the generation of high spatial resolution multispectral compositions by adopting image fusion approaches. The final multispectral fused images have an absolute spatial resolution of 5m. Thus, it was possible to generate multispectral compositions suitable for detailed mapping of soil cover in extensive regions of Amazon, which is essential for the refinement of current maps available in this region. The generation of high spatial resolution multispectral compositions by combining remote sensing data with different spatial, temporal, radiometric and spectral resolutions, is suggested as an efficient low cost method to map soil cover in Amazon region.

In the second chapter, we discussed the application of remote sensing imagery to spatialize and to map soil groups in Amazon Podzols. Remote sensing images and field sample data were used to estimate the SOC stock at the studied area. The distribution of Podzols is closely related to the topography due to local variations on the groundwater level. Given this, a high variability on SOC content was found within Podzol areas, where we observed regions of well drained Podzols, seasonally flooded Podzols and overflooded Podzols. The availability of soil maps at regional map scale is an urgent matter to spatialize soil units within Podzols, which allows for a better understanding of the environmental attributes related to the lateral variation of SOC stock in Amazon Podzols.

The chapters three and four refer to the estimation and mapping of SOC stock in the region of Rio Negro basin, taking into account deep soil horizons at 1m and 3m soil depths. The final maps were obtained by regression kriging of predicted values of SOC stock. The proper estimation of SOC stock in Rio Negro basin at the abovementioned soil depths depends on the modeling of the SOC stock by pedotransfer techniques and the interpolation of the resulted values by regression kriging, which allows the association of SOC stock with ancillary datasets. Thus, the presented method is complex and deals with a certain level of uncertainties; however the associated errors can be quantified. Given this, we were able to estimate and map the amount of SOC stored in deep soil horizons of Rio Negro basin, with an unprecedented precision.

## 1.1. INTRODUÇÃO

De acordo com inúmeras pesquisas desenvolvidas nas últimas décadas (POST et al., 1982; BURINGH, 1984; KIMBLEET et al., 1990; SOMBROEKET et al., 1993; ESWARANET et al., 1993; BATJES, 1996), os solos de todo o globo tem a capacidade de estocar em torno de 2.2 Gt (Gigatoneladas) de carbono, o que os tornam uma das mais importantes reservas de CO<sub>2</sub> do planeta. A quantidade de carbono estocada nos solos representa o dobro daquela encontrada na atmosfera e é três vezes maior que o carbono armazenado na forma de biomassa vegetal. Na região Amazônica, a maior parte do carbono orgânico do solo (CO) está estocada na forma de matéria orgânica (MOS) nos primeiros 0,3m de profundidade (BATJES; SOMBROEK, 1997). Diante disso, destaca-se a necessidade de estudos mais aprofundados voltados a estimativa do estoque de CO do solo, considerando-se cenários futuros de mudanças climáticas.

A maioria dos estudos voltados a estimativas locais e regionais do estoque de CO do solo levam em conta os estoques de regiões temperadas da Europa e América do Norte. Poucas pesquisas foram desenvolvidas em áreas tropicais e equatoriais (BATJES, 1996). No ano de 1996, Batjes (1996) estimou os estoques de CO de solos tropicais em regiões da América, África e Ásia, o que possibilitou a atualização de valores já apresentados em pesquisas anteriores (POST et al., 1982; BURINGH, 1984; KIMBLE et al., 1990; SOMBROEK et al., 1993; ESWARAN et al., 1993). Por meio deste trabalho, Batjes (1996) concluiu que os estoques de CO de solos tropicais variam de 1,46 a 1,54 Pg (Petagramas), considerando-se as profundidades de 0 a 0,3m e 0 a 0,5m, respectivamente. Foram considerados, portanto horizontes pedológicos superficiais e sub-superficiais, pois tais profundidades estão diretamente relacionadas com as interações com a atmosfera e são mais sensíveis a mudanças na cobertura do solo. Porém, grandes quantidades de CO podem ser encontradas em horizontes pedológicos com até 2m de profundidade em áreas de Argissolos, Latossolos e Neossolos de regiões tropicais, bem como, em Espodossolos intrazonais (SOMBROEK et al., 1993).

São poucas as pesquisas que levam em conta o papel dos estoques de carbono profundo dos Espodossolos na reserva global de carbono do solo (MONTES et al., 2011; PEREIRA et al., 2015). Sendo assim, novas estimativas mais precisas do estoque de CO são necessárias para a implementação de mecanismos de comercialização de carbono, o que depende de uma melhor estimativa da distribuição espacial de tais estoques no contexto da bacia amazônica.

A bacia do Rio Negro, localizada na alta bacia Amazônica, é marcada pela ocorrência expressiva de Argissolos e Latossolos com Espodossolos associados. Os Latossolos são comumente encontrados às margens dos platôs dissecados de baixa altitude pertencentes ao pediplano pan-americano Ucayali (CAMPBELL et al., 2006). Já os Espodossolos, ocorrem em depressões pouco drenadas nas regiões centrais dos platôs. A formação desses solos está relacionada à migração lateral e vertical de ácidos orgânicos (PEDRO, 1987) em horizontes pedológicos arenosos altamente porosos durante as fases de rebaixamento do lençol freático suspenso. Os tipos de solo encontrados na bacia do Rio Negro são compostos basicamente por dois grandes grupos compreendendo a associação Latossolo/Argissolo e os Espodossolos. Pesquisa recente desenvolvida nessa região (MONTES et al., 2011) constatou que a transição entre solos ferralíticos (Argissolos e Latossolos) e Espodossolos é marcada por um aumento significativo na quantidade de MOS estocada em horizontes profundos.

A ocorrência de Espodossolos intrazonais amazônicos foi constatada em áreas de relevo cristalino e em zonas de rochas sedimentares. A formação destes solos é devida à migração lateral e vertical de ácidos orgânicos através de materiais arenosos, durante o rebaixamento do lençol freático (BRAVARD; RIGHI, 1989; NASCIMENTO et al., 2004). A acumulação de ácidos orgânicos leva a formação de compostos organo-metálicos ricos em matéria orgânica dissolvida (MOD) (NASCIMENTO et al., 2004). Os compostos de Al e Fe previamente incorporados à fase mineral em ambientes ferralíticos (Latosolos e Argissolos), se tornam predominantemente ligadas à matéria orgânica nos Espodossolos. A alta porosidade do horizonte eluvial (E) dos Espodossolos, explica o curto tempo de permanência do lençol freático suspenso (NASCIMENTO et al., 2008) e a prevalência de fluxos laterais rápidos (LUCAS et al., 1996) que acentuam os processos de lixiviação e acidificação desses solos, ocasionando a formação de horizontes espódicos espessos ricos em matéria orgânica. Na região da alta bacia do Rio Negro, há uma grande área coberta por Espodossolos (Espodossolos Gigantes), onde o tempo de evolução pedológica foi suficiente para possibilitar a formação de tais solos, com horizontes espódicos profundos e espessos (DUBROEUCQ; VOLKOFF, 1998). Nesta área, Montes et al. (2011) constatou que o estoque CO do solo está em torno de  $13.6 \pm 1.1$  PgC, valor que é 12.3 PgC superior àquele apresentado em pesquisas anteriores (BATJES; DIJKHOORN, 1999). No entanto, as estimativas apresentadas por Montes et al. (2011) foram desenvolvidas com base em um conjunto de amostras coletadas na alta bacia do Rio Negro, sem possibilidade de uma extrapolação precisa para toda a região da bacia.

Atualmente, com o avanço nas técnicas de geoprocessamento e sensoriamento remoto, algumas pesquisas (VAN-MEIRVENNE et al., 1996; POST et al., 2001) ressaltaram a importância do mapeamento digital do estoque de CO do solo em diferentes escalas espaciais desde mapeamentos locais até mapas em escala continental. Com relação à floresta amazônica, a sua porção brasileira ocupa uma área aproximada de  $3.84 \cdot 10^6$  km<sup>2</sup>, compreendendo a maior bacia hidrográfica continental do planeta. Nesta região, a escassez de amostras de solo de horizontes pedológicos profundos pode ser apontada como a principal razão para a ausência de mapeamentos sistemáticos do estoque de CO do solo em escala regional. A adequada estimativa deste estoque depende de uma série de fatores que podem afetar a ocorrência de CO, tais como, tipos de solo, uso do solo, biomassa vegetal, relevo, vegetação, litologia e clima (POST et al., 2001).

A distribuição espacial dos tipos de solo na bacia amazônica está relacionada a uma série de variáveis ambientais que podem ser indiretamente mapeadas por meio de interpretação de imagens de sensoriamento remoto (Ex.: relevo, umidade superficial, temperatura superficial e cobertura do solo) adquiridas por diferentes sistemas sensores (PEREIRA et al., 2015). A porção noroeste da bacia amazônica é caracterizada pela ocorrência de uma floresta densa que dificulta o mapeamento direto dos tipos de solo por meio de produtos de sensoriamento remoto (DUBROEUCQ; VOLKAFF, 1999; PEREIRA et al., 2015). Sendo assim, a segmentação lateral dos tipos de solo pode ser feita através da associação de imagens orbitais (sistemas de sensores ativos e passivos), amostras de campo e mapas temáticos da área de interesse em diferentes escalas (IBGE, 2008; EMBRAPA, 2014). Portanto, a associação de informações de diferentes fontes oferece uma alternativa interessante para o refinamento dos mapas de solo atualmente disponíveis em território amazônico. Tal aspecto é essencial para o mapeamento e quantificação do estoque de CO em Espodossolos da bacia do Rio Negro.

A disponibilidade de dados gratuitos em território amazônico nas últimas décadas (IBGE, 2008), possibilitou o refinamento de mapas temáticos de solo, vegetação, geomorfologia e geologia em escalas menores (1:1.000.000) provenientes de estudos desenvolvidos na década de 70 no âmbito do projeto RADAMBRASIL (BRASIL, 1977). Atualmente, são disponibilizados mapas em escala absoluta de 1:250.000, envolvendo toda a Amazônia legal brasileira, o que representa um avanço importante no entendimento do funcionamento do bioma amazônico. Além disso, o IBGE (2008) disponibilizou uma série de amostras de solo envolvendo mais de 400 perfis na região da alta bacia do Rio Negro. Tais amostras compreendem mais de 2300 horizontes amostrados para profundidades que variam

de 0,5 a 3m. Este banco de dados é compatível com sistemas de informação geográfica (SIG), o que possibilita a espacialização das amostras e aplicação de estatísticas multivariadas relacionada a atributos ambientais com distribuição contínua (Ex.: relevo e produtos derivados, temperatura superficial, refletância superficial absoluta nos comprimentos de onda do visível e infravermelho), como também dados categóricos (Ex.: geologia, solos, cobertura do solo e geomorfologia).

Na região amazônica, dados sistemáticos disponibilizados por agências de pesquisas (EASTER et al., 2007; IBGE, 2008, EMRAPA, 2014) podem ser adotados para espacialização do estoque de CO viabilizando o refinamento das atuais estimativas já disponíveis (MONTES et al., 2011), sobretudo com relação ao estoque de horizontes profundos. Algumas pesquisas anteriores buscaram estimar o estoque de CO em solos da Amazônia, porém, as amostras consideradas foram extraídas de banco de dados continentais do projeto SOTER-LAC (*Soils and Terrain Database/Latin America*) (BATJES, 2002; EASTER et al., 2007), que é compatível com escala de mapeamento de 1:1.000.000. Com base em dados do IBGE (2008) combinados às técnicas de estatística multivariada, é possível a obtenção de mapas do estoque de CO com escala absoluta de 1:250.000. A quantificação do carbono profundo de solos amazônicos é de indispensável importância, considerando a ausência de estudos sistemáticos voltados à estimativa de tais estoques.

Mesmo os mapeamentos sistemáticos mais refinados da região amazônica (IBGE, 2008), tem uma pequena quantidade de amostras de solo coletadas em áreas de Espodosolos. Além disso, os perfis de solo limitam-se, em sua maioria, aos primeiros 1,5m de profundidade, desconsiderando os horizontes espódicos profundos (Bh). Diante disso, é necessária a obtenção de uma base de dados mais detalhada em áreas de Espodosolos buscando-se a estimativa do estoque de CO profundo. Tais estimativas são essenciais para a modelagem de cenários futuros de mudanças climáticas. A alta pluviosidade, em torno de 2600 mm por ano e a densa cobertura vegetal são os principais fatores que possibilitam a podzolização em condições hidromórficas em áreas de Espodosolos amazônicos. Modelos climáticos que sugerem uma diminuição nas taxas anuais de chuvas para esta área, normalmente consideram estoques de CO do solo em profundidades de 0 a 0,3m e de 0 a 1m (CERRI et al., 2007).

A diminuição da pluviosidade anual na região da bacia do Rio Negro poderia acarretar um drástico abaixamento no nível médio do lençol freático em zonas de Espodosolos, o que causaria uma mudança no regime hídrico do solo levando à oxigenação dos horizontes eluviais e espódicos (Bh) (MONTES et al., 2011). O aumento na disponibilidade de oxigênio

atmosférico no solo pode acentuar a atividade microbiana em horizontes espódicos causando a mineralização do CO e a emissão de grandes quantidades de CO<sub>2</sub> para a atmosfera. Sendo assim, é importante realizar a adequada estimativa e mapeamento do estoque de CO do solo em áreas de Espodossolos, buscando-se aprimorar modelos climáticos futuros que considerem diminuição gradativa nos índices de pluviosidade da Amazônia, para as próximas décadas.

A principal hipótese é que a quantidade de carbono estocada nos Espodossolos da bacia do Rio Negro é mais elevada quando comparada aquela presente em solos adjacentes (Argissolos e Latossolos) e que a distribuição espacial destes estoques pode ser correlacionada com variáveis da paisagem representadas em mapas temáticos e imagens de sensoriamento remoto. Portanto, o objetivo desta pesquisa foi quantificar e mapear o estoque de CO de solos Amazônicos na região da bacia do Rio Negro, considerando-se horizontes pedológicos profundos.

Este trabalho está organizado em quatro capítulos que tratam do uso de sensoriamento remoto, dados de campo e mapas temáticos, para a estimativa do CO estocado em solos da bacia do Rio Negro. No primeiro capítulo foi explorada a aplicação de imagens orbitais obtidas por diferentes sistemas de sensores para a obtenção de composições multiespectrais com alta resolução espacial, por meio da adoção de métodos de fusão de bandas. As imagens fusionadas resultantes têm uma resolução espacial absoluta de 5m. Com isso, foi possível a geração de composições multiespectrais compatíveis com mapeamento detalhado da cobertura do solo em grandes áreas de Amazônia. Tal aspecto é essencial para o refinamento dos atuais mapas de cobertura do solo disponíveis para essa região. Sendo assim, neste trabalho sugere-se a geração de composições multiespectrais com alta resolução espacial por meio da combinação de imagens de sensoriamento remoto oriundas de diferentes sistemas de sensores com resoluções espacial, temporal, espectral e radiométrica, distintas. Esta metodologia pode ser adotada para o mapeamento detalhado da cobertura do solo na Amazônia com baixos custos.

No segundo capítulo, discutiu-se a aplicação de imagens de sensoriamento remoto para a espacialização e mapeamento de grupos de solo em áreas de Espodossolos amazônicos. Imagens orbitais e dados de campo foram associados para estimar o estoque de CO na área estudada. Observou-se que a distribuição dos Espodossolos está intimamente relacionada à topografia devido a variações locais no nível dos lençóis freáticos. Foi constada uma alta variabilidade na quantidade de CO do solo em áreas de Espodossolos, onde se observou a ocorrência de Espodossolos bem drenados, Espodossolos sazonalmente alagados e Espodossolos alagados. Concluiu-se neste estudo, que a disponibilidade de mapas detalhados

de cobertura do solo para a região avaliada é de indispensável importância para a especialização dos diferentes tipos de Espodossolos que ocorrem nesta área, o que pode viabilizar uma melhor compreensão dos aspectos ambientais relacionados à variação do estoque de CO nos Espodossolos amazônicos.

Os capítulos três e quatro abordam a estimativa e mapeamento do estoque de CO dos solos na região da bacia do Rio Negro, considerando-se os estoques até 1m e até 3m de profundidade. Os mapas finais foram obtidos por meio de krigagem por regressão dos valores preditos do estoque de CO. A adequada estimativa de tais estoques para a bacia do Rio Negro, de acordo com as profundidades abordadas, depende da modelagem da quantidade de CO na escala do perfil de solo, com base em métodos de pedotransferência e interpolação dos valores resultantes por krigagem, possibilitando a correlação com variáveis auxiliares. Diante disso, o método utilizado é complexo e considera uma série de incertezas, porém, o erro associado pode ser quantificado. Com isso, foi possível estimar e mapear os estoques de CO profundo na região da bacia do Rio Negro, com uma precisão que ainda não havia sido alcançada.

## 2. New Approaches to Evaluate Fusion Algorithms Using Landsat 8 and CBERS 2B Images in Natural Regions of Amazon Forest and Zambebian Flooded Grasslands<sup>1</sup>

### Abstract

Advances in remote sensing technology and the release of new sensor systems have been providing a wide range of optical and radar satellite images. The availability of such images gives new options for mapping relatively remote and sparsely settled territories. Given this, the main goal of this research was to perform a qualitative and quantitative assessment of the quality of a set of fused images obtained by CBERS 2B (HRC) and Landsat 8 (OLI) satellites over natural regions of Amazon Forest and Zambebian Flooded Grasslands, by the adoption of zonal and global quality evaluation approaches. Through the applied methodology we were able to combine the spectral resolution of Landsat 8 OLI images with the spatial resolution of the CBERS/HRC panchromatic band. The quantitative evaluation of the spectral and spatial quality of the fused products based on a zonal approach was essential to understand the spatial variability of the fusion quality, according to each fusion method. Based on the obtained results we observed that CBERS/HRC panchromatic bands can be satisfactorily applied in substitution of Landsat 8/OLI panchromatic band according to the ATWT and Ehlers fusion methods, allowing the generation of multispectral fused images with a 5-m spatial resolution.

**Keywords:** Image Fusion, Landsat 8, CBERS, Image Enhancement, Regional soil cover maps

### 2.1. Introduction

The Brazilian Institute for Space Research (INPE) and the China Academy of Space Technology (CAST) have launched to space the China–Brazil Earth Resources Satellite (CBERS-4) in December 2014, allowing the continuity of the CBERS program. This mission has the main goal of providing high resolution multispectral and panchromatic images to develop researches concerning natural areas still poorly mapped, especially in Latin America and Africa (ABDON et al., 2009; DAL’ASTA et al., 2012). The CBERS-4 sensors will provide 3 bands on visible (green and red) and near infrared ranges of electromagnetic spectrum with a 10-m spatial resolution and one panchromatic band (Multispectral and Panchromatic Camera: PAN) with a 5-m spatial resolution. The CBERS-4 platform has three other cameras covering visible, near infrared, shortwave infrared (SWIR) and thermal infrared (TIR) regions of the electromagnetic spectrum (MUX, IRS and WFI sensors).

---

<sup>1</sup> Submitted to the International Journal of Image Fusion (Taylor & Francis. ISSN: 1947-9832), in July 2015. Authors: O. J. R. Pereira, C. R. Montes, T. C. T. Pissarra, Y. Lucas, A. Minghelli-Roman, A.J.Melfi

Around 70% of optical satellites on orbit nowadays generate bands on both modes, panchromatic and multispectral, usually with high and moderate spatial resolution, respectively (ZHANG 2004; 2008). Among them, it is important to highlight the Landsat program. The first Landsat satellite (Landsat 1/ MMS), was placed in orbit in 1972. Ever since, 8 satellites were launched until the Landsat 8 mission, currently in operation (USGS, 2014). The ETM+ Landsat 7 (Enhanced Thematic Mapper plus) sensor was the first one to generate multispectral and panchromatic images with 30-m and 15-m spatial resolution, respectively. The availability of a panchromatic band was important to allow the generation of regional and detailed soil cover maps, considering the capability of applying image fusion techniques (FAUNDEEN et al., 2004). Due to the advances of the Landsat 7 mission, the panchromatic camera was then added to the OLI (Operational Land Image) sensor on Landsat 8 platform. However, the band width was reduced to the visible range of the electromagnetic spectrum (0.50 to 0.68  $\mu\text{m}$ ) which might derail the application of fusion methods in the near infrared (NIR) and short wave infrared (SWIR).

Image fusion techniques has been used as the combination of two or more images (MARCELINO et al., 2009; CHENG et al., 2011), and there is a great variety of fusion methods comprising different procedures with diverse equations from simple arithmetic operations to more sophisticated algorithms that applies a series of transformation on the original multispectral and panchromatic bands (DONG et al., 2013; KHALEGHI et al., 2013; XIA; LEUNG, 2014). The simplest techniques are easier to implement and can be promoted in most of GIS currently available. The more accurate fusion techniques might be difficult to apply due to the complexity of the base equations, but usually generate better results when compared to simpler algorithms. Some studies have investigated the quality of multispectral fused images based on qualitative and quantitative approaches (RANCHIN; WALD, 2000; RANCHIN et al., 2003; ALPARONE et al., 2007; GALIANO et al., 2012). The results of the quantitative approach are expressed as average values representing an overall index for the fused images compared to the reference image (global quality index).

The quantitative evaluation of the quality of the fused images depends on the availability of a reference multispectral composition representing an optimal scenario regarding the spectral and spatial attributes (WALD et al., 1999; WANG; BOVIK, 2002; WANG et al., 2004). Thus, a feasible way to obtain the reference image would be through degradation of all available data to a coarser resolution and carrying out the fusion from the resulting degraded multispectral composition (WALD et al., 1997). To access and compare results, the average quality values are generated by a set of equations that evaluate the spatial

and spectral quality of the fused bands. The representation of the arithmetic mean for the whole scene might result in biased values, due to a considerable lateral variation of soil cover. Such problem can be solved by adopting a zonal approach (zonal sampling windows), which allows the generation of intermediate quality evaluation values for specific image features.

The main goal of this research was to perform a qualitative and quantitative assessment of the quality of a set of fused images obtained by CBERS 2B (HRC) and Landsat 8 (OLI) satellites over natural regions of Amazon Forest and Zambebian Flooded Grasslands, by the adoption of zonal and global evaluation approaches. The present paper intended also to investigate the applicability of CBERS panchromatic band in substitution of the panchromatic Landsat OLI band.

## 2.2. Methodology

The satellite images were acquired in two distinct areas with a great diversity of soil covers, from high dense rainforest in Amazon, to shrub savannas and seasonally flooded grassland fields in Zambia (Figure 2.1). The Amazon study area (Figure 2.1a) is located near the city of Santa Isabel do Rio Negro (Amazonas State). The mainly vegetation classes are the high dense rainforest, in regions of Acrisols and Ferralsols, associated to low convex hills and sclerophyllic vegetation (Campinarana) related to soils in hydromorphic conditions (Podzols and Gleysols) (BRASIL, 1977; IBGE, 2008).

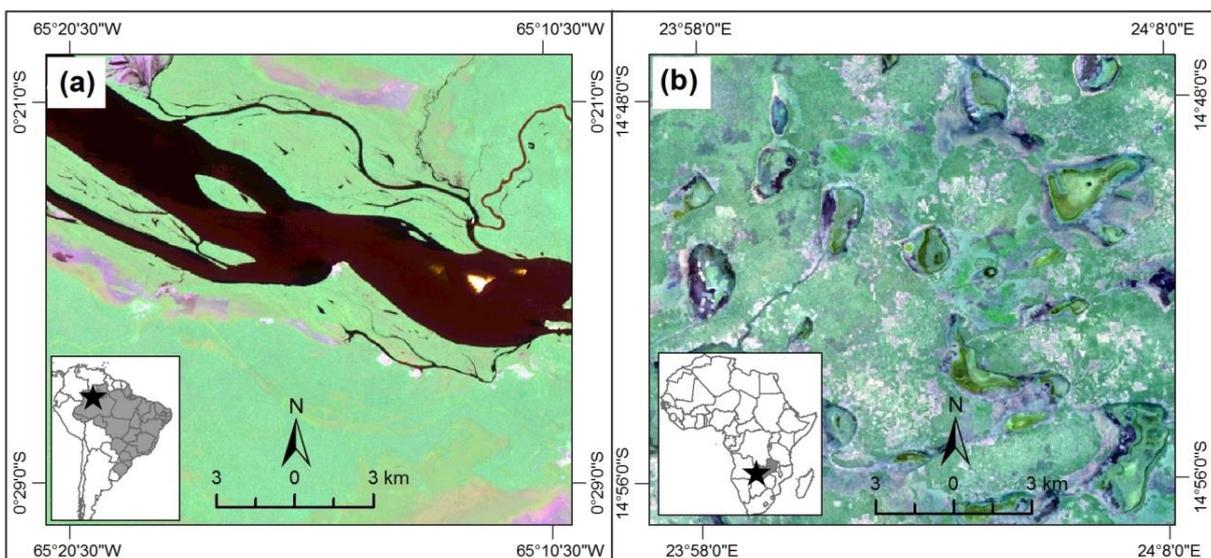


Figure 2.1 - Studied areas. (a): Amazon Forest site; (b) Zambebian Flooded Grasslands (Zambia site). Multispectral composition (reference image) for both regions (Synthetic OLI bands: Blue/Green; Red/Near Infrared and SWIR 1/SWIR 2 - RGB).

The Zambia investigation site is located a few miles west from Mongu village (Figure 2.1b). There is a great diversity of soil covers from flooded grasslands (Zambeian Flooded Grassland) related to seasonally flooded Podzols, to bare ground sites, usually associated to agricultural activity. The natural vegetation fields grade from grasses and lichens to shrubs and patches of savanna woodlands (BOOTH et al., 1994).

### 2.2.1. Remote Sensing Data Acquisition and Preprocessing

We adopted Landsat 8 OLI multispectral compositions (USGS, 2014) and CBERS HRC panchromatic bands (INPE, 2014) as shown in Table 2.1. The images were geometrically co-registered to a previously rectified Landsat 7 ETM+ composition (Universal Transverse Mercator, datum: WGS 1984), corrected and orthorectified to level 1B products, with a sub-pixel precision, considering ETM+ images with a 30-m spatial resolution (GLCF, 2009).

Table 2.1 - Summary of remotely sensed data.

Satellite Images*	Bands	Spatial Resolution	Spectral Resolution (PAN)	Radiometric Resolution
Landsat Multispectral	Blue, Green, Red, NIR, SWIR1, SWIR2	30-m	-	16 bits
Landsat PAN.	Panchromatic	15-m	0.50-0.68 $\mu\text{m}$	16 bits
CBERS/HRC PAN	Panchromatic	2.5-m**	0.50-0.80 $\mu\text{m}$	8 bits

\*Acquisition Date: Amazon site – Landsat (29/09/2014); CBERS 2B (04/01/2010). Zambia site - Landsat (08/12/2013); CBERS (15/07/2008). \*\* Resampled to 5-m in this study.

The positional accuracy has a root mean square Error (RMSE) better than 30m, (GLCF, 2009) on multispectral image, in both studied areas (Figure 2.1). There is a significant temporal difference between CBERS and Landsat OLI images, as shown in Table 2.1, due to a low availability of cloud free scenes in Amazon region, and the failure of the CBERS 2B satellite in 2010. In Amazon area, there is no human disturbance on natural vegetation and the most significant spectral distortion between scenes collected in different periods is caused by seasonality. The area in Zambia has some agricultural activity, which might leads to seasonal changes on soil cover associated to the dynamic of those lands (HUEMMRICH et al., 2005; JIN et al., 2003). Therefore, atmospheric correction for the multispectral and panchromatic images was applied (ADLER-GOLDEN et al., 1999; BERK et al., 2002).

The Landsat OLI bands have been atmospherically corrected using FLAASH Model (ENVI®), after conversion of image digital values to top of atmosphere radiance (TOA). The parameters used in this correction were set to tropical model. No cloud masking process was required, since the images were cloud-free scenes in both studied areas. The 5S physically-

based calibration model MODTRAN (ADLER-GOLDEN et al., 1999; BERK et al., 2002) was applied with the aerosol model set to rural, the initial visibility was set to 40km and 35km for Amazon and Zambia sites, respectively. This method was applied for calibrating the following bands: blue, green; red; NIR; SWIR 1; SWIR 2 and OLI panchromatic band (OLI bands 2, 3, 4, 5, 6, 7 and 8, respectively). After calibration, the resulting reflectance images were rescaled to 8 bits.

The atmospheric correction applied for panchromatic CBERS HRC bands, is based on an algorithm developed by Carlotto (1999). The algorithm allows dynamic selection of grey values as reference to be subtracted on the original scene. The chosen targets were selected in areas of clear water bodies, where pixel subtraction was carried out. After calibration, the resulting images were rescaled to 8 bits and resampled to a 5-m spatial resolution (bicubic resampling) to simulate the panchromatic band (HRC) that will be provided by the CBERS-4 satellite.

### ***2.2.2. Description of the Applied Fusion Algorithms***

The images were resampled according to the flowchart on Figure 2.2, to match the resolution of the fused images with a 30-m spatial resolution (reference composition: Landsat OLI). The panchromatic channels of CBERS 2B and Landsat OLI satellites have different spectral, radiometric and spatial resolution. The method presented in Figure 2.2, was conducted in order to standardize the spatial resolution, according to the reference image.

The ratio between the optical Landsat bands blue/green, red/NIR, SWIR 1/SWIR 2 (RGB), allows synthesizing the spectral information of those channels in three bands, which makes it easier the application of fusion algorithms and the quantitative evaluation of results, saving computational time (RODRIGUEZ-GALIANO et al., 2012). The above-mentioned multispectral composition was used to simulate the reference image to be fused with the panchromatic HRC an OLI bands (Figure 2.2).

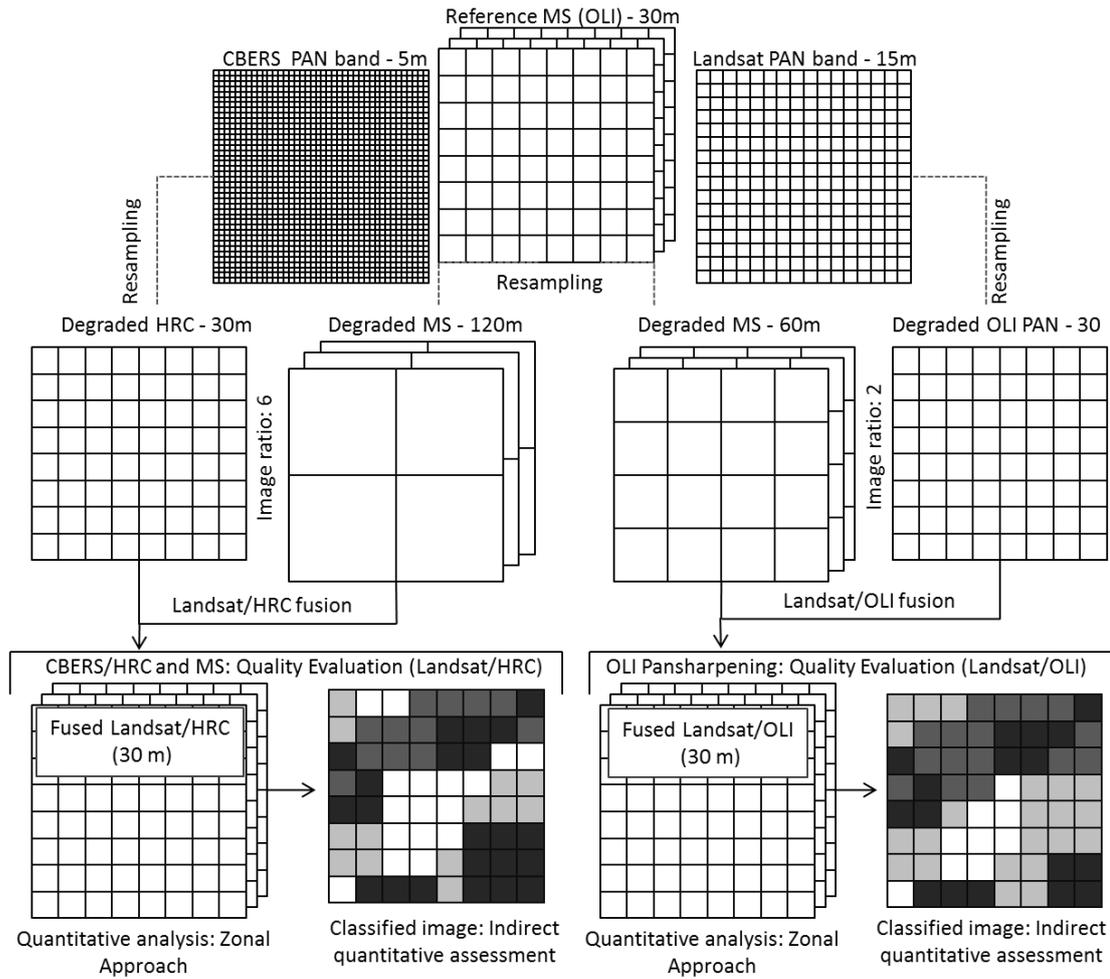


Figure 2.2 - Flowchart of the methodology applied to evaluate the fusion algorithms by a zonal approach and by unsupervised classification of fused compositions.

The flowchart in Figure 2.2 illustrates the applied method, carried out for evaluating quantitatively the fused compositions. Two steps were considered, one based on the direct zonal quantitative assessment and another one carried out by unsupervised classification (ISODATA) of the degraded fused images and the reference images. The results were generated for both studied areas (Figure 2.1), taking into account the fusion algorithms briefly described below:

1. *Color Normalized - Brovey*. It is a simple and ease to implement method for combining data from different sensors. The algorithm consists of an arithmetic normalization of the spectral information on the three multispectral (MS) channels (RGB). After normalization, the resulting images are multiplied with the panchromatic high spatial resolution band (PAN), which allows the maintenance of the spectral feature of each pixel and the refinement of the spatial resolution of the fused MS composition. The general equation uses red, green, and blue (RGB) and the PAN bands as inputs to output new red, green, and blue bands with refined spatial resolution.

2. *Principal Components*: The principal component method is based on the transformation of the spectral information of the MS bands in principal components. Most of the spectral variability of the MS channels can be synthesized on the first component (CHÁVEZ et al., 1991; SHAH, 2008). After a principal component analysis, the resulting first component of the low resolution MS composition is substituted by the PAN high resolution band. Finally the inverse transformation is applied to generate the RGB fused MS composition.

3. *Gram Schmidt*: In this method, the spatial resolution of the MS image is enhanced by merging the high resolution single band (PAN) with the low spatial resolution MS bands (EASTMAN KODAK COMPANY, 2000). A lower spatial resolution panchromatic band (GS) is simulated using the multispectral bands as input. The simulated high resolution GS band is employed as the first band. Then, the PAN band is swapped with the first GS band. Finally, the inverse GS sharpening transform is applied to form the fused MS spectral bands.

4. *Intensity Hue and Saturation (IHS)*: The IHS method converts a three bands MS composition from the red, green and blue (RGB) space into the IHS color space. The intensity band (I) in the IHS space is replaced by a high-resolution PAN image. The final fused composite image is obtained by converting back the IHS components into the original RGB space (CHOI, 2006).

5. *Wavelet Transform*: In Wavelet transformation the images are decomposed in high and low-frequency components, comprising the higher and the lower spatial resolution images, respectively. In this study we applied two different wavelet algorithms based on Mallat's (DWT) (ZHOU, 1998; RANCHIN; WALD, 2000) and à trous wavelet (ATWT) transformations (NUÑEZ, 1999; RANCHIN et al., 2003). Each of the above-mentioned methods has specific mathematical properties and results in different image decompositions.

6. *Ehlers*: The Ehlers method has the advantage of preserving the spectral information of the MS channels, while refining the spatial resolution based on the panchromatic image (EHLERS, 2004; EHLERS et al., 2006). To generate the high resolution fused composition, a series of algorithms are applied on the original MS and panchromatic bands. First, the spectral and spatial attributes are separated in different components, then the spatial information (PAN) is altered to allow adaptive enhancement of the images (Figure 2.3).

A multiple IHS transformation is applied over the MS images until the total number of bands is exhausted, since the usual IHS algorithm is limited to three MS bands (RGB). After a Fast Fourier Transformation (FFT), the intensity component (I) is filtered using a Low Pass Filter (LPF) and the panchromatic image (PAN) is filtered with a High Pass Filter (HPF). These images are converted back into the spatial domain using an inverse Fast Fourier

Transformation ( $\text{FFT}^{-1}$ ) and combined to generate a fused intensity channel. An inverse IHS transformation ( $\text{IHS}^{-1}$ ) is performed to produce the final fused image that contains the spatial information from the panchromatic image and the spectral resolution of the multispectral image (Figure 2.3).

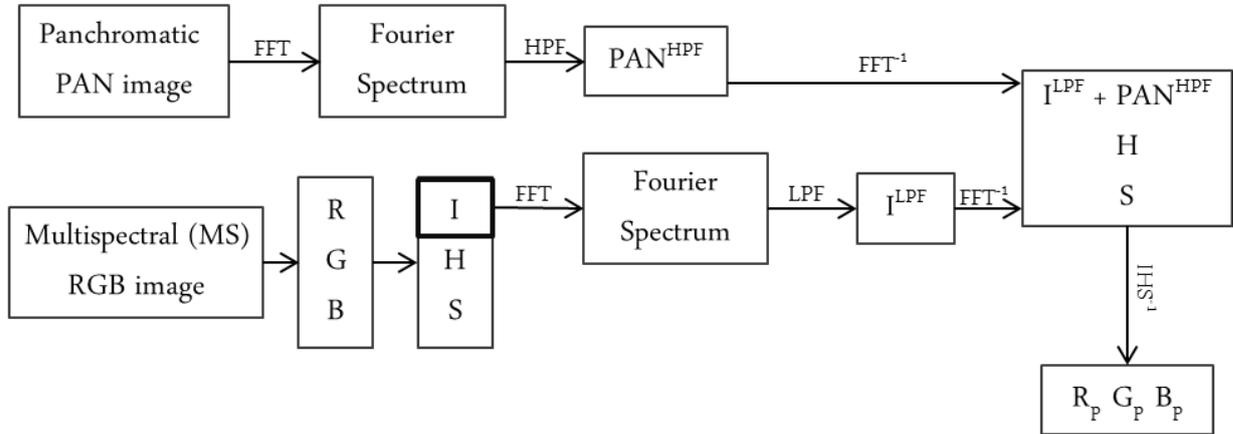


Figure 2.3 - Flowchart illustrating the Ehlers pansharpening procedure (EHLERS et al., 2006).

Different fusion methods were used in order to generate seven fused MS composition for each studied area considering the following algorithms: Brovey (Br.); Principal Components (PC); Gram Schmidt (G.S.); Intensity Hue and Saturation (IHS); Discrete Wavelet Transform (DWT); À Trous Wavelet Transform (ATWT) and Ehlers (Eh). All resulting images are presented in 8 bits format (0 to 255 gray values), which allows the standardization of results for quantitative comparison and unsupervised classification.

### 2.2.3. Qualitative Assessment

The qualitative approach was carried out to evaluate visually the quality of fused MS images, considering the reference MS composition as an optimal scenario. The qualitative assessment refers to a subjective approach and depends on the experience of the person who is analyzing the images (RODRIGUEZ-GALIANO et al., 2012). A series of variables can be used to assess the quality of the images. In this study we considered variations in tone, contrast, saturation, sharpness and texture of the fused images compared to the reference image.

#### 2.2.4. Quantitative Assessment

A zonal quantitative investigation was carried out by fusion quality evaluating algorithms regarding the assessment of spectral distortions: Correlation Coefficient (CC) (ZHANG, 2008); Erreur Relative Globale Adimensionnelle de Synthèse (ERGAS) proposed by Wald (2000); Structural Similarity Index (SSIM) (WANG et al., 2004) and spatial distortions: Spatial Quality Metric index (SM) (OTAZU et al., 2005).

##### 1. Correlation Coefficient (CC)

Given two images  $x$  and  $y$  the correlation coefficient CC is given as:

$$CC(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad \text{Eq. 2.1}$$

Where  $\bar{x}$  and  $\bar{y}$  are the mean gray values of the reference and the fused bands respectively, and CC is estimated globally for each MS band. The result of this equation shows similarity in the small structures between the original and fused images. The resulting values ranges from -1 to 1.

##### 2. ERGAS

ERGAS (Equation 2.2) is an acronym in French for “*Erreur relative globale adimensionnelle de synthese*” which translates to relative dimensionless global error in synthesis. ERGAS is used to calculate the amount of spectral distortion between fused and reference images and is given by the following equation:

$$ERGAS = 100 \frac{h}{l} \sqrt{\frac{1}{N} \sum_{n=1}^N \left( \frac{RMSE_{(n)}}{\mu_0(n)} \right)^2} \quad \text{Eq. 2.2}$$

Where  $h/l$  is the ratio between the spatial resolution of the panchromatic image and the MS image,  $N$  is the number of spectral bands of the fused image,  $\mu_0$  is the mean value of each spectral band, and RMSE (relative root mean square error) represents the difference of standard deviation and mean of the fused and the original image. The best possible value for ERGAS is zero. According to experiments carried out by Wald et al. (1999) an ERGAS value below 3 corresponds to fused images of satisfactory quality.

### 3. Structural Similarity Index (SSIM)

The Structural Similarity index (Equation 2.3) was proposed by Wang et al. (2004) and takes into account the luminance, contrast and structure differences, between each band of the fused and original multispectral images, resulting in a global index representing the fusion quality.

$$SSIM(x, y) = \frac{(2 * \mu_x * \mu_y + C_1)(2 * \sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad \text{Eq. 2.3}$$

Where  $x$  and  $y$  are two non-negative image signals (grey value) for fused and reference bands, respectively.  $\mu_x$  and  $\mu_y$  are the mean luminance of images  $x$  and  $y$ , while  $\sigma_{xy}$ ,  $\sigma_x$  and  $\sigma_y$  are the covariance and the variances of image  $x$  and  $y$ , respectively.  $C_1$  and  $C_2$  are constants defined by Wang et al. (2004). The index result takes values between 0 and 1. The closer the SSIM index to one the better the fused image.

### 4. Spatial Quality Metric Index (SM)

The extraction of spatial features is necessary before carrying out the SM algorithm. This procedure is done by applying a Laplacian filter over all fused and reference MS bands. The correlation coefficient (CC) is then estimated for corresponding bands and the derived values are averaged resulting in the overall spatial measure. The SM index is mathematically expressed as follow (Equation 2.4):

$$SM(x, y) = \frac{1}{N} \sum_{i=1}^n CC_i \quad \text{Eq. 2.4}$$

Where  $N$  is the total number of MS bands and  $CC$  represents the correlation coefficient obtained by the Equation 2.5:

$$CC_i(X, Y) = \frac{\sigma_{Xi Yi}}{(\sigma_{Xi} \sigma_{Yi})} \quad \text{Eq. 2.5}$$

$$Xi = X * l; \quad Yi = Y * l$$

Where  $X$  and  $Y$  are the  $i_{th}$  bands in the reference and fused images, respectively,  $*$  is the convolution operator,  $l$  is a Laplacian filter, set by Otazu et al. (2005) as:

$$l = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

The spatial quality (SM) results in values within a range of -1 to 1 where 1 indicates the best spatial quality.

The quantitative evaluation algorithms (Equation 2.1 to Equation 2.4) result in a single quality index for each fused MS band. Nevertheless, local variations in fused bands might occur due to differences related to specific image features, depending on the fusion algorithm. Thus, the numerical expression of the fusion quality by means of a global value might lead to biased results, especially considering fusion methods that cause a high standard deviation between the reference and the fused bands. Given this, the zonal approach applied in this study, have generated a range of values for each fused band. The results of the quantitative approach evolving each fusion method were compared by descriptive and multivariate statistics (Cluster Group Analysis), taking into account the spatial behavior of the quantitative index, considering the diversity of targets identified in both studied areas.

A principal component analysis (PCA) was carried out between the covariance (CC) and SSIM indexes in order to estimate the variance among the fusion methods for each fused composition in the two studied areas. After PCA analysis, we were able to identify the quantitative index that would be grouped together to run the grouping cluster analysis. Given this, the index results were standardized and integrated (PC and IHS methods were not used due to the high standard deviation when compared to the other methods) to generate the cluster group map. We divided the groups into four levels organized by overall quality for the Landsat/HRC and Landsat/OLI fused images in Amazon and Zambia region. The cluster loadings were associated to each sampling window with values ranging from 0 to 3 (low to high quality, respectively). The resulting cluster maps were obtained by ordinary kriging interpolation of the cluster values associated to the sampling windows.

### ***2.2.5. Quality Evaluation by Unsupervised Classification***

The indirect quantitative analysis was carried out by comparing the classification results of the reference and fused images through the application of unsupervised classification algorithms (CONGALTON, 1991; CONGALTON; GREEN, 1999; RODRIGUEZ-GALIANO et al., 2012). The ISODATA unsupervised classification (LILLESAND et al., 1987), was applied in order to classify the fused images. The reference MS composition was classified using the same algorithm, with the same parameters and the resulting classified image was used as truth image. Classification accuracy has been assessed by calculating the confusion matrix and the Kappa index (CONGALTON, 1991; CONGALTON; GREEN, 1999).

## 2.3. Results

The selected fusion algorithms comprise easy to implement methods, extensively applied to merge moderate and high resolution satellite images. These methods range from simple arithmetic operations to complex algorithms, based on the application of convolutions and filters. A significant difference between fused MS compositions was observed, which emphasizes the effectiveness of the quality assessment by a zonal approach, which adopts local sampling windows to estimate quantitative.

### 2.3.4. Qualitative Assessment

The MS compositions (Landsat OLI) fused with the panchromatic bands CBERS/HRC and PAN/OLI, for both investigation areas, are referred as Landsat/HRC and Landsat/OLI, respectively. The image fusion process has generated a great number of MS compositions (14 fused MS images for each area). The qualitative assessment was carried out by means of a general inspection in order to evaluate, visually, the quality of the fused compositions. After analyzing the images, the most significant results were selected to describe the algorithms that caused significant visual distortions.

*Amazon Site:* The dense vegetation cover and the drainage network are the most remarkable features in this area. The E.h. and DWT methods resulted in less spectral distortion for both fused compositions (Landsat/HRC and Landsat/OLI), and a significant change in spatial quality (blurred effect) was observed, mostly regarding water bodies and sand banks. The ATWT and G.S. methods returned satisfactory results enhancing the spatial quality and maintaining the spectral pattern, for the Landsat/HRC fusion. However, the Landsat/OLI fused images have significant spectral distortion, due to the spectral resolution of the Landsat panchromatic band (0.50 to 0.68  $\mu\text{m}$ ), which caused the loss of spectral information of the infrared OLI channels. The HSV, PC and CN methods refined the spatial resolution but caused a significant negative impact on the image spectral quality, causing color saturation and contrast decreasing.

*Zambia Site:* In Zambia region the overall visual quality of fused MS images was better when compared to the results obtained in Amazon, probably due to soil coverage characterized by sclerophyllic vegetation (shrubs and grasses) and a vast area of bare soils and flooded lands. The land features have a lower overall reflectance in infrared range of the electromagnetic spectrum, which causes a lower influence of NIR and SWIR channels on the merging process. We observed similar results for Landsat/HRC and Landsat/OLI fused compositions, highlighting a good performance of E.h., ATWT and G.S. methods. The HSV,

PC and CN methods have returned unsatisfactory results due to the same reasons observed in Amazon site.

Taking into account the qualitative assessment of fused images, we observed that the best results were achieved for the E.h. and ATWT methods. The DWT, G.S. and PC methods might generate satisfactory results, depending on the area and the images used in the fusion process. Ultimately, the IHS and B.r. methods resulted in images with significant spectral distortions for all fused compositions.

### **2.3.2. *Quantitative Assessment***

The quantitative assessment depends upon the availability of a reference image that can be obtained according to method presented in Figure 2.2. After the fusion procedure, a statistical global parameter is obtained from the fused and the reference MS images. Traditionally, the statistical results are presented as unique values for each fused band (CONGALTON, 1991; CONGALTON, 1999; ALPARONE et al., 2007). A mask was applied in order to generate statistical parameters in zonal sampling windows (each sampling window has an area of 200 m<sup>2</sup>, which covers 255 pixels for Landsat OLI multispectral bands). The obtained spatially dependent statistical index varies according to the spectral and spatial behavior of the image features. Moreover, the results are expressed as a range of values instead of just one global value, allowing a more accurate investigation of the quality of the fused compositions.

The per-band analysis was carried out by evaluation of correlation coefficients (Table 2.2), considering the global average. High correlation values were observed for the synthetic Red/NIR fused band (B2, Table 2.2), followed by the SWIR1/SWIR2 band (B3, Table 2.2). The worst performance was observed for the Blue/Green band (B1, Table 2.2), in Zambia site, probably due to the interference of atmospheric aerosols and the high reflectance related to the blue and green ranges of the electromagnetic spectrum (extensive area of bare soils). With regards to differences between fusion algorithms, the Eh. method returned the best result followed by the ATWT method. The IHS and Br. algorithms are the less correlated to the MS reference image with a significant deviation for B1 and B3. Ultimately, the DWT, G.S. and PC methods had similar results, for all fused MS images in the two studied areas.

Table 2.2 - Band by band correlation values between the original multispectral image and the degraded fused images.

	<b>Zambia Site</b>						<b>Amazon Site</b>					
	Landsat/HRC			Landsat/OLI			Landsat/HRC			Landsat/OLI		
	B1*	B2	B3	B1	B2	B3	B1	B2	B3	B1	B2	B3
Eh.	0.93	0.91	0.87	0.52	0.92	0.53	0.95	0.87	0.94	0.97	0.89	0.96
ATWT	0.53	0.89	0.53	0.53	0.91	0.54	0.75	0.85	0.76	0.76	0.88	0.76
G.S.	0.53	0.88	0.53	0.53	0.90	0.53	0.75	0.85	0.76	0.76	0.85	0.76
PC	0.53	0.89	0.53	0.53	0.90	0.53	0.74	0.78	0.75	0.74	0.79	0.75
DWT	0.49	0.84	0.51	0.52	0.89	0.53	0.72	0.81	0.74	0.73	0.85	0.75
IHS	0.38	0.64	0.40	0.39	0.66	0.42	0.69	0.60	0.44	0.28	0.65	0.47
Br.	0.15	0.44	0.34	0.15	0.44	0.28	0.19	0.35	0.67	0.28	0.52	0.22

\*Synthetic bands: B1-blue/green; B2-red/infrared; B3: SWIR 1/SWIR 2.

The quantitative assessment of fused images in different scenarios is an openly debated topic, considering that the studies concerning quantitative evaluation of fused images, have not established yet which algorithm would be more accurate (THOMAS; WALD, 2005; KHALEGHI et al., 2013). The proposal of spatially dependent evaluation algorithms helps on a better comparison between methods and a detailed visualization of the behavior of each method according to different soil covers. The zonal results of Correlation Coefficient (CC), ERGAS, SSIM and SM are shown in Figure 2.4 and Table 2.3. A range of values (2000 sampling windows) were generated according to the zonal sampling windows, band by band, referent to each image fusion method. The descriptive statistic was applied in order to evaluate qualitatively the fused images (Table 2.3; Figure 2.4).

Table 2.3 – Descriptive statistics (arithmetic mean, variance and standard deviation) for the fused images according to the four quantitative evaluation algorithms.

		Amazon (Landsat/HRC)			Zambia (Landsat/HRC)			Amazon (Landsat/OLI)			Zambia (Landsat/OLI)		
		Mean	Var.	SD									
Correlation	ATWT	0.24	0.05	0.23	0.47	0.02	0.14	0.39	0.05	0.22	0.74	0.01	0.09
	PC	0.28	0.08	0.27	0.40	0.02	0.15	0.20	0.02	0.14	0.57	0.01	0.11
	DWT	0.27	0.07	0.26	0.48	0.02	0.13	0.34	0.05	0.22	0.63	0.01	0.12
	Eh.	0.29	0.04	0.19	0.50	0.01	0.12	0.43	0.02	0.15	0.78	0.01	0.08
	G.S.	0.25	0.08	0.28	0.41	0.02	0.15	0.27	0.05	0.23	0.66	0.01	0.11
	IHS	0.26	0.07	0.26	0.42	0.03	0.16	0.35	0.04	0.20	0.71	0.01	0.09
	Br.	-0.21	0.09	0.31	0.41	0.02	0.15	-0.24	0.07	0.26	0.66	0.01	0.11
ERGAS	ATWT	2.38	3.21	1.79	6.83	0.80	0.90	2.29	3.00	1.73	6.83	0.80	0.90
	Br.	226.37	348.20	59.01	125.70	8.66	2.94	74.15	298.96	17.29	125.70	8.66	2.94
	DWT	2.68	5.13	2.27	8.15	54.09	7.35	9.03	181.60	42.63	8.15	54.09	7.35
	Eh.	0.16	0.01	0.12	0.06	0.00	0.02	0.10	0.00	0.05	0.06	0.00	0.02
	G.S.	2.60	3.62	1.90	6.78	0.54	0.74	2.52	1.83	1.35	6.78	0.54	0.74
	IHS	6.70	656.69	25.63	16.01	278.23	16.68	4.17	6.54	2.56	16.56	490.91	22.16
	PC	3.78	3.08	1.75	6.79	0.52	0.72	3.41	7.21	2.68	6.79	0.52	0.72
SSIM	ATWT	0.78	0.01	0.08	0.65	0.00	0.03	0.80	0.01	0.09	0.66	0.00	0.03
	Br.	0.39	0.00	0.05	0.48	0.00	0.02	0.54	0.00	0.05	0.49	0.00	0.01
	DWT	0.76	0.01	0.09	0.61	0.00	0.05	0.78	0.01	0.09	0.64	0.00	0.05
	Eh.	0.91	0.00	0.06	0.90	0.00	0.03	0.93	0.00	0.07	0.66	0.00	0.03
	G.S.	0.78	0.01	0.08	0.65	0.00	0.02	0.78	0.01	0.07	0.66	0.00	0.03
	IHS	0.60	0.01	0.07	0.47	0.00	0.04	0.62	0.01	0.10	0.49	0.00	0.05
	PC	0.76	0.01	0.07	0.65	0.00	0.02	0.76	0.01	0.07	0.66	0.00	0.03
SM	ATWT	0.12	0.02	0.15	0.29	0.01	0.11	0.24	0.03	0.18	0.42	0.01	0.09
	Br.	0.10	0.02	0.12	0.29	0.01	0.11	0.13	0.02	0.15	0.33	0.01	0.09
	DWT	0.12	0.02	0.14	0.29	0.01	0.10	0.28	0.02	0.15	0.35	0.01	0.09
	Eh.	0.07	0.01	0.10	0.29	0.01	0.11	0.23	0.01	0.11	0.46	0.01	0.11
	G.S.	0.10	0.02	0.13	0.29	0.01	0.11	0.23	0.02	0.15	0.38	0.01	0.11
	IHS	0.11	0.02	0.13	0.26	0.01	0.10	0.34	0.02	0.15	0.38	0.01	0.09
	PC	-0.10	0.03	0.16	0.29	0.01	0.12	-0.14	0.03	0.17	0.39	0.01	0.11

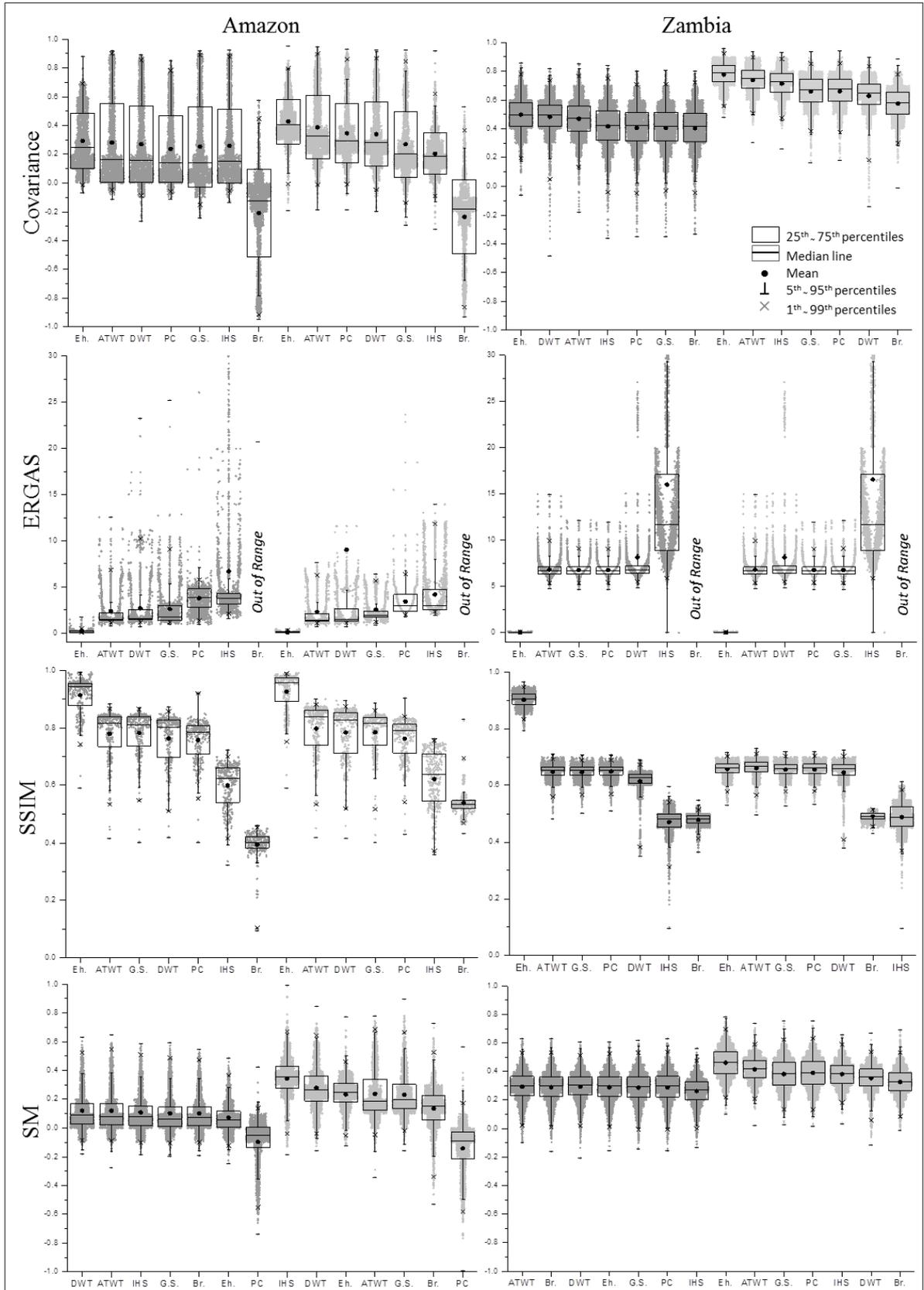


Figure 2.4 - Box-plots with jitters. Dark gray jitters (Landsat/HRC); Light gray jitters (Landsat/OLI). The “out of range” boxes represent Br. algorithm with high ERGAS values.

The merging procedure (Table 2.3) has generated a great volume of data. Thus, an easy way to verify and to compare the quality of the fused images is through the interpretation of box-plots (Figure 2.4) with *jitters* (grey dots: Figure 2.4). Based on a quick inspection of the plots we were able to identify and group the fused images by quality. A similar behavior was observed on the fused images Landsat/HRC and Landsat/OLI in both Amazon and Zambia regions, even considering an image ratio of 6 for the Landsat/HRC against 2 for the Landsat/OLI. However, the Landsat/OLI images have slightly better SM coefficients when compared to Landsat/HRC images.

Analyzing the covariance in Amazon area, we observed a biased behavior with a significant difference between mean and median, for all fused images (Figure 2.4). The distribution is skewed downward the median (*jitters*) with a few larger values closer to 1. The SSIM for Amazon has the opposite behavior and most of the measured data trends to values closer to one, which can be clearly inferred by the *jitters*, interloped under the boxplots. However, in both charts we observed a similar behavior regarding the overall accuracy of the merging algorithms. The ERGAS algorithm maximized the difference between methods, with a considerable high deviation for IHS and Br. in both studied areas (Table 2.3; Figure 2.4), but the mean and median values are always below 20. The SSIM and ERGAS indexes have shown the same pattern in Amazon and Zambia region with good performances for Eh. and ATWT methods and intermediated values for G.S., DWT and PC. Unsatisfactory results were obtained for IHS and Br. methods due to color saturation and introduction of chromatic distortions (Table 2.3; Figure 2.4).

The SM index indicates the amount of spatial detail of the panchromatic image injected on the MS composition. Concerning this index, we observed a different pattern when compared to the other evaluation indexes (covariance, ERGAS and SSIM). The overall spatial quality of the Landsat/OLI images was better than Landsat/HRC images in Amazon region and very similar in Zambia region. The best methods were ATWT, DWT and IHS in Amazon, but in Zambia all methods resulted in very close indexes with a better performance for the Eh. method (Landsat/OLI fusion). The quantitative evaluation was useful on highlighting the statistical behavior of the fused images according to each method. However, we were not able to identify how the lateral variation of the image features interferes on the merging process. Thus, we applied a grouping cluster analysis in order to evaluate the interference of soil cover in the fused images.

The spatial variability of the fusion quality follows the same pattern in both studied areas with a higher performance in homogenous regions of vegetation and lower fusion

quality in zones of water bodies and bare soil. In Amazon region the higher quality is related to the rainforest cover, while the lower values occur in zones of water bodies and in fluvial sand deposits of the Rio Negro (Figure 2.5a). In Zambia the lateral variability of the fusion quality is more complex, with satisfactory values associated to extensive areas of dryer grasslands and pastures and lower quality in zones of flooded depression (Figure 2.5c and 2.5d). The results presented in the cluster maps were useful on describing the quality of the fused compositions according to different soil cover and to highlight the similarity between the Landsat/HRC and Landsat/OLI images. The fusion trends to better results in areas of homogeneous features (grasslands and rainforest) and has lower quality in zones with higher variety of targets (river/lake borders and regions of human intervention: crops and roads).

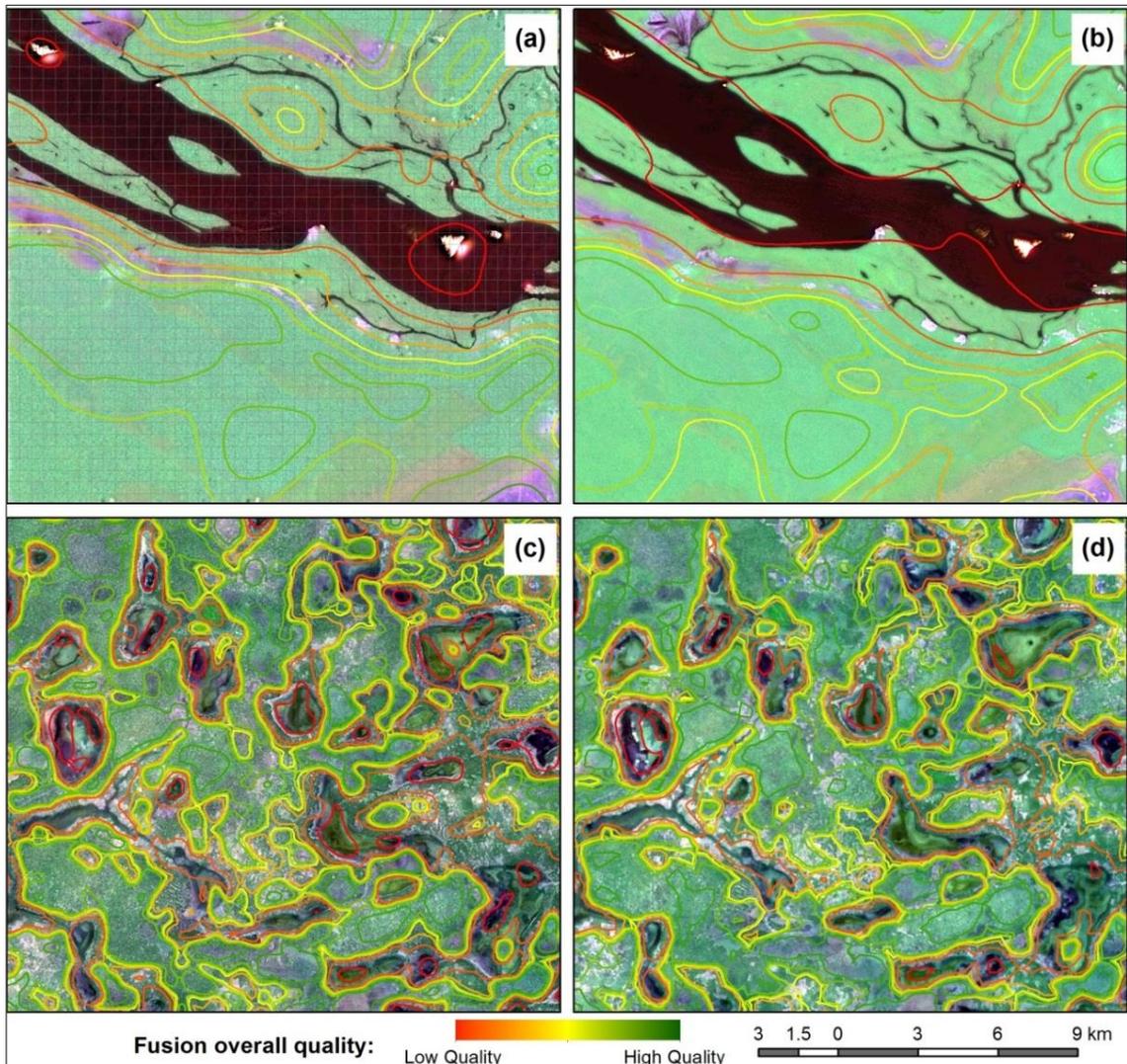


Figure 2.5 - Overall quality of the fusion methods ATWT, Eh., DWT, GS and PC grouped by clusters. The MS compositions illustrate the ATWT fused images: (a) Amazon – Landsat/HRC; (b) Amazon – Landsat/OLI; (c) Zambia – Landsat/HRC and (d) Zambia – Landsat/OLI. A representation of the zonal sample windows is shown in Figure 2.5a, as light gray lines.

### 2.3.3. Indirect Quantitative Assessment

The realizing of 5-m spatial resolution HRC panchromatic bands in 2015 (CBERS-4) by free, represents an important step on the understanding of natural environments in regional and local map scales in South America and Africa. Therefore, the evaluation of classified fused MS compositions is a key aspect to describe the use of fused data from different sensor systems to map soil cover in natural regions. Given this, we carried out an ISODATA unsupervised classification with the same parameters for the reference and the degraded fused images, in both studied areas (LILLESAND et al., 1987). The parameter used to access the quality of the classified images was relative to the ground truth image (Landsat/OLI MS composition) that was obtained by a fully automated process. Thus, the classes of the reference image have not specified topology according to the soil cover, but represent the classes obtained by the ISODATA classification of the reference image.

Table 2.4 - Overall accuracy (O.A. in %) and kappa coefficient for the classified images according to different fusion methods.

	Amazon: Landsat/HRC		Amazon: Landsat/OLI		Zambia: Landsat/HRC		Zambia: Landsat/OLI	
	O.A.	Kappa	O.A.	Kappa	O.A.	Kappa	O.A.	Kappa
ATWT	60.32	0.48	60.23	0.47	34.08	0.20	50.73	0.41
Br.	6.16	-0.04	3.36	-0.05	26.05	0.10	30.33	0.16
DWT	58.67	0.33	55.04	0.26	29.29	0.14	32.56	0.18
Eh.	73.22	0.55	77.83	0.62	39.40	0.26	59.66	0.51
G.S.	44.21	0.27	29.37	0.12	25.14	0.09	30.46	0.16
IHS	41.07	0.21	14.24	0.01	25.15	0.09	25.73	0.09
PC	48.64	0.30	39.87	0.20	32.44	0.18	49.37	0.39

We specified a maximum of 7 and a minimum of 5 classes and the process has generate 6 classes within the reference and all fused images for ISODATA classification of the Amazon area (Table 2.4). In this region the O.A. and kappa coefficient are higher for the Landsat/HRC images, due to the high reflectance of the Rainforest that is better represented in the CBERS HRC panchromatic band. Therefore, the overall results of Landsat/HRC are higher than the ones obtained for the Landsat/OLI images (Table 2.4), with satisfactory values related to the Eh. and ATWT methods. The other fusion algorithms have lower O.A. and kappa values with considerable deviation for Br. (Landsat/HRC) and IHS (Landsat/OLI) algorithms (Table 2.4).

In Zambia region, a maximum of 8 and a minimum of 5 classes were specified and the ISODATA classification resulted in 6 classes for all images. The overall quality of the classified images is lower if compared to Amazon area, due to the complexity of the soil cover with a higher variety of image features. In Zambia the Landsat/OLI images resulted in better O.A. and kappa values. The reflectivity of bare soils and branches of herbaceous

vegetation is considerably higher in VIS (blue/green) OLI bands, which would result in good fused images for Landsat/OLI fusion due to the spectral range of the PAN OLI band (0.50-0.68  $\mu\text{m}$ ). In this region the better results were also obtained for the algorithms Eh. and ATWT, but the overall O.A. and kappa values are lower with a weak correlation to the reference classified image in both fused compositions (Landsat/HRC and Landsat/OLI). This behavior can be explained by the complexity of the soil cover with a higher variety of targets, if compared to Amazon region (see Figure 2.5).

The results obtained by the indirect quantitative assessment have a significant difference between fusion methods with O.A. values ranging from 73% (Eh. in Amazon) to 3% (Br. in Amazon). However, there is a great similarity between Landsat/HRC and Landsat/OLI fused compositions, with better results for Landsat/HRC images in vegetated areas (Amazon) and Landsat/OLI in regions of bare soils (Zambia). Thus, CEBERS/HRC panchromatic images can be applied to generate local and regional soil cover maps, through the fusion with MS Landsat OLI images.

We observed that some fusion methods might causes a significant deviation from the arithmetic mean related to color saturation and introduction of chromatic distortions associated to some specific image targets. Therefore, a detailed spatial dependent evaluation (zonal approach) of the fusion quality is interesting to compare algorithms and to highlight where the fusion process returned better results. Besides, a zonal quantitative assessment gives us statistical parameters that allow the application of basic statistic or multivariate statistic in order to compare fusion methods.

The zonal assessment of the fusion quality in Amazon and Zambia sites allowed us to the divide the resulting fused images into three groups: (1) satisfactory results obtained by Eh. and ATWT algorithms; (2) intermediary results for G.S., PC and DWT algorithms and (3) poor results for IHS and Br. algorithm in all evaluated fused images. However, we observed that the ATWT algorithm introduces more spatial information of the panchromatic band if compared to Eh. algorithm, which might causes a blurred effects in Eh. fused compositions. This aspect has to be considered when applying fusion techniques considering that Eh. algorithm might causes the loss of spatial information from the panchromatic band.

## 2.4. Conclusions

In this paper we evaluated the quality of fused MS compositions obtained according to different fusion techniques, through the assessment of their accuracies according to qualitative and quantitative approaches based on a zonal investigation. By the proposed method we have identified how different images features are affected by the application of fusion algorithms, considering areas of Amazon Forest and Zambia Grasslands. We observed that local distortions in fused bands might occur due to differences related to specific image targets, depending on the applied fusion algorithm. Thus, the numerical expression of the fusion accuracy by means of a global averaged value might lead to biased results, especially considering fusion methods that cause a high standard deviation between the reference and the fused MS images. Moreover, we were able to apply cluster grouping analysis based on the zonal quality indexes.

The availability of optical satellite images provided by free in Latin America and Africa is essential in order to enhance regional and local maps in remote areas. The results presented in this research have shown that high-resolution HRC CBERS-4 image can be satisfactorily applied in substitution of Landsat OLI panchromatic bands, by generating sharpened MS composition with a 5-m spatial resolution using ATWT and E.h fusion algorithms. CBERS-4 satellite will provide panchromatic images with a 5-m spatial resolution covering the VIS and NIR ranges of the electromagnetic spectrum that would be properly fused with VIS and IR bands of Landsat OLI sensor.

## References

- ABDON, M.M.; OLIVEIRA, M.; LUCIANO, A.C.S.; SILVA, J.S.V. Identificação e mapeamento de pastagens degradadas nos municípios de Corguinho e Rio Negro, MS, utilizando fusão de imagens CBERS-2B -CCD e HRC). In: SIMPÓSIO DE GEOTECNOLOGIAS NO PANTANAL, 22., 2009, Corumbá. **Proceedings...** Corumbá: Embrapa Pantanal, 2009. p. 343-352.
- ADLER-GOLDEN, S.M.; MATTHEW, M.W.; BERNSTEIN, L.S.; LEVINE, R.Y.; BERK, A.; RICHTSMEIER, S.C.; ACHARYA, P.K.; ANDERSON, G.P.; FELDE, G.; GARDNER J.; HOKE, M.; JEONG, L.S.; PUKALL, B.; RATKOWSKI, A; BURKE, H.K. Atmospheric correction for shortwave spectral imagery based on MODTRAN4. In: IMAGING SPECTROMETRY, 5., 1999, Denver, Colorado. **Proceedings...** Bellingham, WA: SPIE, 1999. p. 61-69. (Proceedings of SPIE, v. 3753).
- ALPARONE, L.; WALD, L.; CHANUSSOT, J.; THOMAS, C.; GAMBA, P.; BRUCE, L. M. Comparison of Pansharpening Algorithms: Outcome of the 2006 GRS-S Data-Fusion Contest. **IEEE Transactions on Geoscience and Remote Sensing**, New York, v. 45, p. 3012-3021, 2007.

- BERK, A.G.P.; ADLER-GOLDEN, S.M.; RATKOWSKI, A.J.; FELDE, G.W.; ANDERSON, G.P.; HOKE, M.L.; COOLEY, T.; CHETWYND, J.H.; GARDNER, J.A.; MATTHEW, M.W.; BERNSTEIN, L.S.; ACHARYA, P.K.; MILLER, D.; LEWIS, P. Exploiting MODTRAN radiation transport atmospheric correction: the FLAASH algorithm. In: INTERNATIONAL CONFERENCE ON INFORMATION FUSION, 5., 2002, Annapolis, MD. **Proceedings...** New York: IEEE, 2002. v. 2, p. 798-803.
- BOOTH, A.; MCCULLUM, J.; MPINGA, J.; MUKUTE, M. **State of the environment in Southern Africa: A report by the Southern African Research and Documentation Centre**. Harare, Zimbabwe: SARDC: IUCN, Regional Office for Southern Africa; Maseru, Lesotho: SADC, Environment and Land Management Sector Coordination Unit, 1994. 332 p.
- BRASIL. Ministério das Minas e Energia. **Projeto RADAMBRASIL. Folha SA.19-Içá-AM: Geomorfologia**. Rio de Janeiro, 1977. p. 125-180. (Levantamento dos Recursos Naturais, v. 14).
- CARLOTTO, M. J. Reducing the effects of space-varying, wavelength-dependent scattering in multispectral imagery. **International Journal of Remote Sensing**, London, v. 20, p. 3333-3344, 1999.
- CHAVEZ, P.S.; SIDES, S.C.; ANDERSON, J.A. Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic. **Photogrammetric Engineering & Remote Sensing**, Washington, DC, v. 57, p. 295-303, 1991.
- CHENG, W.; ZHOU, J.; YANG, C.; ZHOU, W. Analysis of Image Fusion of TM and CEBERS Based on Pixel Level. **Procedia Environmental Sciences**, Amsterdam, v. 10, p. 1674-1679, 2011.
- CHOI, M.A. New Intensity-Hue-Saturation Fusion Approach to Image Fusion With a Tradeoff Parameter. **IEEE Transactions on Geoscience and Remote Sensing**, New York, v. 44, n. 6, p. 1672-1682, 2006.
- CONGALTON, R. A review of assessing the accuracy of classifications of remotely sensed data. **Remote Sensing of Environment**, New York, v. 37, p. 35-46, 1991.
- CONGALTON, R.G.; GREEN, K. **Assessing the Accuracy of Remotely Sensed data: principles and practices**. Boca Raton: CRC Press, 1999. 160 p.
- DAL'ASTA, A.N.; BRIGATTI, N.; AMARAL, S.; ESCADA, M.I.S.; MONTEIRO, A.M.V. Identifying Spatial Units of Human Occupation in the Brazilian Amazon Using Landsat and CBERS Multi-Resolution Imagery. **Remote Sensing**, Basel, v. 4, p. 68-87, 2012.
- DONG, Z.; WANG, Z.; LIU, D.; ZHANG, B.; ZHAO, P.; TANG, X.; JIA, M. SPOT5 multi-spectral (MS) and panchromatic (PAN) image fusion using an improving wavelet method based on local algorithm. **Computer & Geosciences**, Amsterdam, v. 60, p.134-141, 2013.
- EHLERS, M. Spectral characteristics preserving image fusion based on Fourier domain filtering. In: REMOTE SENSING FOR ENVIRONMENTAL MONITORING, GIS APPLICATIONS, AND GEOLOGY, 4., 2004, Bellingham, WA. Bellingham, WA: SPIE, 2004. 13 p. (Proceedings of SPIE, v. 5574).

EHLERS, M.; GREIWE, A.; TOMOWSKI, D. On Segment Based Image Fusion. In: INTERNATIONAL CONFERENCE ON OBJECT-BASED IMAGE ANALYSIS – OBIA, 1., 2006, Salzburg. **Proceedings...** Salzburg, 2006. p. 11.

FAUNDEEN, J.L.; WILLIAMS, D.L.; GREENHAGEN, C.A. Landsat yesterday and today: An American vision and an old challenge. **Journal of Map & Geography Libraries**, London, v. 1, n. 1, p. 59-73, 2004.

GALIANO, V.F.R.; PARDO, E.I.; OLMO, M.C.; MATEOS, J.; SÁNCHEZ, J.P.R.; VEGA, M. A comparative assessment of different methods for Landsat 7/ETM+ pansharpening. **International Journal of Remote Sensing**, London, v. 33, n. 20, p. 6574-6599, 2012.

GLOBAL LAND COVER FACILITY - GLCF. **Landsat GeoCover**. Maryland, 2009. Available at: <<http://glcf.umiacs.umd.edu/data/landsat/>>. Accessed in: Sep. 10, 2014.

HUEMMRICH, K.F.; PRIVETTE, J.L.; MUKELABAI, M.; MYNENI, R.B.; KNYAZIKHIN, Y. Time-series validation of MODIS land biophysical products in a Kalahari Woodland. Africa. **International Journal of Remote Sensing**, London, v. 26, p. 4381–4398, 2005.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. Geoscience Division - DGC. Coordenação de Recursos Naturais e Estudos Ambientais - CREN. **Digital Maps of Natural Resources**. Map Scale: 1:250,000. Digital Format: shp. Rio de Janeiro, 2008. Available at: <[ftp://geofp.ibge.gov.br/mapas/banco\\_dados\\_georeferenciado\\_recursos\\_naturais/](ftp://geofp.ibge.gov.br/mapas/banco_dados_georeferenciado_recursos_naturais/)>. Accessed in: Jun. 10, 2014.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS - INPE. **Image Generation Division/DGI**. Sao Jose dos Campos, 2014. Available at: <http://www.dgi.inpe.br/CDSR/>. Accessed in: Sep. 11, 2014.

JIN, C.; XIAO X.; MERBOLD, L.; ARNETH, A.; VEENENDAAL, E.; KUTSCH, W.L. Phenology and gross primary production of two dominant savanna woodland ecosystems in Southern Africa. **Remote Sensing of Environment**, New York, v. 135, p. 189-201, 2013.

KHALEGHI, B.; KHAMIS, A.; KARRAY, F.O.; RAZAVI, S.N. Multisensor data fusion: A review of the state-of-art. **Information Fusion**, Amsterdam, v. 14, p. 28-44, 2013.

EASTMAN KODAK COMPANY. (New York, NY). C.A. Laben, B.V. Brower. **Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening**. US Patent # 6011875. 4 Jun. 2000.

LILLESAND, T.M.; KIEFER, R.W.; CHIPMAN, J.W. **Remote Sensing and Image Interpretation**. 2. ed. New York: John Wiley and Sons, 1987.

MARCELINO, E.V.; FORMAGGIO, A.R.; MAEDA, E.E. Landslide inventory using image fusion techniques in Brazil. **International Journal of Applied Earth Observation and Geoinformation**, Amsterdam, v. 11, p. 181-191, 2009.

- NÚÑEZ, J.; OTAZU, X.; FORS, O.; PRADES, A.; PALA, V.; ARBIOL, R. Multiresolution based image fusion with additive wavelet decomposition. **IEEE Transaction on Geoscience and Remote Sensing**, New York, v. 37, p. 1204-1211, 1999.
- OTAZU, X.; GONZALEZ-AUDICANA, M.; FORS, O.; NUNEZ, J. Introduction of sensor spectral response into image fusion methods. application to wavelet-based methods, **IEEE Transactions on Geoscience and Remote Sensing**, New York, v. 43, n. 10, p. 2376-2385, 2005.
- RANCHIN, T.; WALD, L. Fusion of high spatial and spectral resolution images: the ARSIS concept and its implementation. **Photogrammetric Engineering and Remote Sensing**, Washington, DC, v. 66, p. 49-61, 2000.
- RANCHIN, T.; AIAZZI, B.; ALPARONE, L.; BARONTI, S.; WALD, L. Image fusion-the ARSIS concept and some successful implementation schemes. **ISPRS Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 58, p. 4-18, 2003.
- RODRIGUEZ-GALIANO, V.F.; GHIMIRE, B.; ROGAN, J.; CHICA-OLMO, M.; RIGOLSANCHEZ, J.P. An assessment of the effectiveness of a random forest classifier for land-cover classification. **ISPRS Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 67, p. 93-104, 2012.
- SHAH, V.P.; NICOLAS, H.; YOUNAN, R.; KING, L. An Efficient Pan-Sharpening Method Via a Combined Adaptive PCA Approach and Contourlets. **IEEE Transaction on Geoscience and Remote Sensing**, New York, v. 46, n. 5, p. 1323-1335, 2008.
- THOMAS, C.; WALD, L. Assessment of the Quality of Fused Products. In: EARSel SYMPOSIUM: NEW STRATEGIES FOR EUROPEAN REMOTE SENSING, 24., 2005, Dubrovnik, Croatia. Dubrovnik, Croatia: Millpress, 2005. p. 317-325.
- USGS. Earth Resources Observation and Science. **EROS Center**. Reston, VA, 2010. Available at: <<http://eros.usgs.gov/>>. Accessed in: Mar. 10, 2014.
- XIA, Y.; LEUNG, H. Performance analysis of statistical optimal data fusion algorithms. **Information Sciences**, New York, v. 277, p. 808-824, 2014.
- WALD, L. Some terms of reference in data fusion. **IEEE Transactions on Geoscience and Remote Sensing**, New York, v. 37, p. 1190-1193, 1999.
- WALD, L. Quality of high resolution synthesized images: is there a simple criterion? In: CONFERENCE ON FUSION OF EARTH DATA, 3., 2000, Sophia Antipolis, France. Merging point measurements, raster maps and remotely sensed images. Sophia Antipolis, France: SEE/URISCA, 2000. p. 99-103.
- WALD, L.; RANCHIN, T.; MANGOLINI, M. Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. **Photogrammetric Engineering & Remote Sensing**, Washington, DC, v. 63, p. 691-699, 1997.

WANG, Z.; BOVIK, A. C. A. Universal image quality index. **IEEE Signal Processing Letters**, New York, v. 9, p. 81-84, 2002.

WANG, Z.; BOVIK, A.C.; SHEIKH, H.R.; IMONCELLI, E.P. Image quality assessment: from error visibility to structural similarity. **IEEE Transactions on Image Processing**, New York, v. 13, p. 600-612, 2004.

ZHANG, Y. Methods for Image fusion Quality Assessment – A Review, Comparison and Analysis **The International Achieves of the Photogrammetry, Remote Sensing Information Sciences**, Sydney, v. 37, pt. B7, p. 1101-1110, 2008.

ZHANG, Y. Understanding Image Fusion. **Photogrammetric Engineering and Remote Sensing**, Washington, DC, v. 70, p. 657–661, 2004.

ZHOU, Y.; CIVCO, D.L.; SILANDER, J.A. A wavelet transform method to merge Landsat TM and SPOT panchromatic data. **International Journal of Remote Sensing**, London, v. 19, n. 4, p. 743–757, 1998.

### 3. A multi-sensor approach for mapping plant-derived carbon storage in Amazonian Podzols<sup>2</sup>

#### Abstract

The Rio Negro basin is characterized by the extensive occurrence of Podzol-type soils that store large amounts of organic matter in deep thick spodic horizons, resulting in the storage of great amounts of soil organic carbon that can be mineralized and released to atmosphere with climate change. The quantification of this carbon requires determination of Podzol types and their spatial distribution. Remote sensing techniques would be helpful in indirect spatializing and segmentation of soil groups in Amazon Podzols. Here we associated remote-sensing images (Shuttle Radar Topographic Mission (SRTM), Operational Land Imager sensor/Landsat 8, and Thermal Infrared Sensor/Landsat 8) and field sample data in order to achieve carbon stock mapping. We found that a multi-sensor approach was critical for a proper segmentation of vegetation groups and spatial distribution of areas with different hydrologic soil regimes.

**Keywords:** Landsat 8, Deep-SOC stock, Podzols, Remote Sensing

#### 3.1. Introduction

The Amazon region is in urgent need of detailed soil mapping covering hitherto relatively undiscovered remote areas, such as those located in the high Rio Negro basin. These areas were mapped between 1970 and 1985 within the framework of the RADAMBRASIL project, the first effort aimed at mapping the whole Brazilian Amazon region, through a combination of aerial Synthetic Aperture Radar (SAR) imagery and scarce soil and rock field controls. The resulting maps were published at a scale of 1:1,000,000 and covered the whole Brazilian Amazon forest area. In 2008, the Brazilian Institute of Geography and Statistics (IBGE) provided a new soil mapping of the Amazon region at a scale of 1:250,000 (IBGE, 2008). The RADAMBRASIL data were refined using Landsat imagery and new field data. Although these maps represented a major advance when compared with the RADAMBRASIL maps, they are restricted by a low level of detail, insufficient with regard to requirements related to soil and forest protection and management purposes.

The Rio Negro basin is located in the northwestern part of the central Amazon plain. It mainly consists of an extensive, low-altitude peneplain, some 10–20 m above average river levels (60–90 m above sea level) with scarce relictual inselbergs and mesas. Elementary landscape units are mainly flat plateaux with dispersed and ramified depressions in their centre and, at the edge of the plateaux and in more dissected areas, flat-top to convex hills.

---

<sup>2</sup> Paper published in the International Journal of Remote Sensing (DOI: 10.1080/01431161.2015.1034896).  
Authors : O. J. R. Pereira, C. R. Montes, Y. Lucas, R. C. Santin, A. J. Melfi

Narrow alluvial terraces are observed in major river corridors. Soils in the dissected, better-drained areas are Ferralsols and Acrisols while the plateaux are occupied by Gleyic Plinthosols and hydromorphic Podzols (BRASIL, 1977; DUBROEUCQ; VOLKOFF, 1998; NASCIMENTO et al., 2004). Such soil diversity reflects on the physiognomy and spectral signatures of the vegetation that grades from typical evergreen forest (high rainforest) to forest with a higher density of smaller trees (campinarana) and shrub savannah (campina). Maps from the RADAMBRASIL project (1972–1978) reveal such diversity and the close relationship between vegetation and soil types.

Podzolization in Amazonia has been studied by several authors (LUCAS et al., 1984; DUBROEUCQ; VOLKOFF, 1998; DUBROEUCQ et al., 1999; NASCIMENTO et al., 2004; MONTES et al., 2007; BUENO, 2009), who have shown its importance as the main process of differentiation of Amazonian ecosystems. Podzols develop with time at the expense of clayey soils, constituting an endmember of soil evolution in such areas. The evolution of the soil systems in the Rio Negro basin is closely connected to the geomorphological pattern of the region. Red and yellow clayey Ferralsols and Acrisols are found at the edge of dissected low plateaux; by contrast, Podzols are found in poorly drained depressions in the central parts of the plateaux. In Podzols, the organic matter produced in the topsoil is transferred at depth through sandy eluviated horizons and accumulates at a depth varying from 1 m to more than 10 m, forming thick horizons rich in organic matter, called spodic horizons (Bh). The resultant Podzols can store considerable amounts of carbon at depth and thus represent an important carbon pool at the global scale (MONTES et al., 2011).

The complexity of tropical ecosystems, with their high diversity of plant species, decreases the accuracy level of the quantification of carbon stored in vegetation and soil. Therefore, some studies developed over recent decades have been searching for innovative techniques to spatialize carbon with the aid of multi-criterial analysis, which facilitates the correlation between environmental variables that can be inferred by remote-sensing techniques and the concentration of organic carbon stored in vegetation and soil.

Most spatialization methods disregard the occurrence of soil organic carbon (SOC) stored at depths greater than 0.3 m. According to most studies conducted in the Amazon, the carbon stock in soils was estimated around 6–9.4 kg C m<sup>-2</sup> considering the first 0.3 m of soil (BERNOUX et al., 2002), but recent research (MONTES et al., 2011) showed that in hydromorphic Podzols the carbon stock can exceed 66.7 kg C m<sup>-2</sup> (0–5 m). Such values suggest that the carbon stored in hydromorphic Podzols might represent a significant part of the total organic carbon (TOC) stored in the Amazonian biome. Thus, the

carbon stored in thick spodic horizons should not be overlooked in studies that aim to estimate CO<sub>2</sub> emissions in tropical regions while considering the scenario of global climate change.

Taking into account the scarcity of studies that quantify and spatialize the TOC stored in Amazon Podzols, the present research aimed to propose association methods between biophysical variables derived from remote sensing imagery and field sample data to identify areas of Podzols under different hydrologic soil regimes.

## **3.2. Methodology**

### **3.2.1. Study Area**

The study area was selected after interpretation of spectral vegetation signatures (Landsat/OLI) and with the aid of soil maps (IBGE, 2008). It covers an area of 71 km<sup>2</sup> located north of Barcelos City, Amazon State, Brazil, at the central coordinates 0°15'18"N and 62°46'36"W (Figure 3.1). The geological substratum is the sedimentary cover of the rivers Branco and Negro, with some younger depositional areas surrounding the Demeni river (Holocene alluvium of Demeni river; BRASIL, 1977). Three soil types developed in the area according to IBGE (2008): Ferralsols, Gleysols, and Podzols.

The light green region on the multispectral composition (Figure 3.1) indicates the soil association Podzol/Ferralsol, while the darker green and reddish regions indicate Podzols at different stages of evolution (soil drainage conditions). Considering the difficulties involved in accessing the whole study area, a representative zone of soil lateral variability was selected for field investigations and further extrapolation.

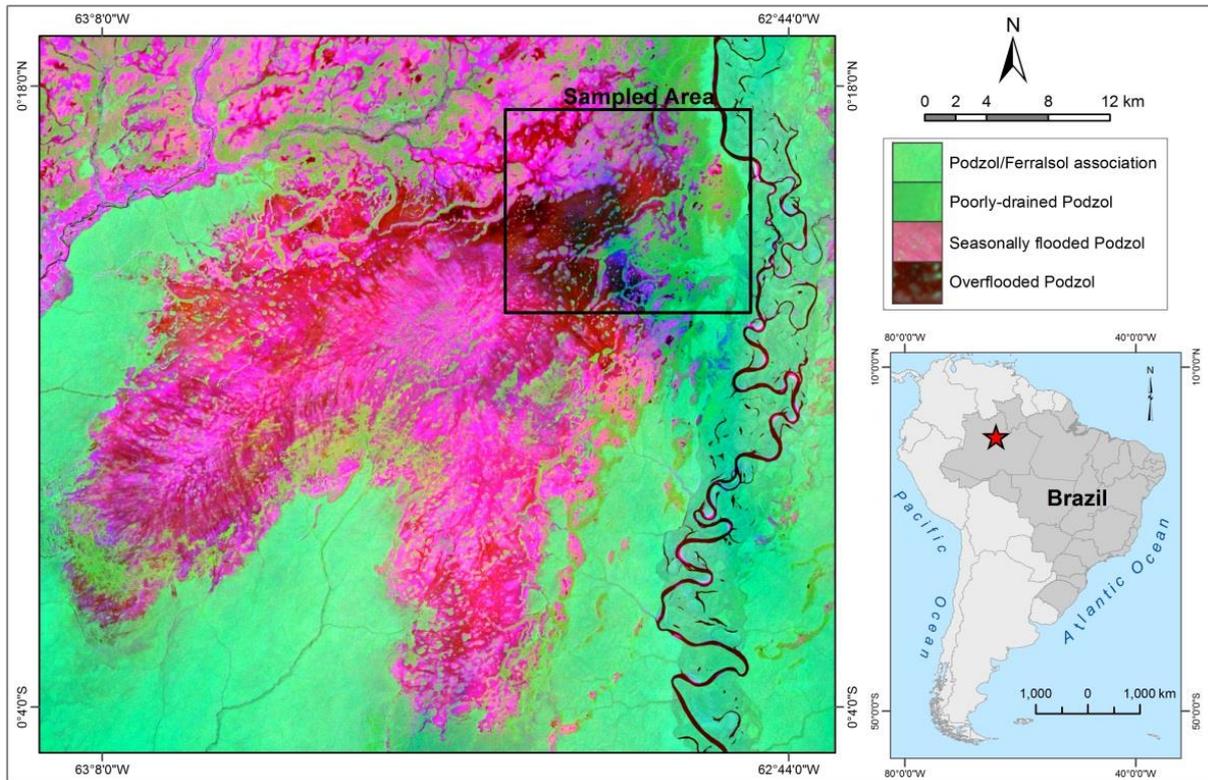


Figure 3.1 - Situation of the study area. The sampled area represents the region where the soil samples were collected. The map illustrates the extrapolation area (multi-sensor composition: Land Surface Temperature, SAVI, and NDMI – R, G, B, respectively).

### 3.2.2. Field data

Ten profiles representative of the three soil units were selected for detailed soil description and sampling. Several other observation points were also selected to determine the occurrence of Podzols. Observations and sampling were done by hand-auger drilling. Casing the auger holes with PVC tubes was necessary because of the collapse of the sandy material overlying the spodic horizons when digging or trading. TOC in samples was measured by the dry combustion technique using a Shimadzu TOC-5000 apparatus.

### 3.2.3. Image data and processing methods

Podzol mapping was achieved using remote sensing images from the Landsat 8 (Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS)) and Shuttle Radar Topographic Mission (SRTM) digital elevation models (DEMs), provided by USGS (2014). The Landsat 8 cloud-free composition (path 233, row 60) was acquired on 25<sup>th</sup> January 2014 at the central coordinates  $63^{\circ}1'55.81''\text{W}$ ;  $0^{\circ}3'33.08''\text{N}$ . The Landsat bands and the SRTM image were registered by image-to-image registration, to the Landsat 7 ETM+ (Enhanced Thematic Mapper Plus) composition corrected and orthorectified (GLCF, 2009). For this composition the positional accuracy on the final image product always has a root mean square

error (RMSE) below 50 m. Given the foregoing, all the derived products were referenced according to the horizontal accuracy achieved based on the Landsat 7 ETM+ composition systematically corrected and orthorectified. After these procedures, the methods shown in Figure 3.2 were adopted.

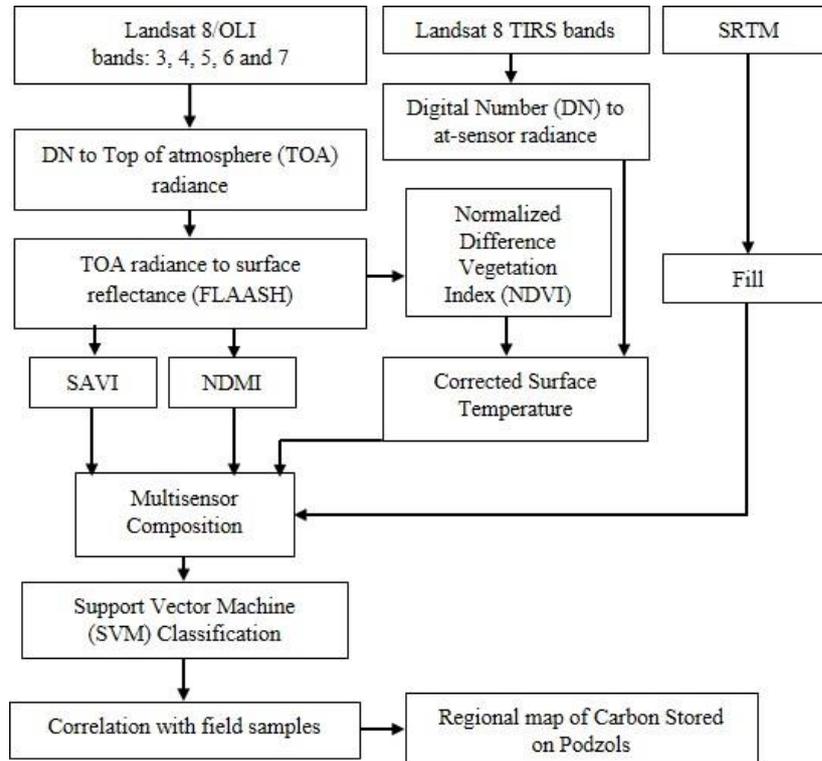


Figure 3.2 - Flow chart showing the methodology employed in this work for generating the regional map of carbon stock. SAVI, soil adjusted vegetation index; NDMI, normalized difference moisture index.

### 3.2.3.1. SAVI and NDMI indices

The soil cover in the study area is a mixture of natural vegetated regions and bare soil, mostly white sand zones, with different moisture concentrations. The soil water content, the phytophysiology, and the soil type vary according to the topography and hydrologic soil regime. Such aspects can be mapped by remote sensing images: the products described below were generated from the atmospherically calibrated Landsat/OLI-TIRS bands and SRTM DEM. All derived products were generated taking into account the lateral variation in biophysical properties of soils and vegetation (e.g. topography, surface moisture, surface temperature, and vegetation density).

$$SAVI = \frac{(1 + L) + (NIR_{\rho_{sur}} - R_{\rho_{sur}})}{(NIR_{\rho_{sur}} + R_{\rho_{sur}} + L)} \quad \text{Eq. 3.1}$$

Where  $L$  (correction factor according to vegetation cover density) is set to 0.5, and  $NIR_{\rho_{sur}}$  and  $R_{\rho_{sur}}$  are, respectively, the NIR and Red OLI bands atmospherically corrected.

$$NDMI = \frac{OLI\ 5 - OLI\ 6}{OLI\ 5 + OLI\ 6} \quad \text{Eq. 3.2}$$

This index contrasts the NIR band ( $OLI\ 5$ ), which is sensitive to the reflectance of leaf chlorophyll content, to the mid-infrared band ( $OLI\ 6$ ), which is sensitive to the absorbance of leaf and soil moisture.

### 3.2.3.2. Surface Temperature

The thermal bands (TIRS/Landsat 8) 10 and 11 were calibrated to surface temperature as follows. The digital numbers (DNs) of the TIRS bands were converted to the top-of-atmosphere (TOA) radiance from Equation 3.3.

$$L_{TOA} = M * DN + b \quad \text{Eq. 3.3}$$

Where  $L_{TOA}$  is the cell value as radiance ( $Wm^{-2} sr^{-1} \mu m^{-1}$ ), DN is the digital number, M is the radiance multiplier (0.0003342), and b is the radiance add (0.1). The equation was applied for both thermal raw TIRS bands (10 and 11).

After conversion of DN to TOA, the surface radiance was calculated using the Reference Channel Emissivity (RCE) method. According to this procedure, all the pixels of the thermal channel have a constant emissivity ( $\epsilon$ ) usually set to 0.95. According to another method described by Sobrino, Jimenez-Munoz and Paolini (2004), the emissivity may depend on the soil cover. It is thus necessary to consider three different cases: (1) bare ground, (2) fully vegetated, and (3) mixture of bare soil and vegetation. The third case was applied in this study, and hence Equation 3.4 was used to calculate the emissivity ( $\epsilon_v$ ):

$$\epsilon_v = 0.004 P_v + 0.986 \quad \text{Eq. 3.4}$$

Where  $\epsilon_v$  is the land surface emissivity (LSE); 0.986 is the standard emissivity value for vegetation and soils; 0.004 is the mean standard deviation value for the emissivity of soils included in the ASTER spectral library (<http://asterweb.jpl.nasa.gov>) and filtered according to the band TM6 (Landsat 5) filter function (SOBRINO; JIMENEZ-MUNOZ; PAOLINI, 2004) and  $P_v$  is the vegetation proportion obtained from Equation 3.5.

$$P_v = \left[ \frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \right] \quad \text{Eq. 3.5}$$

Where  $NDVI$  (normalized difference vegetation index), ranging between  $-1$  and  $+1$ , is calculated from Equation 3.6;  $NDVI_{min}$  is 0.2 and  $NDVI_{max}$  is 0.5 for a mixture of bare soil and vegetation:

$$NDVI = \left[ \frac{OLI\ 5 - OLI\ 4}{OLI\ 5 + OLI\ 4} \right] \quad \text{Eq. 3.6}$$

Where  $OLI\ 4$  and  $OLI\ 5$  correspond to the red and infrared channels of the OLI sensor (Landsat 8), respectively. As indicated by Sobrino, Jimenez-Munoz and Paolini (2004), a more accurate measurement of  $NDVI$  can be obtained by the use of bands atmospherically calibrated.

The final step involves the conversion of the radiance image to spectral emissivity obtained from Equation 3.3, according to the nature of the surface. For this purpose, the TIRS 10 and 11 radiance bands were converted to TOA temperature by the Equation 3.7 (NASA, 2009):

$$T = \frac{K2}{\ln\left(\frac{K1}{L_\lambda} + 1\right)} \quad \text{Eq. 3.7}$$

Where  $K1$  is the calibration constant 1 ( $774.89\ \text{W m}^{-2}\ \text{ster}^{-1}\ \mu\text{m}$ , for band 10 and  $480.89\ \text{W m}^{-2}\ \text{ster}^{-1}\ \mu\text{m}$  for band 11);  $K2$  is the calibration constant 2 ( $1321.08$  for band 10 and  $1201.14$  for band 11); and  $L_\lambda$  is the at-sensor radiance calculated from Equation 3.3. After this conversion, the surface temperature images were generated from Equation 3.8 (WENG; LU; SCHUBRING, 2004):

$$St = \frac{T_B}{1 + (\lambda [T_B/\rho]) \ln_\epsilon} \quad \text{Eq. 3.8}$$

Where  $T_B$  is the blackbody temperature from Equation 3.7;  $\lambda$  is the wavelength of emitted radiance (band 10:  $10.89\ \mu\text{m}$ ; band 11:  $12.15\ \mu\text{m}$ );  $\rho$  is the multiplication of the Boltzmann constant by Planck's constant divided by the velocity of light  $-0.01438$ ; and  $\ln_\epsilon$  is the land surface emissivity ( $\epsilon_v$ ) calculated from Equation 3.4. The resulting images from Equation 3.6 are given in kelvin.

### 3.2.3.3. *SRTM-filled DEM*

The main purpose of this procedure was to create a SRTM DEM free of sinks, which simplifies interpretation of the drainage pattern and the identification of the different topographic gradients in the study area. It is therefore necessary to identify the sinks and to compare with the DEM to determine whether they have to be eliminated. Given this, a depression-less SRTM DEM was created and the result was used to fill the original DEM.

The study area is a flat zone covered by rainforest with strongly dissected residual hills and poorly drained depressions at the central parts of the Podzol plateaux. Because the SRTM DEM was acquired by SAR sensors on the X band, which has a low capacity for penetration in the upper tree canopy, it is sensitive to the texture of the vegetation canopy. Thus, the main purpose of the applied method was to remove the small sinks associated with textural variation in the tree canopies.

The generation of drainage flow direction allowed the estimation of the sinks used to determine the fill limit. After sink identification, their central areas were used to estimate the pour points, allowing the generation of each sink watershed. Untimely, the minimum elevation in each sink watershed was calculated to estimate the average fill limit. The resulting image is a DEM of 30 m spatial resolution (1 arc second data set: SRTM GL1 provided by USGS, 2014).

### 3.2.3.4. *Image classification and generation of soil maps*

This study attempted to classify the objects into seven regions of interest (ROIs) according to the soil cover associated with the vegetation, surface temperature, surface moisture, and topography: (1) Ferralsol–Podzol association covered by high, dense rainforest (HRF); (2) poorly drained Podzols associated with campinarana forest; (3) depositional zones (alluvial Gleysols covered by dense forest and hygrophilous vegetation); (4) incised plains and depressions with seasonally flooded Podzols; (5) overflowed Podzols covered by herbaceous vegetation; (6) regions of water bodies and permanent flooded areas; and (7) sandbanks and bare soils.

The ROIs were identified in the study area from field observations, regional maps (IBGE, 2008), visual interpretation of the spectral signature of the targets (Landsat 8 multispectral composition), and visual interpretation of IKONOS II mosaics in specific zones (where cloud-free scenes were available). The classification of the multi-sensor composition was carried out using the Support Vector Machine (SVM) technique, according to the seven ROIs described above.

A SVM algorithm separated the different ROIs by a hyperplane (VAPNIK, 1998). The points lying on the boundaries were the support vectors and the middle of the margin was the optimal separating hyperplane. An optimum hyperplane was determined using a training data set (ROI), and its generalization ability was verified using a validation data set (field-truth). The study used a polynomial kernel and employed a ‘one-against-one’ technique to allow the multi-class classification. The SVM algorithm was implemented in ENVI® software (ITT, 2009).

Classified images using the SVM classifier were generated for both multi-sensor composition (NDMI, SAVI, land surface temperature, and filled SRTM) and multispectral OLI composition for statistical comparison. Besides, the SVM classifier was applied for both compositions according to the ISODATA unsupervised classification procedure in order to verify the quality of the multi-sensor composition in comparison to multispectral data of OLI Landsat bands. This procedure does not require human intervention that potentially biases classification and determines the differentiability among spectral classes, giving a better comparison parameter.

Considering that ISODATA is a fully automated method, there is no possibility of interference with the designation of the classes and thus the validation of the classified images was done by comparison between the automatically generated classes and the field-truth established for the study areas.

#### **3.2.4. Soil map and correlation with field sample data**

For each individual profile, the organic carbon stock was estimated by the following equation:

$$SOC_s = \sum_{i=1}^n B_i C_i D_i \quad \text{Eq. 3.9}$$

Where  $SOC_s$  is the SOC stock ( $\text{kg C m}^{-2}$ );  $B_i$  is the soil bulk density ( $\text{mg m}^{-3}$ ) of layer  $i$ ;  $C_i$  is the proportion of organic carbon ( $\text{g C g}^{-1}$ ) in layer  $i$ ; and  $D_i$  is the thickness (m) of layer  $i$ . The average soil density for Podzol horizons ( $D_{bi}$ ) was calculated from surveys previously carried out in the Amazon region by Du Gardin, Grimaldi and Lucas (2002) and Montes et al. (2011), and directly determined by the Kopeck ring method, with a  $70.49 \text{ cm}^3$  cylinder (3.8 cm height and 4.86 cm diameter). The SOC for each soil map unit was then estimated according to its corresponding area in the study site, and then extrapolated at the regional scale.

Soil mapping and segmenting at the regional scale was achieved using products derived from the Landsat sensors and related to the soil cover: biophysical properties of the vegetation and soils (SAVI/NDMI indices), thermal behavior of the soil surface (temperature images), and surface texture derived from the filled SRTM DEM for segmentation of topographic gradients. The resulting images were applied to estimate the regional behavior of the biophysical characteristics related to soil type.

### **3.3. Results and Discussion**

#### ***3.3.1. Vegetation and topographic features related to lateral variation in podzols***

According to field observations, changes in soil cover occur abruptly in accordance with soil type. Given this, we identified three major soil domains related to Podzols.

The first group is dominant in the landscape, and consists of seasonally flooded and overflowed Podzols. The vegetation is strongly correlated with the topography, with herbaceous campina in depressed flooded areas and scrubland in adjacent plateaux and, in some regions, patches of bare white sand where grasses and lichens grow according to variations in soil surface moisture. According to both IBGE (2008) and field observations, bare white sand patches can be considered as Gleysols or Podzols with low SOM content in topsoil horizons due to occasional dryness of topsoil material.

The second soil group belongs to the Rio Branco geomorphological domain, with poorly drained Podzols. According to IBGE (2008), this area is dominated by tabular hills with vast flat interzones. Field surveys, however, have shown that this domain is a flat landform whose elevation ranges from 50 to 60 m above sea level and covered by campinaranam with scattered HRF patches related to better-drained Podzols. Therefore, the phytophysionomies comprise an ecotone. Ferralsols were not observed in this group. The third soil group comprises well-drained Podzols and Ferralsols covered by HRF. This area is characterized by a slight inclination towards the drainage network. Ferralsols may occur punctually in scattered hills 2–10 m above flat, sandy inter-hill surfaces.

The segmentation of the three major groups mentioned above was carried out by a clustering group analysis, considering a Delaunay triangulated relationship between 1000 random samples for the following images: SAVI, relief (SRTM), NDMI, and land surface temperature (arithmetic mean of approximately 186 pixels for each sample). Temperature and SAVI demonstrated an inverse polynomial relationship (Figure 3.3a); the

correlation between relief and SAVI is better illustrated by a third-order polynomial regression (Figure 3.3b) and by a linear relationship for the variables NDMI and SAVI (Figure 3.3c).

The inverse correlation between LST and vegetation density (SAVI) indicates that the areas with the highest vegetation indices (HRF from Group 3 and campinarana from Group 2) also have the lower canopy temperatures, which is consistent with their location on hills and tabular tops. Group 2 has a spectral behavior similar to Group 3, although it may be found in low lands as a sclerophyllous campinarana with slightly higher canopy temperatures. Group 1 has a wide range of temperature, moisture indices, and vegetation indices, which is consistent with the variety of soil cover and the high surface temperatures that can occur in herbaceous or bare sand areas under drier conditions.

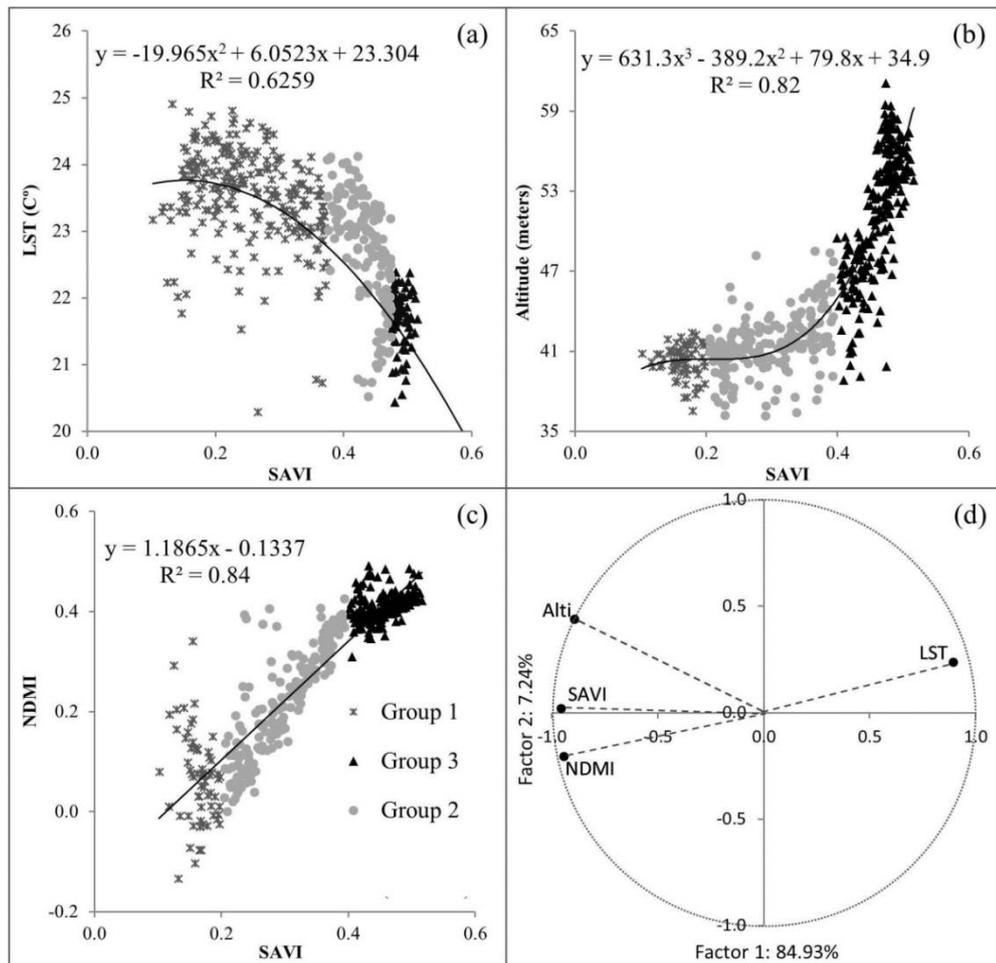


Figure 3.3 - Scatterplot showing the relation between soil adjusted vegetation index (SAVI) and the following variables: (a) land surface temperature (LST); (b) altitude; and (c) normalized difference moisture index (NDMI). (d) Projection of the normalized factor coordinates of variables (biophysical variables) in the 1 × 2 factor plane obtained by principal component analysis. Group 1: seasonally flooded and overflowed Podzols. Group 2: poorly drained Podzols. Group 3: well-drained Podzols and Ferralsols.

Taking into account the principal component analysis of the remote sensing variables, the first principal component (factor 1) accounted for a variance of 84.9% and the second of 7.2%, which corroborates the assumption of a gradual lateral variation in biophysical parameters, from regions of HRF to zones with bare soils and grasses. We observed that this pattern is properly modeled by a general approach involving the study area at the regional scale. However, smaller variations associated with each group must be explained by local models within each cluster group.

According to the distribution of the clusters, Group 1 corresponds to extensive Podzols areas located in depressed regions. Both field surveys and remote-sensing images have shown that the surface moisture in such areas can vary laterally according to the topography and vegetation cover (Figure 3.3). The differences in relief were not detectable in the SRTM images considering that the larger height differences are around 1.5–2 m. Such height differences, however, are not negligible in the context of Podzol lateral differentiation: the wettest areas allow humic topsoil horizons that are absent in drier areas.

It is important to highlight that lateral variations in the soil mantle in Group 1 are not abrupt like in other areas (Groups 2 and 3). As observed in the field, all intermediaries exist between soils with low SOM content in the topsoil (bare white sand) and those with a high-SOM topsoil horizon (O horizon). This has to be considered for segmentation and to estimate the soil carbon stock in different environments.

### ***3.3.2. Classification of soil cover and generation of regional soil map***

The land cover/soil thematic maps were produced from the multi-sensor (OLI/TIRS/SRTM) and OLI multispectral composition according to the SVM classification. The same method, based on ROIs selected for the same known targets (field-truth), was applied to the Landsat OLI multispectral composition (Green, Red, Near Infrared, SWIR 1, and SWIR 2 bands). The ISODATA classifier was applied to both multi-sensor and OLI multispectral compositions (the ISODATA clustering method was used simply for comparison between compositions). The accuracy assessment was then based on the computation of the overall accuracy (OA), producer's accuracy (PA), user's accuracy (UA), and kappa coefficient (Kc) (CONGALTON; GREEN, 1999), in order to evaluate and compare compositions adopting field surveys as ground truth.

The overall indices of the ISODATA classification were 53.0% and 63.9% for the OLI multispectral composition and the multi-sensor composition, respectively. Given this, and taking into account an optimized unsupervised classification with identical parameters for both compositions, we observed that the classifications were similar. The multi-sensor composition, however, gave results closer to the reality (ROI) with a better Kc when compared with the OLI composition (0.52 versus 0.45), and better PA and UA for most classes (Table 3.1).

Both compositions were able to segment the three major clustering groups and the variations within each group, but returned high levels of omitted targets for those classes that were correlated to similar targets (Table 3.1). The unsupervised classification of the OLI multispectral composition has areas of seasonally flooded podzols (Class 4) associated with other targets having closer spectral behaviour, such as regions of overflooded podzol (Class 5) and depositional zones (Class 3). With regard to the multi-sensor composition, we observed a higher level of confusion related to Classes 2 and 3 due to their close spectral similarity. ISODATA clustering was helpful in an unbiased comparison between the two compositions. However, better results were achieved through the application of supervised classification algorithms (Figure 3.4).

Table 3.1 - Producer and user's accuracy (PA and UA, respectively) for ISODATA clustering according to the field-truth (ROI). The classes of water bodies and bare soils are not shown.

Classes <sup>a</sup>	Multisensor Composition		OLI Composition	
	PA (%)	UA (%)	PA (%)	UA (%)
Class 1	99.90	47.41	94.11	94.47
Class 2	36.00	90.78	86.50	65.74
Class 3	28.12	16.70	19.54	15.12
Class 4	95.48	86.09	0.06	0.10
Class 5	90.09	95.23	73.03	95.14

<sup>a</sup>Class 1: Ferralsol/Podzol association covered by HRF (High Dense Rainforest); Class 2: Poorly-drained Podzols associated to Campinarana Forest; Class 3: depositional zones (alluvial Gleysols covered by dense forest and hygrophilous vegetation); Class 4: incised plains and depressions with seasonally flooded Podzols; Class 5: overflooded Podzols covered by herbaceous vegetation.

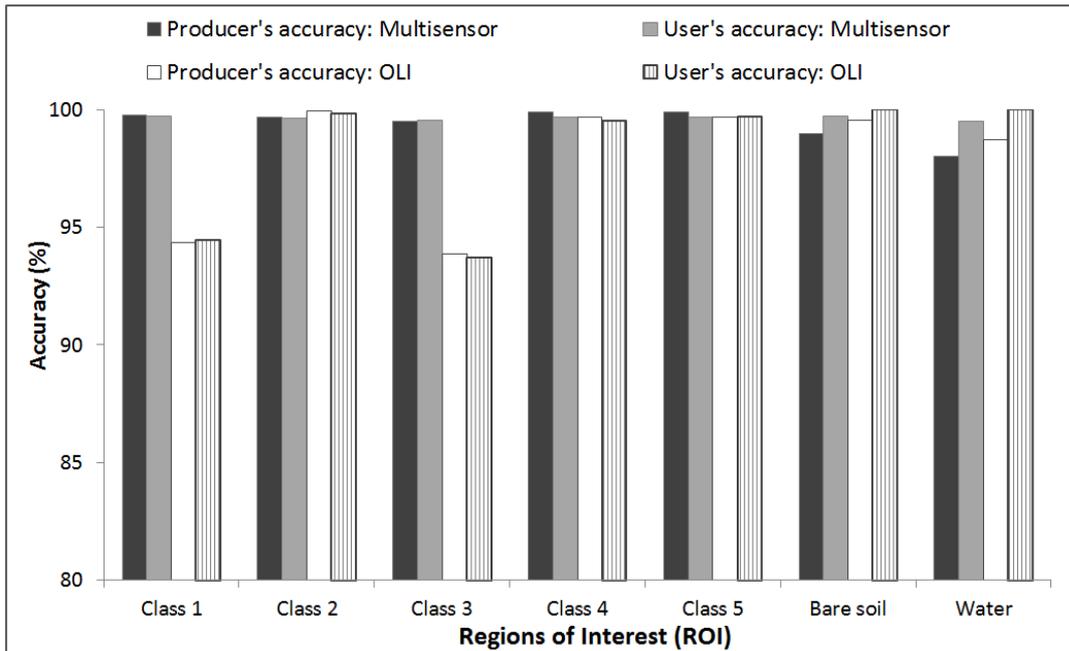


Figure 3.4 - Producer's and user's accuracies for the SVM classification of multi-sensor and OLI Landsat 8 compositions. The designation of each class is shown in Table 3.1.

All of the classes in Figure 3.4 have a PA above 98% for the multi-sensor composition. Such values were achieved according to ground truth observed in field surveys and regional soil maps (IBGE, 2008) and thus the classifications might be biased, which must be considered in final maps of soil carbon stock. It is important to highlight the enhanced efficiency of the multi-sensor composition when compared with the OLI multi-spectral composition using the same ROI for both images (Figure 3.4).

To access an optimum SVM classification algorithm, a range of kernel types, penalty parameters, pyramid levels, and classification probability thresholds was tested until reaching the value closest to the ground truth (SHAFRI; RAMLE, 2009). The best result was obtained through a radial basis kernel with a penalty parameter of 100, a pyramid level of 2, and a probability threshold of 0. The OA of such classification was 99.65% with a Kc of 0.99. For the SVM classification of the OLI multispectral composition, obtained OA and Kc were 96.87% and 0.96, respectively.

Table 3.2 - Confusion matrix of the multisensor classified image, representing the classes' similarity.

		Ground Truth ROI (%)						
Classes <sup>a</sup>		Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Classified (SVM)	Class 1	99.8	0.0	0.3	0.0	0.0	0.0	0.0
	Class 2		99.7	0.3	0.0	0.0	0.0	0.1
	Class 3			99.5	0.0	0.0	0.0	0.0
	Class 4				99.0	0.0	1.0	0.0
	Class 5					98.0	0.0	1.0
	Class 6						99.0	0.0
	Class 7							99.0

<sup>a</sup>Class 1: Ferralsol/Podzol association covered by HRF; Class 2: Poorly-drained Podzols associated to Campinarana Forest; Class 3: depositional zones (alluvial Gleysols covered by dense forest and hygrophilous vegetation); Class 4: incised plains and depressions with seasonally flooded Podzols; Class 5: overflooded Podzols covered by herbaceous vegetation; Class 6: bare soils; Class 7: water.

The SVM algorithm returned refined results when compared with the ground truth (Table 3.2), with all classes having a similarity above 98%. However, the ground truth data may be biased due to difficulty in processing detailed field surveys and the low spatial scale of the reference soil map (1:250,000). A detailed visual inspection over the classified image was thus carried out to verify errors and occurrences of incorrect classification. Errors were corrected by changing the pixel values of the original GRID. The regional soil map was then elaborated on the basis of the classified image, with a nominal spatial resolution of 30-m.

### 3.3.3. Mapping the deep-SOC stock in Podzol regions

Our studies have shown that topsoil and deep SOM-rich Podzol horizons can vary laterally in both thickness and carbon content, which means that the carbon stored in Podzols is sensitive to local environmental variables such as soil moisture, topography, hydrological regime, geologic substratum, and vegetation cover, as well as regional variables such as temperature and rainfall. The most refined maps available for this region (IBGE, 2008) do not consider local variables as relevant factors in regard to the lateral distribution of Podzols. The local environmental variables are considered hereafter to allow correlation between soil carbon stocks and biophysical features.

The lateral and vertical variation in carbon content in Podzol profiles is related to the relief and the local hydrologic regime. The first area we investigated was designated as a HRF over Podzols and Ferralsols. Detailed study of a range of profiles gave a good approximation of the average carbon stock in this domain (Figures 3.5 and 3.6). The better-drained areas (HRF over low-hill Ferralsols and well-drained Podzols) have a large amount of carbon in the first few centimeters of soil (O and A horizons). Deep Bh (depth 1.5–4.5 m) was observed but with a low content of organic carbon (Figure 3.5). At the transition between this area and the

poorly drained podzol region, the deep SOM-rich layers increase and show a higher content of carbon. The third main area comprises extensive regions of overflooded Podzols where the occurrence of Bh was observed.

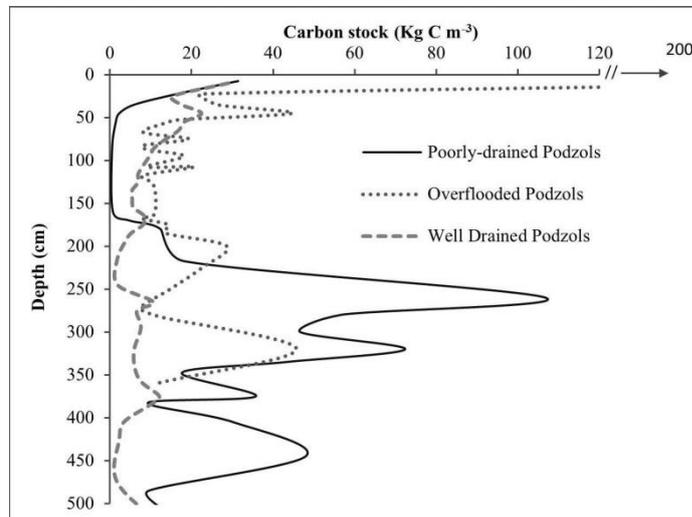


Figure 3.5 - Average carbon stock for the three main Podzol groups described in the study area.

The final group consists of herbaceous and flooded areas. The presence of Podzols in such areas was disregarded in previous studies (BATJES; DIJKSHOORN, 1999; BERNOUX et al., 2002; CERRI et al., 2007; SAATCHI et al., 2007). The topsoil horizons comprise high SOM content. Podzols with bare topsoil store some carbon in their deep Bh, although in smaller quantities than more vegetated Podzols, probably because of low production in the topsoil horizons of humic substances likely to accumulate in deep Bh. The high level of moisture in such areas allows the accumulation of a considerable amount of SOC in the topsoil. In this domain we observed that topsoil (first 5 cm) may store more than 200 t C ha<sup>-1</sup> (Figure 3.5).

From vegetated Podzol areas to bare soil, the amount of carbon stored decreases gradually, until the zones of open sandy fields. According to Stropp et al. (2011), white sand fields have a sparse distribution over the region of the Rio Negro basin. In the study area, however, these zones follow an insular pattern with highly weathered bare soils in the depressed centre surrounded by SOC-rich Podzols on flat areas, then Ferralsols at the slightly incised, low-hill borders. Similar patterns were previously observed in other Podzol areas (BOULET et al., 1997; NASCIMENTO et al., 2004; MONTES et al., 2007; 2011). These authors pointed out that the lateral organization of the Ferralsol/Podzol soil system indicates its stage of evolution, which is important in estimating carbon stock.

The average carbon stock for each soil group was estimated taking into account the clustering analysis and the soil map derived from the multi-sensor image composition (Figure

3.6). The greatest amount of carbon stored in depth horizons was found in the poorly drained Podzol areas, the lowest in Podzols/Ferralsols, and intermediate amounts in flooded Podzols (Table 3.3).

Table 3.3 - Average Carbon Stock for Podzols.

Soil Groups	0 to 0.50 m	0.50 to 4 m
	Carbon (ton ha <sup>-1</sup> )	Carbon (ton ha <sup>-1</sup> )
Poorly-drained Podzols <sup>a</sup>	62.80 ±10.03	477.75±63.9
Flooded Podzols <sup>b</sup>	249.12±18.83	161.22±45.35
Podzols/Ferralsols association	84.35±12.4	188.77±38.5

<sup>a</sup>It was considered the average thickness observed in filed for surface and depth horizons, according to the soil types. <sup>b</sup>This class comprises the soil units seasonally flooded Podzols and overflooded Podzols.

The areas of alluvial Gleysols and Ferralsols have a small amount of carbon stored in horizons below 0.50 m soil depth, mostly located in the first few centimetres in the organic layer according to our investigations and values provided by IBGE (2008) for soil samples collected in areas of Ferralsols and alluvial Gleysols (0–0.5 m: 70.9 ± 27 t C ha<sup>-1</sup> for Ferralsols and 89.9 ± 35.4 t C ha<sup>-1</sup> for Gleysols; below 0.5 m: 31.5 ± 10.7 t C ha<sup>-1</sup> for Ferralsols and 25.8 ± 8.5 t C ha<sup>-1</sup> for Gleysols). In flooded Podzols we observed a large amount of carbon stored as poorly decomposed organic matter on the surface (Table 3.3), due the low microbial activity in this domain. These zones are strongly related to wetlands covered by grasses and scrubs. Finally, we did not identify a significant amount of carbon stored in areas of bare soil. The range of the soil units and their carbon stock are shown in Table 3.4.

Table 3.4 - Total carbon stock according to each soil unit. The stock is represented in Teragram (10<sup>12</sup> grams) and the area in hectares.

Class <sup>a</sup>	C Stock (Tg)	Area (ha)	Area (%) <sup>b</sup>	C Stock (%)
Class 1	17.9±6.4	6.47 10 <sup>4</sup>	26.8	21.7
Class 2	16.1±2.5	3.07 10 <sup>4</sup>	12.8	19.6
Class 3	4.0±0.4	2.81 10 <sup>4</sup>	11.7	5.0
Class 4	21.8±5.7	7.88 10 <sup>4</sup>	32.7	26.5
Class 5	9.7±0.7	3.19 10 <sup>4</sup>	13.2	11.8
Class 6	-	3.70 10 <sup>3</sup>	1.6	-
Class 7	-	3.17 10 <sup>3</sup>	1.3	-

<sup>a</sup>The designation of each class is shown in Table 3.2. <sup>b</sup>The percentage values indicate the scope of each unit, relative to the total area and carbon stock.

Previous carbon stock maps for these regions (BERNOUX et al., 2002) estimated that Podzol areas store about  $275 \text{ t C ha}^{-1}$  (0–1 m soil thickness), which is lower than our estimation of  $415 \text{ t C ha}^{-1}$  as average carbon stock in the Podzol areas (0–4 m soil thickness) when we take into account the deep SOM-rich Podzol horizons.

The most detailed map available for this area (IBGE, 2008) shows generalized classes of Podzols, Ferralsols, and alluvial Gleysols (IBGE limits are also reported in Figure 3.6). The subunits within each soil unit are not designated due to the low level of detail in the soil map. The definition of soil units, according to surveys carried out in previous studies (IBGE, 2008), was based on a small number of soil samples. Extrapolation was performed by interpretation of topography (SAR) and the spectral behavior of vegetation (passive remote-sensing systems). The lack of soil samples and the absence of detailed biophysical data are the main reasons for the misclassification of some soil units. As an example, some Ferralsol/Podzol associations represented on Figure 3.6 were defined by IBGE (2008) as Ferralsols. According to our observations, Ferralsols in these areas occur as scattered, well-drained hills that can only be mapped at a local scale.

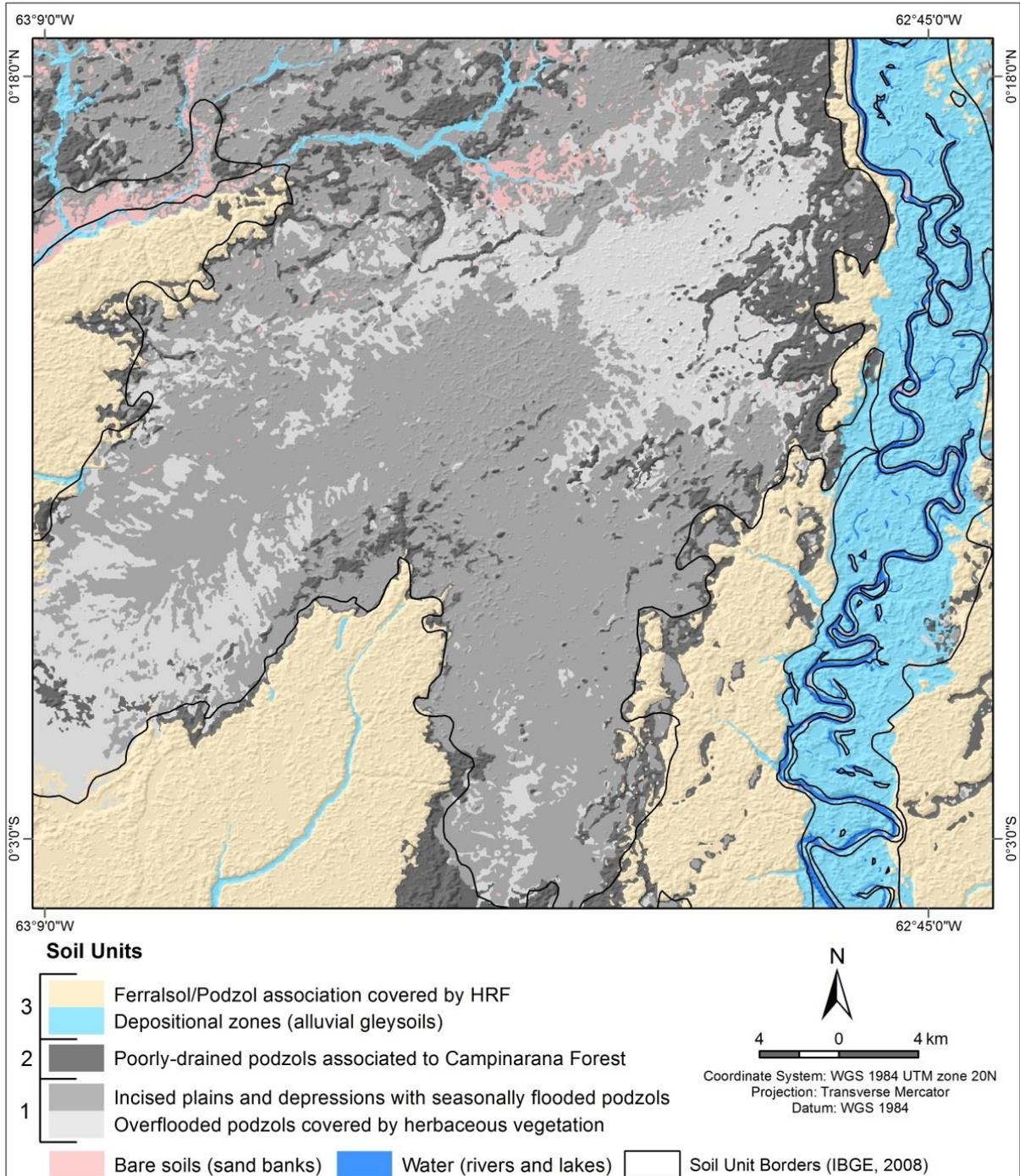


Figure 3.6 - Soil map illustrating the spatial distribution of the soil types. The numbers for the soil units represent the cluster groups described in Figure 3.3.

Taking into account the carbon soil contents defined in this study and using the IBGE (2008) soil delimitation, the calculation of soil carbon stored in the study area yields 0.65 Pg when using the IBGE (2008) soil delimitations and 0.81 Pg when using the soil map derived from the multi-sensor image composition. Such a discrepancy highlights the need for mapping the high Rio Negro basin at a regional scale, allowing for a more accurate estimation of the carbon stored in soils. The correct estimation of SOC stock at a regional scale in such a

complex environment depends on both the availability of soil samples suitable to represent both the lateral and vertical variability of the soil units and the association with biophysical parameters inferred by remote sensing data.

### 3.4. Conclusions

A multi-sensor approach proved crucial to mapping the soil carbon stock at the regional scale in regions of hydromorphic soils in the high Rio Negro basin characterized by wide lateral and vertical variability in carbon stock. The quantity of carbon stored in soils is related to environmental aspects such as topography, vegetation type, and soil surface moisture, which can be indirectly inferred by remote sensing images through a range of data sets collected by various sensor systems and orbital satellite platforms. The combination of these data sets allowed a better understanding of the aspects related to soil variability, when combined with field sample data.

According to our study, Podzols from depressed overflooded and poorly drained areas of sclerophyllic vegetation store considerable amounts of carbon in deep horizons that can range from 2 m to more than 10 m. Pre-existing maps have only a low level of detail, and field-truth was based on scarce soil observations. Moreover, the soil profiles used in previous field surveys are limited to depths of no more than 1 m, which is inadequate to investigate the level of SOC stored in Podzol's Bh horizons.

### References

BATJES, N.H.; DIJKSHOORN, J.A. Carbon and nitrogen stocks in the soils of the Amazon region. *Geoderma*, Amsterdam, v. 89, p. 273–286, 1999.

BERNOUX, M.; CARVALHO, M.D.S.; VOLKOFF, B.; CERRI, C.C. Brazil's Soil Carbon Stocks. *Soil Science Society of America Journal*, Madison, v. 66, p. 888–896, 2002.

BOULET, R.; CHAUVEL, A.; HUMBEL F.X.; LUCAS, Y. Analyse structurale et cartographie en pédologie: I – Prise en compte de l'organisation bidimensionnelle de la couverture pédologique: les études de toposéquences et leurs principaux apports à la connaissance des sols. *Cahiers ORSTOM, Série Pédologie*, Paris, v. 19, p. 309–321, 1982.

BOULET, R.; LUCAS, Y.; FRITSCH, E.; PAQUET, H. Geochemical Processes in Tropical Landscapes: Role of the Soil Covers. In: PAQUET, H.; CLAUER, N. (Ed.). *Soils and sediments: mineralogy and geochemistry*. Heidelberg: Springer-Verlag, 1997. p. 67–96.

BRASIL. Ministério das Minas e Energia. **Projeto RADAMBRASIL**: Volume SA.19: Pedologia. Rio de Janeiro, 1977. p. 181–237. (Levantamento dos Recursos Naturais, 14).

BUENO, G.T. **Appauvrissement et podzolisation des latérites du bassin du Rio Negro et genèse des Podzols dans le haut bassin amazonien**. 2009. 193 p. Thesis (Ph.D.) - Institut de Physique du Globe de Paris (IPGP), Paris, 2009.

CERRI, C.E.P.; EASTER, M.; PAUSTIAN, K.; KILLIAN, K.; COLEMAN, K.; BERNOUX, M.; FALLOON, P.; POWLSON, D.S.; BATJES, N.H.; MILNE, E.; CERRI, C.C. Predicted Soil Organic Carbon Stocks and Changes in the Brazilian Amazon between 2000 and 2030. **Agriculture Ecosystems & Environment**, Amsterdam, v. 122, n. 1, p. 58-72, 2007.

CONGALTON, R.G.; GREEN, K. **Assessing the Accuracy of Remotely Sensed Data: principles and practices**. Boca Raton: Lewis, 1999. 183 p.

DU GARDIN, B.; GRIMALDI, M.; LUCAS, Y. Effets de la déshydratation sur les sols du système ferral-sol-podzols d'Amazonie centrale. Reconstitution de la courbe de désorption d'eau à partir de la porosimétrie au mercure. **Bulletin de la Société Géologique de France**, Paris, v. 173, p. 19-34, 2002.

DUBROEUCQ, D.; VOLKOFF, B. From Oxisols to Spodosols and Histosols: Evolution of the Soil Mantles in the Rio Negro Basin (Amazonia). **Catena**, Amsterdam, v. 32, p. 245-280, 1998.

DUBROEUCQ, D.; VOLKOFF, B.; FAURE, P. Les couvertures pédologiques à podzols du bassin du haut Rio Negro (Amazonie). **Etude et Gestion des Sols**, Orléans, v. 6, p. 131-153, 1999.

GLOBAL LAND COVER FACILITY - GLCF. **Landsat GeoCover**. Maryland, 2009. Available at: <<http://glcf.umiacs.umd.edu/data/landsat/>>. Accessed in: Sep. 10, 2014.

GOODWIN, N.R.; COOPS, N.C.; WULDER, M.A.; GILLANDERS, S.; SCHROEDER, T.A.; NELSON, T. Estimation of Insect Infestation Dynamics Using a Temporal Sequence of Landsat Data. **Remote Sensing of Environment**, New York, v. 112, p. 3680-3689, 2008.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. Geoscience Division - DGC. Coordenação de Recursos Naturais e Estudos Ambientais - CREN. **Georefered Maps of Natural Resources**: Scale 1:250:000. Digital Format: shp. Rio de Janeiro, 2008. Available at: [ftp://geoftp.ibge.gov.br/mapas/banco\\_dados\\_georeferenciado\\_recursos\\_naturais](ftp://geoftp.ibge.gov.br/mapas/banco_dados_georeferenciado_recursos_naturais). Accessed in: Aug. 15, 2014.

ITT VISUAL INFORMATION SOLUTIONS. **ENVI®**. Atmospheric Correction Module. Version 4.7. Boulder, CO, Aug. 2009. Available at: [http://www.envi.com.br/index.php?link=Atmospheric\\_Correction](http://www.envi.com.br/index.php?link=Atmospheric_Correction). Accessed in: Jan. 20, 2014..

LUCAS, Y.; CHAUVEL, BOULET, A., R.; RANZANI, G.; SCATOLINI, F. Transição latossolos-podzols sobre a formação Barreiras na região de Manaus, Amazônia. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 8, p. 325-335, 1984.

MONTES, C.R.; LUCAS, Y.; MELFI, A.J.; ISHIDA, D.A. Systèmes sols ferrallitiques-podzols et genèse des kaolins. **Comptes Rendus Geoscience**, Paris, v. 339, p. 50-56, 2007.

MONTES, C.R.; LUCAS, Y.; PEREIRA, O.J.R.; ACHARD, R.; GRIMALDI, M.; MELFI, A.J. Deep Plant-Derived Carbon Storage in Amazonian Podzols. **Biogeosciences**, Göttingen, Germany, v. 8, p. 113–120, 2011.

NASA. **Landsat 7 Science Data Users Handbook**. Maryland, 2009. Available at: [http://landsathandbook.gsfc.nasa.gov/pdfs/Landsat7\\_Handbook.pdf](http://landsathandbook.gsfc.nasa.gov/pdfs/Landsat7_Handbook.pdf). Accessed in: Jan. 15, 2014.

NASCIMENTO, N.R.; BUENO, G.T.; FRITSCH, E.; HERBILLON, A.J.; ALLARD, T.; MELFI, A.; ASTOLFO, R.; BOUCHER, H.; LI, Y. Podzolization as a Deferralitization Process: A Study of an Acrisol-Podzol Sequence Derived from Palaeozoic Sandstones in the Northern Upper Amazon Basin. **European Journal of Soil Science**, Oxford, v. 55, p. 523–538, 2004.

SAATCHI, S.S.; HOUGHTON, R.A.; DOS SANTOS ALVALÁ, R.C., SOARES, J.V.; YU, Y. Distribution of Aboveground Live Biomass in the Amazon Basin. **Global Change Biology**, Oxford, v. 13, p. 816–837, 2007.

SHAFRI, H.Z.M.; RAMLE, F.S.H. A Comparison of Support Vector Machine and Decision Tree Classifications Using Satellite Data of Langkawi Island. **Information Technology Journal**, New York, v. 8, p. 64–70, 2009.

SOBRINO, J.A.; JIMENEZ-MUNOZ, J.C.; PAOLINI, L. Land Surface Temperature Retrieval from LANDSAT TM 5. **Remote Sensing of Environment**, New York, v. 90, p. 434–440, 2004.

STROPP, J.; VAN DER SLEEN, P.; ASSUNÇÃO, P.A.; SILVA, A.L.; TER-STEEGE, H.T. Tree Communities of White-Sand and Terra-Firme Forests of the Upper Rio Negro. **Acta Amazônica**, Manaus, v. 41, p. 521–544, 2011.

USGS. Earth Resources Observation and Science. **EROS Center**. Reston, VA, 2010. Available at: <http://eros.usgs.gov/>. Accessed in: Mar. 10, 2014.

VAPNIK, V. **Statistical Learning Theory**. New York: John Wiley and Sons, 1998.

WENG, Q., LU, D.; SCHUBRING, J. Estimation of Land Surface Temperature–Vegetation Abundance Relationship for Urban Heat Island Studies. **Remote Sensing of Environment**, New York, v. 89, p. 467–483, 2004.

#### 4. Evaluation of pedotransfer equations to predict deep soil carbon stock in tropical Podzols compared to other soils of Brazilian Amazon forest<sup>3</sup>

##### Abstract

According to the soil measurement procedures proposed by the Intergovernmental Panel on Climate Change (IPCC), the sampling depth for SOC stock estimation is centered on the upper soil horizons where root biomass and organic matter inputs are concentrated, depending on soil type and ecosystem, typically between 0 and 0.3 m. However, recent research in areas of Amazon Podzols has shown that these soils store a great amount of carbon in thick spodic horizons (Bh). The amount of carbon stored in deep Bh horizons of Podzols (down to 3 m) may exceed 80 kg C m<sup>-2</sup> in some regions of the Amazon. Thus, a better understanding of the vertical distribution of the SOC in Amazon soils is an urgent matter considering the volume of carbon stored in Podzols. Given this, the main goal of this research was to test and to propose pedotransfer functions based on collection of Amazon soil profiles in order to estimate SOC stock and evaluate different soil attributes that could be used to infer indirectly, soil bulk density. For this propose, we selected around 320 pedons that were collected in the region of the Rio Negro basin, to model the vertical distribution of SOC stock using a series of negative exponential profile depth functions and parametric/non-parametric functions for Podzols. The derived function parameters were used to predict carbon stock in deep horizons for all studied profiles and to explain the vertical behavior of the SOC stock in Podzol profiles. The soil bulk density of Amazon soils was properly modeled by symbolic regression, considering pH, clay content and SOC as the most relevant variables likely to affect soil bulk density values. We observed that the SOC stored in deep horizons of non-Podzol soils can be modeled by exponential decay equations. However, in Podzol, the vertical distribution of carbon stock is highly complex with a significant increase in deep horizons, which cannot be explained by negative exponential functions. According to our research, the SOC stock of Amazon soils excluding Podzols, can be predicted by fitted exponential functions (RMSE: 0.9 kg C m<sup>-2</sup>). However, the vertical variation of SOC stored in Podzol profiles can be modeled just by complex equations (equal-area spline RMSE: 13.6 kg C m<sup>-2</sup>; Fourier RMSE 15.9 kg C m<sup>-2</sup> and Sum of Sines RMSE: 15.0 kg C m<sup>-2</sup>) with a larger number of parameters. According to the results achieved in this research we concluded that the SOC stock of Podzols can be indirectly estimated for the whole soil profile by integrating the Sum of Sines and Fourier equations, which is not possible when applying an equal-area spline fitting due to the absence of model parameters. Moreover, spodic horizons store most of the carbon pool of Podzols areas and have more than twice of the capability of storing carbon when compared to other Amazon soils.

**Keywords:** Soil Organic Carbon Stock, Podzols, Amazon Forest, Pedotransfer Equations

---

<sup>3</sup> This paper will be published as a chapter of the book "Digital Soil Morphometrics" (SPRINGS. Series: *Progress in Soil Science*) in November, 2015. Authors: O. J. R. Pereira, C. R. Montes, Y. Lucas, A. J. Melfi

#### 4.1. Introduction

The Brazilian tropical Podzols cover  $1.36 \cdot 10^5$  km<sup>2</sup> of the Amazon forest, which represents 2.7% of the total area of the Amazon biome and around 20% of the soils of the Rio Negro Basin. The other important soil group in this region comprises ferralitic soils (Acrisols and Ferralsols) that cover 55% of the Rio Negro basin. The remaining soil groups are related to alluvial and litholic soils as well as scattered hydromorphic Plinthosols. Such diversity of soil types reflects on the capacity of the Amazonian biome on storing soil organic carbon (SOC), especially in regions of Podzols. According to recent research, the Amazon Podzols (MONTES et al., 2011) store about  $13.6 \pm 1.1$  PgC, which is at least 12.3 PgC higher than previous estimates (BERNOUX et al., 2002; BATJES; DIJKSHOORN, 1999) that have considered soil depths up to 0.3m.

Several surveys have investigated the capability of soils to store and retain SOC (POST et al., 1982; BURINGH, 1984; KIMBLE, 1990; ESWARAN et al., 1993; BATJES, 1996), but present researches usually consider a fixed soil depth, typically based on the topsoil 0.2 or 0.3 m, where the highest SOC concentrations usually occur (BURKE et al., 1989). Batjes (1996) reported a 60% increase in the global SOC pool when the second meter of soil was included, taking into account the FAO (2012) soil classification system. A few studies of the Amazon forest (MONTES et al., 2011; PEREIRA et al., 2015) have described the capability of Podzols in storing high amounts of C in deep spodic (Bh) horizons. These soils have an average stock of  $70 \text{ kg C m}^{-2}$  and around 80.9% of its C is stored in thick deep Bh horizons in depths ranging from 2 to more than 5m (MONTES et al., 2011; PEREIRA et al., 2015).

The vertical pattern of SOC content in Podzols is highly complex when compared to other Amazon soils, with a significant increase in thick Bh horizons (MONTES et al., 2011). Negative exponential depth function has been successfully applied in several mineral soils to model and predict C stock. However, any local variation affects the quality of the exponential fit everywhere else in the soil profile, as observed by Webster (1978). In this context, the modelling of Podzols SOC stock can be carried out by non-parametric depth function (eg.: equal-area spline) that can result in satisfactory adjustment, but with the disadvantage of not providing any parameters that would allow model generalizations (BISHOP et al., 1999). The summarization of the model by parameters is essential to allow an indirect estimation of the SOC in Podzols and to explain the behaviour of C along the soil profile, by means of a general approach.

Traditionally, the amount of C stored in soil is obtained as C mass per unit area according to a specific profile depth ( $T_p$ ). The calculation is carried out by summing the C stock ( $\text{kg C m}^{-2}$ ) of the measured soil layers ( $1, 2, \dots, N$ ). Thus, the content of C to a given soil profile can be obtained by the following equation:

$$C_s = \sum_{p=1}^N (C * \rho_p) * T_p \quad \text{Eq. 4.1}$$

Where  $C_s$  is the carbon stock ( $\text{kg C m}^{-2}$ ) to a given profile;  $C$  is the carbon content in mass basis ( $\text{kg C kg}^{-1}$ );  $\rho_p$  is the soil bulk density ( $\text{g cm}^{-3}$ ); and  $T_p$  is the layer thickness. Another option to obtain the  $C_s$  value is through the application of a profile depth function fitted to the soil C data in a volumetric basis ( $\text{kg C m}^{-3}$ ) according to specific measured soil layers ( $T_p$ ). The integration of the function is applied in order to obtain the SOC stock ( $\text{kg C m}^{-2}$ ) for the whole profile. The expression of C content as depth function is useful to estimate the C stock down to certain depths and to standardize databases where soil depths are sampled to layers randomly distributed (ARROUAYS; P'ELISSIER, 1994; MINASNY et al., 2006).

Parametric pedotransfer functions (PTF) are widely used in soil science to predict several soil attributes based on empirical equations that result in function parameters that can be easily applied to measure soil attributes (McBRATNEY et al., 2002). Given this, the main goal of this research was to test and to propose PTF functions in several Amazon soil profiles (IBGE, 2008; EMBRAPA, 2014) in order to estimate SOC stock ( $C_s$ ) and evaluate different soil attributes that could be used to infer indirectly, soil bulk density ( $\rho_p$ ).

## 4.2. Methodology

The methods adopted in this research are divided in two general steps. The indirect estimation of soil bulk density by evaluation of traditional PTF functions (BERNOUX et al., 1998; TOMASELLA; HODNETT, 1998; BENITES et al., 2007), compared to the ones developed in the frame of this research. The second step was focused on the estimation of SOC stock by the application of curve fitting models based on different approaches, taking into account the behaviour of SOC along the soil profiles of Podzols and other Amazon soils.

### 4.2.1. Field Sample Data

The studied area is located in the Amazonia State/Brazil (Figure 4.1). The soil database used in this study was provided by IBGE (2008) and Embrapa (2014) as well as collected in filed surveys developed by this research (Figure 4.1).

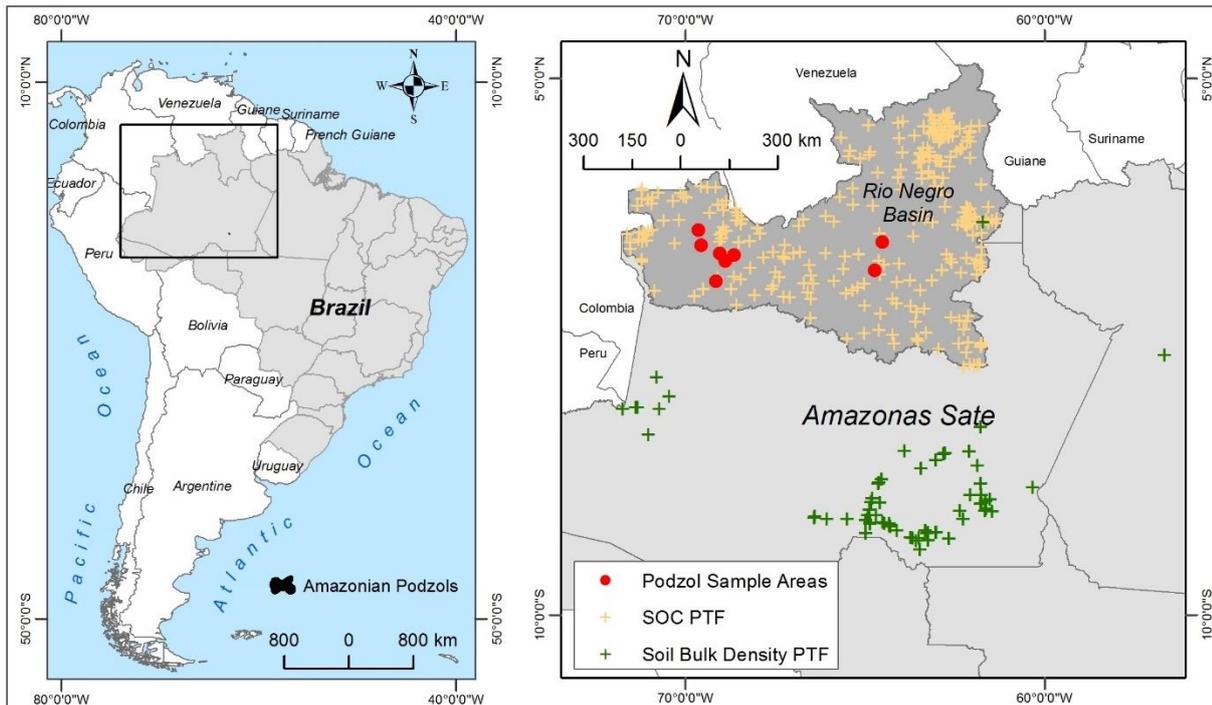


Figure 4.1 - Situation of the studied area highlighting the location of the soil sample profiles used in this study.

The soil profiles in Figure 4.1 are divided in three groups. The “Podzol Sample Areas” that represent the profiles collected by this study in regions of equatorial Podzols of the Rio Negro basin (393 sampled layers in 18 profiles). The “soil bulk density PTF” profiles (Figure 4.1) refer to the soil profiles provided by Embrapa (2014), which were used to develop the soil bulk density PTF functions (668 sampled layers in 129 profiles). Due to the scarcity of samples in Rio Negro basin, we decided to use soil profiles in the entire Amazonia state to validate the bulk density PFT functions. The SOC stock was estimated in profiles limited to the region of the Rio Negro basin based on database provided by IBGE (2008), illustrated in Figure 4.1 by the “SOC PTF” group (1442 sampled layers in 324 profiles).

#### 4.2.1.1. *EMBRAPA Soil Database*

All samples provided by Embrapa contain values of soil pH (water and KCl); organic carbon by dichromate method (SOC); total nitrogen by Kjeldahl digestion; iron oxide ( $\text{Fe}_2\text{O}_3$ ), titanium oxide ( $\text{TiO}_2$ ); aluminium oxide ( $\text{Al}_2\text{O}_3$ ), and silicon oxide ( $\text{SiO}_2$ ) by strong acid digestion; exchangeable cations ( $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Al}^{3+}$ ) by 1 N KCl; soluble potassium and phosphorus by Mehlich 1 method (0.05 N HCl in 0.025 N  $\text{H}_2\text{SO}_4$ ). 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) measured by the hydrometer method using

Na-hexametaphosphate as chemical dispersant; soil bulk density by the core method; and water dispersible clay (WDC). A complete description of the Embrapa soil database can be found in Embrapa (2011).

#### 4.2.1.2. *IBGE Soil Database.*

The IBGE (2008) soil database was developed in the frame of the “Systematization of Natural Resources Information” project coordinated by the Natural Resources and Environmental Studies division/IBGE (CREN). The information of each soil sample was standardized in a harmonized soil geodatabase, which allows interface with GIS (Geographic Information Systems). The samples are divided by horizons according to the Brazilian Soil Classification System (EMRAPA, 2011). The database contains the same information presented in the Embrapa soil database (EMBRAPA, 2011); however, there are no soil bulk density values available for any of the sampled profiles. A full description of the IBGE soil database can be found in IBGE (2008).

#### 4.2.2. *Estimation of Soil Bulk Density*

In Podzol region (Figure 4.1), the soil bulk density was directly determined by the Kopeck ring method described by Blake et al. (1986). The remaining soil orders had their bulk density values estimated and validated based on the Embrapa (2014) database. Two aspects were considered to select the most reliable PTF function: the versatility of the proposed equation and the soil information available in the two databases used in this study (IBGE, 2008; EMBRAPA, 2014). The PTF equations were developed and evaluated using the artificial programming tool *Eureqa* (SCHMIDT; LIPSON, 2009). The database was divided in two datasets (training Dataset 1 and validation Dataset 2).

The independent soil dataset 2 was used in order to compare the proposed model (Dataset 1) with the ones presented in previous research (BERNOUX et al., 1998; TOMASELLA; HODNETT, 1998; BENITES et al., 2007). Therefore, there is no redundancy between Datasets 1 and 2. The Dataset 2 contains 230 soil samples of profiles collected in different regions of Brazilian Amazon forest (EMBRAPA, 2014) excluding the samples collected in the region of the Amazonia state. The descriptive statistics of the soil attributes used to generate the PTF are summarized in Table 4.1.

Unlike previous studies regarding the development of PFT functions in the Amazon region (BERNOUX et al., 1998; TOMASELLA; HODNETT, 1998; BENITES et al., 2007), we applied symbolic regression (SR) analysis (KOZA, 1992) in order to generate PTF equations. SR is a powerful machine learning modelling technique introduced by John and

Koza (1991). Different from linear and nonlinear regression methods, symbolic regression searches both the parameters and the form of equations, which allows the automatic generation of symbolic regression functions (SCHMIDT; LIPSON, 2009).

Table 4.1 - Descriptive statistics of the soil attributes of the training and validation datasets (Datasets 1 and 2).

Soil Attribute	Valid Cases	DATASET 1				DATASET 2				
		Min.	Max	Mean	S.D. <sup>a</sup>	Valid Cases	Min.	Max	Mean	S.D. <sup>a</sup>
Corse Sand (g kg <sup>-1</sup> )	654	1.0	704.0	162.6	160.1	230	10.0	620.0	201.7	134.6
Fine Sand (g kg <sup>-1</sup> )	654	1.0	883.0	219.3	206.0	230	10.0	620.0	243.1	174.8
Total Sand (g kg <sup>-1</sup> )	654	2.0	988.0	381.8	284.9	230	20.0	950.0	444.9	266.4
Silt (g kg <sup>-1</sup> )	654	2.0	806.0	214.6	168.1	230	20.0	482.0	116.1	91.1
Clay (g kg <sup>-1</sup> )	654	10.0	880.0	403.6	228.7	230	20.0	960.0	439.1	246.8
pH	654	3.3	7.4	4.9	0.7	230	3.6	7.3	5.3	0.9
K <sup>+</sup> (cmol <sub>c</sub> kg <sup>-1</sup> )	654	0.0	1.0	0.1	0.1	230	0.0	1.0	0.1	0.2
SiO <sub>2</sub> (g kg <sup>-1</sup> )	654	0.0	325.0	109.5	65.2	230	8.7	379.0	147.7	93.2
Al <sub>2</sub> O <sub>3</sub> (g kg <sup>-1</sup> )	654	0.0	426.0	129.3	72.6	230	6.9	345.1	150.9	101.3
Fe <sub>2</sub> O <sub>3</sub> (g kg <sup>-1</sup> )	654	0.0	467.0	55.5	50.6	230	1.4	259.0	60.9	59.1
SOC (g kg <sup>-1</sup> ) <sup>b</sup>	654	0.2	115.5	8.5	10.7	230	0.2	46.7	8.0	7.8
N (g kg <sup>-1</sup> )	654	0.1	4.7	0.9	0.8	230	0.1	10.0	0.9	1.2
C/N (%)	654	0.0	96.0	8.3	7.4	230	0.1	27.0	10.0	4.7
$\rho_p$ (g cm <sup>-3</sup> ) <sup>c</sup>	654	0.8	1.9	1.3	0.2	230	0.9	1.8	1.3	0.2

<sup>a</sup>Standard Deviation; <sup>b</sup>Total Organic Carbon; <sup>c</sup>Measured soil bulk density

We used the coefficient of determination ( $r^2$ ), the mean squared error (MSE), the root mean square error (RMSE) and Akaike's Information Criterion (AIC) in order to access the accuracy of the proposed PTF against the ones presented in previous research, considering the independent Dataset 2 (Table 4.1). MSE, RMSE and AIC are defined as:

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{\rho}_i - \rho_i) \quad \text{Eq. 4.2}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\rho_i - \hat{\rho}_i)^2} \quad \text{Eq. 4.3}$$

$$AIC = N \ln \left[ \frac{1}{n} \sum_{i=1}^n (\rho_i \hat{\rho}_i)^2 \right] + 2P \quad \text{Eq. 4.4}$$

Where  $\hat{\rho}_i$  and  $\rho_i$  are the observed and predicted soil bulk density values, respectively,  $i$  is the soil sample,  $P$  is the number of parameters used,  $n$  is the total number of observations and  $N \ln$  is the natural logarithm. The best model is that with MSE and RMSE values closer to 0 and the smaller AIC value.

### 4.2.3. Modeling the vertical distribution of SOC

The prediction of soil bulk density values based on PTF was applied to allow the estimation of SOC stock in the region of Rio Negro basin due to the absence of  $\rho_p$  values in IBGE soil database (IBGE, 2008). The  $\rho_p$  values were applied to convert SOC content from a mass basis ( $\text{kg C kg}^{-1}$ ) to a volume basis ( $\text{kg C m}^{-3}$ ). The resulting values were used to predict SOC stock ( $\text{kg C m}^{-2}$ ) in selected profiles (Figure 4.1). We divided the database in two datasets according to the fitting functions applied to model the vertical distribution of SOC stock: (i) Dataset A, which comprises soil profiles that can have their SOC stock modelled by exponential decay equations; (ii) Dataset B, referent to samples collected in field (Podzols), where the vertical distribution of SOC stock cannot be explained by exponential decay equations. The methods concerning each dataset are described below.

#### 4.2.3.1. Exponential Depth Function: Dataset A.

The following negative exponential function was fitted for each sample point in the calibration dataset from the surface to variable soil depth according to each soil profile:

$$C = C_i * \exp\left(\frac{-z}{b}\right) + y_0 \quad \text{Eq. 4.5}$$

Where  $C_i$  is SOC content in volume basis ( $\text{kg C m}^{-3}$ );  $z$  is the soil depth (m) for a given horizon;  $b$  is the SOC decay constant and  $y_0$  is the absolute value of depth (m). The integral of Equation 4.5 (Equation 4.6) represents C stock to depth  $z$  ( $d_z$ ):

$$C_t = \int_0^{d_z} \left( C_i * \exp\left(\frac{-z}{b}\right) + y_0 \right) \quad \text{Eq. 4.6}$$

Where  $C_t$  is the amount of organic C stored per unit area ( $\text{kg C m}^{-2}$ ). Integrating Equation 4.6, the C stock from the soil surface to depth  $z$  ( $d_z$ ) is given by the Equation 4.7:

$$C_t = -C_i b \exp\left(\frac{-z}{b}\right) + y_0 d_z + C_i b \quad \text{Eq. 4.7}$$

The Equation 4.7 was applied to estimate C stock in all profiles excluding regions of Podzols. The equation parameters ( $C_i$ ,  $z$ ,  $b$  and  $y_0$ ) were predicted individually for each profile with variable  $d_z$  values according to the soil depth of each observed profile. We used 25 soil profiles provided by IBGE (2008) to validate the negative exponential function (Equation 4.7). Thus, the parameters for the 25 validation profiles were generated at 1m soil depth in order to predict SOC stock at a 3m soil depth. The validation was based on measured values (IBGE, 2008) down to 3m soil depth.

#### 4.2.3.2. Podzol Depth Functions: Dataset B.

The soil samples provided by the two systematic databases available in the Amazon (IBGE, 2008; EMBRAPA, 2014) have an insignificant number of profiles collected in Podzol areas. Moreover, the sampling soil depth is always up to 2m, which is limited to the topsoil (O/A horizons), elluvial (E) horizon and the first centimetres of the spodic horizon (Bh). Given this, we collected samples in different regions of Podzols, totalizing 18 soil profiles (Figure 4.1). For each profile we have collected from 12 to 36 samples (layers), taking into account deep/thick spodic horizons (from 4.5 to 6m soil depth).

The vertical distribution of SOC stock was modelled according to three fitting models: (i) Non-parametric equal-area splines (BISHOP et al., 1999); parametric (ii) Sum of Sines and (iii) Fourier periodic fitting models (RENSHAW; FORD, 1984). It is important to highlight that periodic fitting models were not applied in previous researches related to vertical distribution of SOC stock. Usually soil attributes have a vertical behaviour that cannot be explained by periodic models. However, we observed that Podzols can have their vertical SOC content distribution fitted in these models due to a specific pattern along the soil horizons. A brief description of the Podzol fitting models is presented below.

*Equal-area splines:* The spline model assumes that soil attributes vary smoothly with depth, which is translated into mathematical terms by denoting depth by  $z$ , and the depth function describing the true attribute values by  $f(z)$ . Given this,  $f(z)$  and its first derivative  $f_0(z)$  are both continuous, and  $f_0(z)$  is square integrable. The depths of the boundaries of the  $n$  layers or soil horizons are given by  $z_0 (< z_1; \dots < z_n)$ . Thus, the measurements of  $C_i$  ( $i = 1; \dots n$ ) are mathematically modelled as:

$$C_i = \bar{f}_i + e_i, \quad \text{Eq. 4.8}$$

Where  $\bar{f}_i = \int_{z_{i-1}}^{z_i} f(z) dz / (x_i - x_{i-1})$  is the mean value of  $f(z)$  considering the interval  $(x_i - x_{i-1})$ . The errors are assumed independent, with mean 0 and common variance  $\sigma^2$ .  $f(z)$  denotes a spline function that can be determined by minimizing:

$$\frac{1}{n} \sum_{i=1}^n (C_i - f_i)^2 + \lambda \int_{z_0}^{z_n} f'(z)^2 dz \quad \text{Eq. 4.9}$$

The first term describes the model fit to data and the second one measures the roughness of function  $f(z)$ , expressed by its first derivative  $f_0(z)$ . Parameter  $\lambda$  controls the trade-off between the fit and the roughness penalty. The solution is a linear-quadratic smoothing spline (BISHOP et al., 1999). The values of SOC were included on the model as

volumetric basis (kg C m<sup>-3</sup>). The total SOC stock of each profile (kg C m<sup>-2</sup>) was calculated summing the resulting values of the fitted model. Bishop et al. (1999) has already discussed the methods to establish a proper  $\lambda$  value. Given this, we considered a standardized  $\lambda$  value of 0.1.

*Fourier series fitting:* Specific Fourier models were developed for each Podzol profile, considering sums of sine and cosine functions (Equation 4.10) assuming the behaviour of SOC along the profile as a periodic signal to a limited soil depth. The Fourier series to  $n$  terms is given by:

$$C = a_0 + \left( \sum_{i=1}^n a_i \text{COS}(i\omega z) + b_i \text{SEN}(i\omega z) \right) \quad \text{Eq. 4.10}$$

Where  $a_0$  models an intercept term in the data and is associated with the  $i = 0$  cosine term,  $\omega$  is the fundamental frequency of the signal,  $n$  is the number of terms in the series, and limited to  $1 \leq n \leq 8$ ,  $z$  is the soil depth interval. We applied four terms in order to achieve the best model adjustment. The resulting parameters were used to estimate the SOC stock in the whole profile, after applying the integration of the Fourier series (Equation 4.11).

$$C_t = \int_0^{d_z} \left( a_0 + \sum_{i=1}^n a_i \text{COS}(i\omega z) + b_i \text{SEN}(i\omega z) \right) \quad \text{Eq. 4.11}$$

Where  $C_t$  is the amount of organic C stored per unit area (kg C m<sup>-2</sup>) and  $d_z$  refers to the profile depth. After integrating the 4 terms of the Fourier series function to  $d_z$  depth, we obtained the following equation:

$$C_t = \frac{0.25a_4 \sin(4\omega d_z)}{\omega} - \frac{0.25b_4 \cos(4\omega d_z)}{\omega} + \frac{0.333a_3 \sin(3\omega d_z)}{\omega} - \frac{0.333b_3 \cos(3\omega d_z)}{\omega} + \frac{0.5a_2 \sin(2\omega d_z)}{\omega} - \frac{0.5b_2 \cos(2\omega d_z)}{\omega} + \frac{1a_1 \sin(\omega d_z)}{\omega} - \frac{1b_1 \cos(\omega d_z)}{\omega} + 1a_0 d_z + \frac{0.25b_4}{\omega} + \frac{0.333b_3}{\omega} + \frac{0.5b_2}{\omega} + \frac{1b_1}{\omega} \quad \text{Eq. 4.12}$$

Where the parameters  $a_0$ ,  $a$  and  $b$  are given for each Fourier term, with 95% confidence bounds.

*Sum of Sines fitting:* The Sum of Sines is similar to the Fourier fitting. However, it includes the phase constant and does not include a constant term. The Sum of Sines function is represented by the following equation:

$$C = \sum_{i=1}^n a_i \sin(b_i z + c_i) \quad \text{Eq. 4.13}$$

Where  $a$  is the amplitude,  $b$  is the frequency, and  $c$  is the phase constant for each sine wave term.  $n$  is the number of terms in the series. We also included four terms to fit the SOC stock in Podzol profiles. The integration to  $d_z$  soil depth is shown in Equation 4.14.

$$C_t = \int_0^{d_z} \left( \sum_{i=1}^n a_i \sin(b_i x + c_i) \right) \quad \text{Eq. 4.14}$$

After integrating the four terms of the Sum of Sines fitting to  $d_z$ , we obtained the Equation 4.15 as follows:

$$C_t = \frac{-(a_4 \cos(b_4 d_z + c_4))}{b_4} - \frac{a_3 \cos(b_3 d_z + c_3)}{b_3} - \frac{a_2 \cos(b_2 d_z + c_2)}{b_2} - \frac{a_1 \cos(b_1 d_z + c_1)}{b_1} + \frac{a_4 \cos(c_4)}{b_4} + \frac{a_3 \cos(c_3)}{b_3} + \frac{a_2 \cos(c_2)}{b_2} + \frac{a_1 \cos(c_1)}{b_1} \quad \text{Eq. 3.15}$$

The parameters  $a$ ,  $b$  and  $c$  are given for each Sum of Sines term with 95% confidence bounds. The evaluation of results was carried out by comparing observed and predicted SOC stock values, considering the coefficient of determination ( $r^2$ ), MSE, RMSE and AIC.

### 4.3. Results

#### 4.3.1. Predicting Soil Bulk Density in Amazon Soils

In the first attempt to generate a PTF function (Dataset 1), we considered all soil attributes presented in Table 4.1. The best model was achieved (Model 1: Eq. 3.16) with the following input data: fine sand, silt, clay, total N and C/N. The symbolic regression analysis considering all input data returned a generalized equation (Equation 4.16) that explained 70% of the soil bulk density variance (Figure 4.2a). The RMSE and MSE between predicted and

$$\rho_p = 1.463 + 0.1998 \tan(1.044 - 0.002(\text{clay})) \cos(0.125 + 0.135(C/N) + (3.543 \cdot 10^{-5})(\text{silt}^2) - 0.013(\text{silt})) \cos(0.004(\text{fine sand}) \cos(0.315 + \tan(0.005(\text{clay}) - 2.317))) - 1.065 \cos(0.315 + \tan(0.005(\text{clay}) - 2.317))) - 0.144(\text{total N}) \quad \text{Eq. 4.16}$$

observed values were  $0.011 \text{ g cm}^{-3}$  and  $0.108 \text{ g cm}^{-3}$ , respectively.

Arithmetic and trigonometric operators were selected by the user and automatically added to the final equation. As pointed out by Benites et al. (2007), a better correlation between N and  $\rho_p$ , when compared to SOC content is observed in Embrapa (2014) database. That might be related to the total SOC measurement procedure (acid-dichromate  $\text{FeSO}_4$  titration procedure) adopted by Embrapa. The use of C/N and N values on the resulting symbolic regression equation (Equation 4.16) might poses a problem towards the proposition

of a general PFT equation for Amazon soils due to the lack of such data in most of soil databases currently available. Thus, we used the following input data to train the symbolic regression model (Model 2: Equation 4.17; Figure 4.2b): total sand, silt, clay, pH and SOC.

$$\rho_p = 1.326 + 0.315 \sin(1.045 - 0.001(\text{clay}) - 0.052(\text{SOC}) + 0.0003(\text{clay}) \sin(\sin(2.561 + 1.287(\text{pH}) - 0.006(\text{clay}))) - 0.134 \sin(\sin(2.561 + 1.287(\text{pH}) - 0.006(\text{clay}))) \quad \text{Eq. 4.17}$$

The Model 2 (Equation 4.17) had a lower correlation with the observed dataset (Figure 4.2b), however It takes into account three soil attributes that are widely available in most of systematic soil databases (EMBRAPA and IBGE legacy data). Therefore, Model 2 was applied in order to predict values of soil bulk density. The Model 2 had an RMSE and MSE of  $0.015 \text{ g cm}^{-3}$  and  $0.123 \text{ g cm}^{-3}$ , respectively. The validation was based on the independent Dataset 2, as described below.

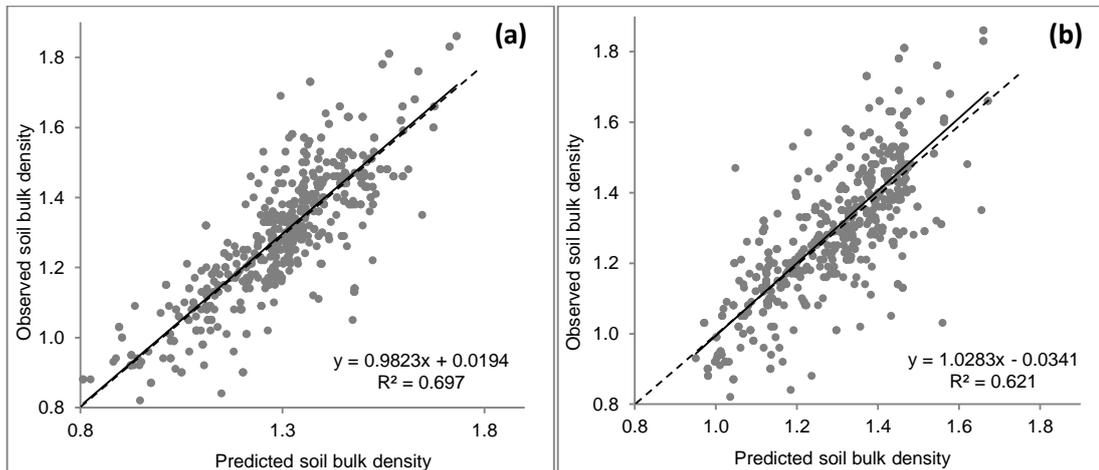


Figure 4.2 - Plot of the predicted data against the observed data. (a): Model 1; (b): Model 2. Dashed lines are the 1:1 lines.

#### 4.3.1.1. Symbolic Regression Model Validation

The proposed Model 2 (Figure 4.3) has shown the best performance, among the evaluated PTF functions, with MSE and RMSE closest to 0 and the lower AIC value. However, the model proposed by Benites et al. (2007) has a similar behavior with close MSE, RMSE and AIC indices. Thus, we observed that Benites' et al. (2007) model could be applied to estimate soil bulk density in Amazon soils, but with the disadvantage of demanding  $\text{Fe}_2\text{O}_3$  values, which are not available in most of the soil profiles provided by IBGE (2008).

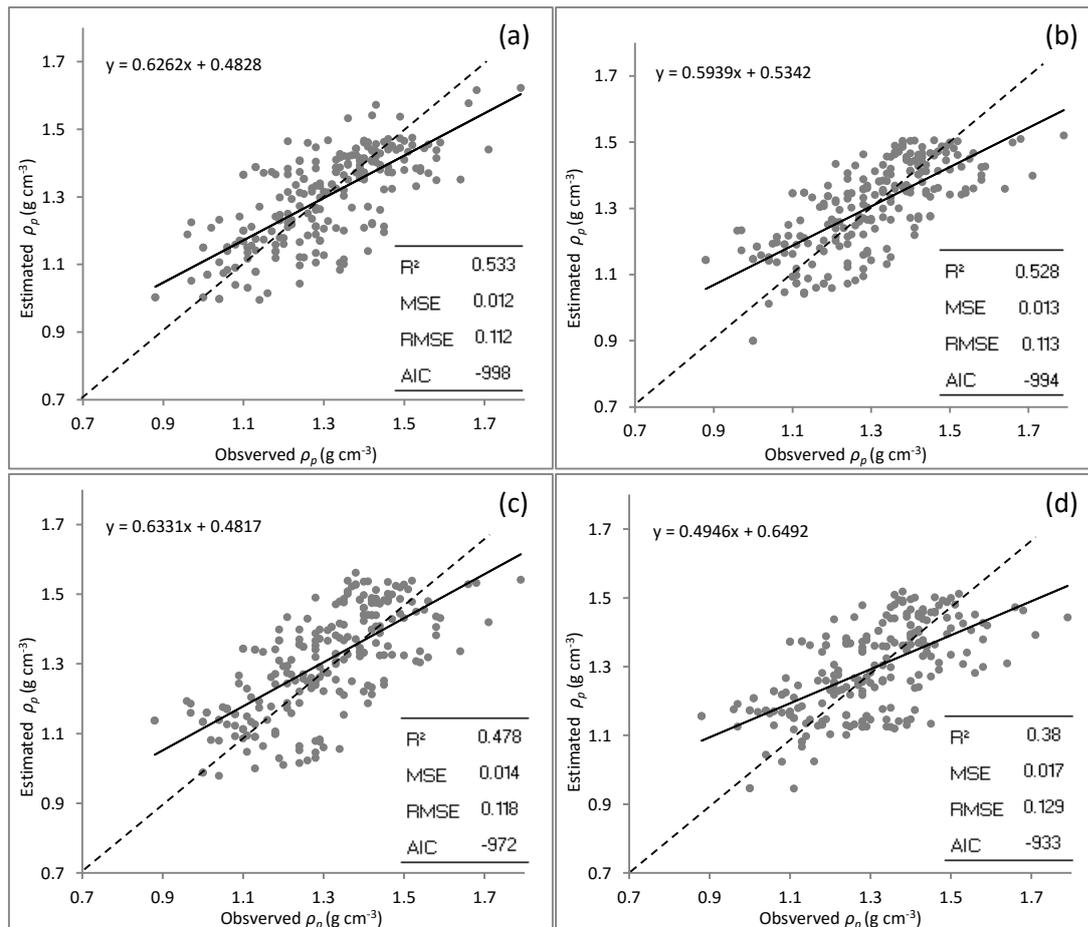


Figure 4.3 - Scatterplots and goodness of fit indexes of proposed and previous soil bulk density PTF functions. (a): Proposed Model 2; (b): Benites et al. (2007); (c) Bernoux et al. (1998) and (d): Tomasella; Hodnett (1998). Dashed lines are the 1:1 lines.

Clay content and SOC have been reported in previous studies as the most relevant attributes to explain soil bulk density variability (BERNOUX et al., 1998; BENITES et al., 2007). Given the availability of soil textural fraction, pH and SOC data in the applied soil databases (IBGE, 2008; EMBRAPA, 2014), we decided to use the Model 2 in order to estimate soil bulk density values. It is important to highlight that this model was developed based on soil samples limited to the region of the Amazon basin, which might derail its application in areas outside Amazon biome.

#### 4.3.2. Modeling the vertical distribution of SOC stock in Amazon soils

The main soil orders in Rio Negro basin are Ferralsols (34% of the region); Acrisols (22% of the region) and Podzols (19% of the region). The remaining orders comprise Gleysols (6%) and Plinthosols (5%). Arenosols, Nitisols and Planosols account to less than 10% of the soils of Rio Negro basin. At the first 0.3m soil depth we observed that Ferralsols and Acrisols have a mean SOC content of  $1.8 \pm 1.4\%$  and  $1.5 \pm 1.1\%$ , respectively. Below 0.3m (0.3 to 0.8m) the SOC content in these soils decays to  $0.57 \pm 0.5\%$  in Ferralsols

and  $0.44 \pm 0.6\%$  in Acrisols. The superficial horizons of Podzols have a higher carbon concentration (0-0.3m:  $2.7 \pm 1.5\%$ ). The same pattern was observed in deep thick Bh horizons (1 to 3m soil depth) where the mean carbon content is  $2.31 \pm 2.1\%$ .

All soils that had their SOC stock modeled by exponential depth functions were grouped together (Dataset A). After integration (Equation 4.9), the exponential functions showed a mean  $r^2$  value of 0.99 and a RMSE of  $0.85 \text{ kg C m}^{-2}$  between the observed and fitted SOC stock (Figure 4.4a). These results indicate that the exponential depth functions fitted the data very well, with an  $r^2$  closer to 1 and an RMSE below  $1 \text{ kg C m}^{-2}$ . It's important to highlight that 5% of the soil profiles originally provided by IBGE (2008) were not fitted to exponential equations due to the low number of observed soil layers (2.4% of profiles) or because of the occurrence of high amounts of SOC content in horizons below 0.3m (2.6% of profiles).

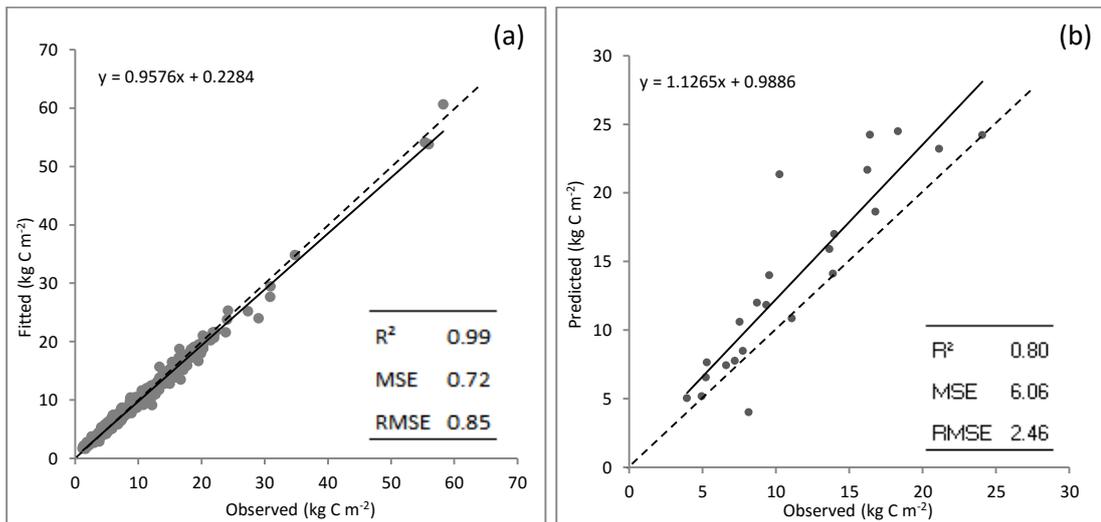


Figure 4.4 - (a) Observed and fitted exponential depth function SOC; (b) Observed and predicted exponential depth function, based on the validation dataset. Dashed line is the 1:1 line.

As we can see in Figure 4.4b the predicted SOC values fitted well to the observed data with an RMSE of  $2.5 \text{ kg C m}^{-2}$  considering a 3m soil depth, which allows the prediction of SOC stock in Amazon soils in deeper horizons (below 1m). The validation dataset comprises soil profiles of Acrisols and Ferralsols, which are dominant in Amazon basin. Therefore, exponential depth functions offer a feasible way to estimate SOC stock in Amazon biome. However, this assumption is not valid for Podzols due to the peculiar distribution of SOC along the profile (Figure 4.5).

The pedogenetic processes involving the formation of Podzols have already been investigated by several studies in Amazon (LUCAS et al., 1984; 1988; 1996; CHAUVEL et al., 1987; BRAVARD; RIGHI, 1990; DUBROEUCQ; VOLKOFF, 1998;

NASCIMENTO et al., 2004; MONTES et al., 2007; PATEL-SORRENTINO et al., 2007; FRITSCH et al., 2009; MONTES et al., 2011). All researches developed in this region have described the occurrence of sandy soil materials (E horizon) that lead to the leaching of Al and Fe organic matter complexes, resulting in the dissolution of clay minerals, Al-hydroxides and Fe-oxides or Fe-oxyhydroxides, causing the formation of illuvial deep Bh rich-SOC horizons. Based on these characteristics we divided Podzol profiles in four systematic horizons according to their SOC content: 1. SOC-rich topsoil horizon (A/O); 2. Eluvial sandy horizon with insignificant amount of SOC; 3. Deep thick SOC-rich Bh horizon; 4. C horizon with a gradual decrease in SOC content (Figure 4.5).

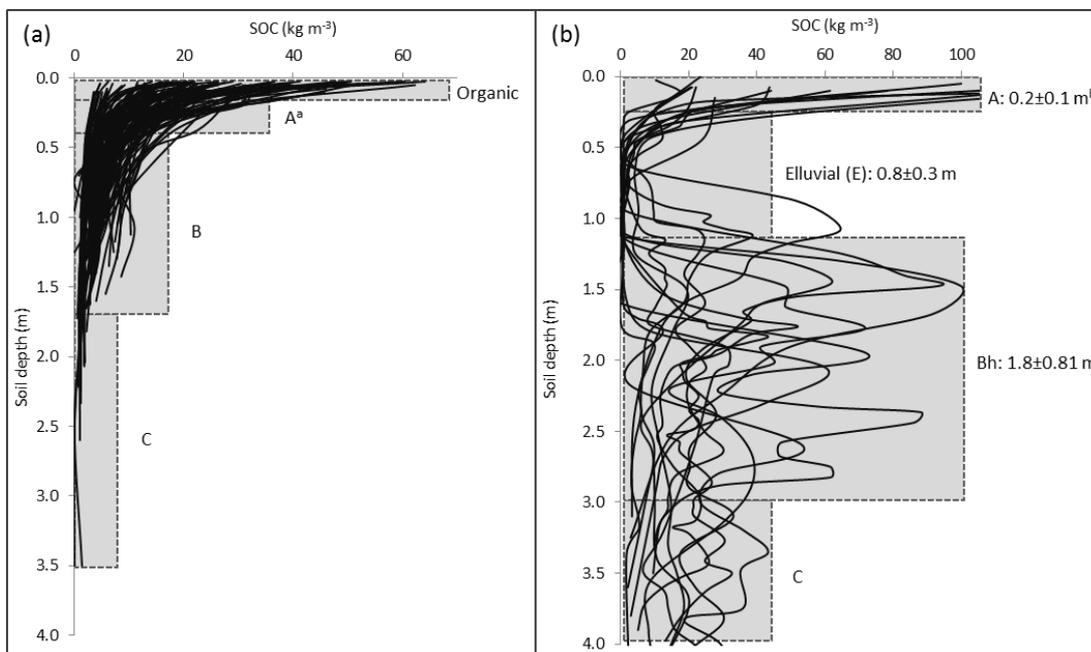


Figure 4.5 - Measured SOC stock. (a): Dataset A (Ferralsols and Acrisols); (b): Dataset B (Podzols). <sup>a</sup>Typical Ferralsol horizons. <sup>b</sup>Typical Podzol horizons with average thickness for evaluated Podzol profiles.

We observed a clear periodical pattern that fit very well in Sum of Sines and Fourier models (Table 4.2). Nevertheless, spline models have generated the best predicted values when compared to measured data (Table 4.2). It's important to emphasise that the establishment of  $\lambda$  (lambda) values is laborious and depends on the availability of several soil samples for each soil profile, which allows an appropriate representation of the soil attribute to be measured. Given this, we decided to apply the  $\lambda$  value of 0.1 as suggest by Bishop et al. (1999). The periodical models were fitted to the observed data with 2, 3 and 4 terms. The best fitting was achieved with 4 terms in both Fourier and Sum of Sines models.

Table 4.2 - Evaluation indices for the three fitting models considering Dataset B (Podzols).

Fitting Model	Observations	R <sup>2</sup>	MSE (kg C m <sup>-2</sup> )	RMSE (kg C m <sup>-2</sup> )	AIC
Equal-area Spline	18	0.85	187.21	13.68	96.06
Sum of Sines	18	0.82	225.38	15.00	99.40
Fourier	18	0.79	255.12	15.97	101.63

The curve fitting of Dataset B was created considering all measured layers of each Podzol profile as shown on the example of Figure 4.6. We observed a complex distribution of SOC along a typical Podzol profile. As we can see in Figure 4.6 the SOC content in Bh horizon is highly variable with abrupt changes in depth intervals lower than 0.05m (Figure 4.5). The variation in C content within Bh horizon is explained by pedogenitic processes of this horizon. What is observed here is a process of reduction and re-oxidation of organometallic complexes leading to the selective accumulation of different amounts of C in Bh horizon.

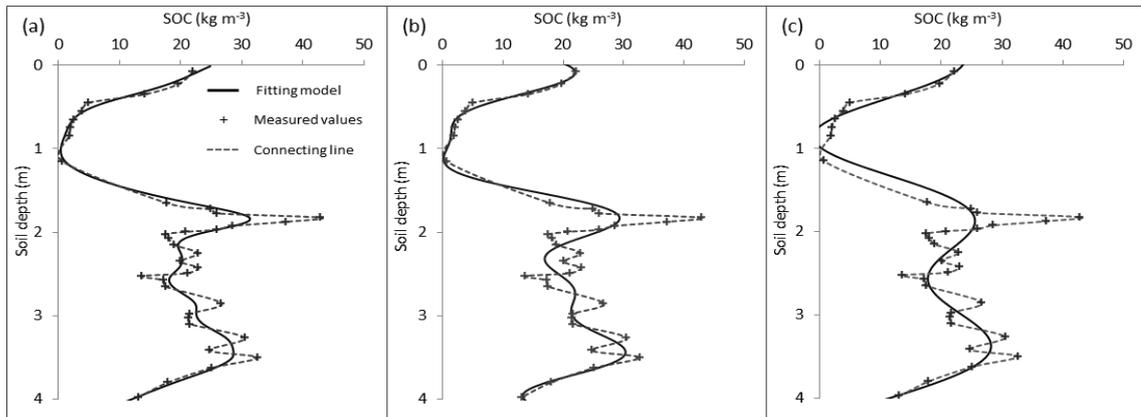


Figure 4.6 - Example of fitting models to a typical Podzol profile. (a) Equal-area Spline; (b) Sum of Sines; (c) Fourier.

Considering the parametric equations, we observed a better performance of the Sum of Sines fitting, confirmed by the quality evaluation indices (Table 4.2). Thus, the integration of the Sum of Sines model to the observed soil depth (Equation 4.15) was capable of representing the complex distribution of SOC stock within Bh horizons (Figure 4.6), which justifies the application of a 4 terms Sum of Sines model. It's interesting to highlight that Sum of Sines as well as Fourier models, deals with trigonometric and circular functions, usually applied to describe attributes with clear periodical behaviour. In this context, we assumed that SOC stock in Podzols has a periodical pattern, which implies in modelling the profile to a limited range according to the observed data, where the assumption of periodicity is attested. Therefore, the prediction of values is limited to the measured soil depth, which means that we cannot predict SOC stock values beyond the measured layers considering the three fitting models: Spline, Sum of Sines and Fourier.

### 4.3.3. *The SOC stock in Dataset A and B*

In Dataset A, we observed that a significant portion of the SOC stock is located at the first 0.3m soil depth with an average stock of  $8.2 \pm 5.0 \text{ kg C m}^{-2}$ . Such value represents 41% of the total SOC stock at 3m soil depth. At the first soil meter we observed that the profiles evaluated in Dataset A store about  $15.2 \pm 6.2 \text{ kg C m}^{-2}$ , which is 76% of the total SOC stored in profiles up to 3m soil depth. Finally, if we take into account the entire soil profile (3m soil depth) the SOC stock increases to a value of  $19.2 \pm 10.7 \text{ kg C m}^{-2}$ . Thus the carbon stored in deep soil horizons, from 2 to 3m soil depth, represents about 19.2% of the total SOC stored in the first 3m soil depth of Dataset A. We observed high standard deviation values due to the grouping of different soil orders into the same dataset in order to compare the vertical behavior of SOC stock distribution in Dataset A against Dataset B (Podzols). Different from other Amazon soils, Podzols have shown a high capacity of storing huge amounts of SOC in deep thick horizons (Bh), with a complex vertical distribution that was not evaluated in previous studies.

In Podzols the organic (O/A) horizon stores higher amounts of C when compared to other Amazon soils due to the prevalence of hydromorphic conditions that leads to the accumulation of fresh MOS in surface. In O/A horizon the average SOC stock is  $17.9 \pm 11 \text{ kg C m}^{-2}$  at 0.3m soil depth. In elluvial sandy horizons (E) the SOC stock abruptly decays to an average value of  $4.8 \pm 2.7 \text{ kg C m}^{-2}$ . The thickness of E horizon is variable ranging from 0.5 to 1.5m, depending on the observed profile. The Bh horizon stores a mean SOC value significantly high with an average stock of about  $83.2 \pm 15.5 \text{ kg C m}^{-2}$  to a 2m horizon thickness. It is important to highlight that the upper and bottom limits of Bh horizon are variable which implies in different Bh thickness according to the observed profiles, however the values presented above were taken to a 3m soil depth. Nevertheless, Bh horizons might extend to a 5m soil depth, which would increase the average SOC stock of Podzols.

## 4.4. **Conclusions**

The good performance of the exponential depth function was attested down to 3m soil depth based on the validation dataset. Thus, the application of exponential models to predicted SOC stock in Amazon soils has proven to be efficient, considering the availability of measured values to the first soil meter. Nevertheless, information of soil bulk density was essential to allow the systematic estimation and prediction of SOC stock. The prediction of soil bulk density data was possible by the application of PTF equations developed according to symbolic regression analysis (SR), which generated a dynamic model suited to predicting

soil bulk density specifically in Amazon soils. Soil bulk density PTF models and exponential depth functions have allowed the estimation of SOC stock in Amazon soils where the assumption of SOC content exponential decay was attested. The prediction of SOC stock by exponential decay equations is simple and can be carried out indirectly by integrating the exponential model to a desired soil depth, which is helpful to estimate deep SOC stock in Amazon soils.

In Podzols, the SOC stock is significantly higher than in other Amazon soils, especially below 1m soil depth. The vertical variation of SOC was successfully modeled by parametric and non-parametric fitting models. The non-parametric equal-area spline model returned the best predicted values. Moreover, the fitted spline curves are not affected by local variations, due to the possibility of fitting piece wise a series of local independent functions over small intervals (soil depths). However, the application of parametric models might be helpful to allow the indirect prediction of SOC stock in Podzols and to describe the general behavior of SOC along the soil profile. In this matter, we observed that Sum of Sines models, yet not explored in predicting soil attributes, can be properly applied to describe and estimate the SOC stock distribution in Amazon Podzols with deep thick Bh horizons.

## References

ARROUAYS, D.; P'ELISSIER, P. Modeling carbon storage profiles in temperate forest humic loamy soils of France. **Soil Science**, New Brunswick, v. 157, p. 185–192, 1994.

BATJES, N.H. Total carbon and nitrogen in the soils of the world. **European Journal of Soil Science**, Oxford, v. 47, p. 151-163, 1996.

BATJES, N.H.; DIJKSHOORN, J.A. Carbon and nitrogen stocks in the soils of the Amazon Region. **Geoderma**, Amsterdam, v. 89, p. 273–286, 1999.

BENITES, V.M, MACHADO, P., FIDALGO E.C.C., COELHO, R.M., MADARI, E.B. Pedotransfer functions for estimating soil bulk density from existing soil survey reports in Brazil. **Geoderma**, v. 139, p. 90-7, 2007.

BERNOUX, M.; ARROUAYS, D.; CERRI, C.; VOLKOFF, B.; JOLIVET, C. Bulk densities of Brazilian Amazon soils related to other soil properties. **Soil Science Society of America Journal**, Madison, v. 162, p. 743–749, 1998.

BERNOUX, M.; CARVALHO, M.D.S.; VOLKOFF, B.; CERRI, C. C. Brazil's soil carbon stocks. **Soil Science Society of America Journal**, Madison, v. 66, p. 888–896, 2002.

BRAVARD, S.; RIGHI, D. Geochemical differences in an oxisolspodosol toposequence of Amazonia (Brazil). **Geoderma**, Amsterdam, v. 44, p. 29–42, 1989.

BISHOP, T.F.A.; MCBRATNEY, A.B.; LASLETT, G.M. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. **Geoderma**, Amsterdam, v. 91, p. 27-45, 1999.

BLAKE, G.R.; HARTGE, K.H. Bulk Density. In: KLUTE, A. (Ed.). **Methods of soil analysis**. Part 1. Madison: SSSA, 1986. p. 363-376.

BURINGH, P. Organic carbon in the soils of the world. In: WOODWELL, G. (Ed.). **The role of terrestrial vegetation in the global carbon cycle: measurement by remote sensing**. Chichester: Wiley, 1994. p. 91-109. (SCOPE, v. 23).

BURKE, I.C.; YONKER, C.M.; PARTON, W.J.; COLE, C.V.; FLACH, K.; SCHIMEL, D.S. Texture, climate, and cultivation effects on soil organic matter content in U.S. grassland soils". **Soil Science Society of America Journal**, Madison, v. 53, p. 800-805, 1989.

CHAUVEL, A.; LUCAS, Y.; BOULET, R. On the genesis of the soil mantle of the region of Manaus, Central Amazonia, Brazil. **Experientia**, Basel, v. 43, p. 234-241, 1987.

DUBROEUCQ, D.; VOLKOFF, B. From oxisols to spodosols and histosols: evolution of the soil mantles in the Rio Negro Basin (Amazonia). **Catena**, Amsterdam, v. 32, p. 245-280, 1998.

EMBRAPA. **Manual de métodos de análises de solos**. 2. ed. Rio de Janeiro: Embrapa Solos, 2011. 230 p.

EMBRAPA. **Digital Soils System Information** (Brazilian Soils Database). Rio de Janeiro: Embrapa Solos, 2014. Available at: [http://www.bdsolos.cnptia.embrapa.br/consulta\\_publica.html](http://www.bdsolos.cnptia.embrapa.br/consulta_publica.html). Accessed in: Oct. 25, 2014.

ESWARAN, H.; VAN DEN BERG, E.; REICH, P. Organic carbon in soils of the world. **Soil Science Society of America Journal**, Madison, v. 57, p. 192-194, 1993.

FAO/IIASA/ISRIC/ISSCAS/JRC. **Harmonized World Soil Database** (version 1.2). FAO, Rome, 2012.

FRITSCH, E.; ALLARD, T.; BENEDETTI, M.F.; BARDY, M.; DO NASCIMENTO, N.R.; LI, Y.; CALAS, G. Organic complexation and translocation of ferric iron in podzols of the Negro River watershed. Separation of secondary Fe species from Al species. **Geochimica et Cosmochimica Acta**, Amsterdam, v. 73, n. 7, p. 1813-1825, 2009.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. Geoscience Division - DGC. Coordenação de Recursos Naturais e Estudos Ambientais - CREN. **Georefered Maps of Natural Resources: Scale 1:250:000**. Digital Format: shp. Rio de Janeiro, 2008. Available at: [ftp://geofp.ibge.gov.br/mapas/banco\\_dados\\_georeferenciado\\_recursos\\_naturais](ftp://geofp.ibge.gov.br/mapas/banco_dados_georeferenciado_recursos_naturais). Accessed in: Aug. 15, 2014.

KIMBLE, J.M.; HEATH, L.S.; BIRDSEY, R.; LAL, R. **The Potential of US. Forest Soils to Sequester Carbon and Mitigate the Greenhouse Effect**. Boca Raton: CRC Press, 1990. 429 p.

KOZA, J. Evolving a computer program to generate random numbers using the genetic programming paradigm. In: INTERNATIONAL CONFERENCE ON GENETIC ALGORITHMS, 4., 1991, La Jolla, CA. **Proceedings...** La Jolla, CA: Morgan Kaufmann, 1991.

KOZA, J. **Genetic Programming: On the Programming of Computers by Means of Natural Selection.** Cambridge, MA: The MIT Press, 1992.

LUCAS, Y.; CHAUVEL, A.; BOULET, R.; RANZANI, G.; SCATOLINI, F. Transição latossolos-podzols sobre a formação Barreiras na região de Manaus, Amazônia. **Revista Brasileira de Ciências do Solo**, Viçosa, v. 8, p. 325-335, 1984.

LUCAS, Y.; BOULET, R.; CHAUVEL, A. Intervention simultanée des phénomènes d'enfoncement vertical et de transformation latérale dans la mise en place de systèmes de sols de la zone tropicale humide. Cas des systèmes sols ferrallitiques – podzols de l'Amazonie Brésilienne. **Comptes Rendus de l'Académie des Sciences, Ser. IIa**, Paris, v. 306, p. 1395–1400, 1998.

LUCAS, Y.; NAHON, D.; CORNU, S.; EYROLLE, F. Genèse et fonctionnement des sols en milieu équatorial. **Comptes Rendus de l'Académie des Sciences, Ser. IIa**, Paris, v. 322, p. 1–16, 1996.

McBRATNEY, A.B.; MENDONÇA-SANTOS, M.L.; MINASNY, B. On digital soil mapping. **Geoderma**, Amsterdam, v. 117, p. 3–52, 2003.

MINASNY, B.; McBRATNEY, A.B.; MENDONÇA-SANTOS, M.L.; ODEH, I.O.A.; GUYON, B. Pre-diction and digital mapping of soil carbon storage in the Lower Namoi Valley. **Australian Journal of Soil Research**, Melbourne, v. 44, p. 233–244, 2006.

MONTES, C.R.; LUCAS, Y.; MELFI, A.J.; ISHIDA, D.A. Systèmes sols ferrallitiques–podzols et genèse des kaolins. **Comptes Rendus Geoscience**, Paris, v. 339, p. 50–56, 2007.

MONTES, C.R.; LUCAS, Y.; PEREIRA, O.J.R.; ACHARD, R.; GRIMALDI, M.; MELFI, A.J. Deep plant-derived carbon storage in Amazonian podzols. **Biogeosciences**, Göttingen, Germany, v. 8, p. 113-120, 2011.

NASCIMENTO, N.R.; BUENO, G.T.; FRITSCH, E.; HERBILLON, A.J.; ALLARD, T.; MELFI, A.; ASTOLFO, R.; BOUCHER, H.; LI, Y. Podzolization as a deferrallitization process: a study of an Acrisol- Podzol sequence derived from Palaeozoic sandstones in the northern upper Amazon Basin. **European Journal of Soil Science**, Oxford, v. 55, p. 523–538, 2004.

PATEL-SORRENTINO, N.; LUCAS, Y.; EYROLLES, F.; MELFI, A.J. Fe, Al and Si species and organic matter leached off a ferrallitic and podzolic soil system from Central Amazonia. **Geoderma**, Amsterdam, v. 137, p. 444–454, 2007.

PEREIRA, O.J.R.; MONTES, C.R.; LUCAS, Y.; SANTIN, R.C.; MELFI, A.J. A multisensor approach for mapping plant-derived carbon storage in Amazonian podzols. **International Journal of Remote Sensing**, London, v. 36, n. 8, p. 2076-2092, 2015.

POST, W.M.; EMANUEL, W.R.; ZINKE, P.J.; STANGENBERGER, A.G. Soil carbon pools and world life zones. **Nature**, London, v. 298, p. 156-159, 1982.

RENSHAW, E.; FORD, E.D. The description of spatial pattern using two-dimensional spectral analysis. **Vegetatio**, Dordrecht, v. 56, p. 75-85, 1984.

SCHMIDT, M.; LIPSON, H. Distilling Free-Form Natural Laws from Experimental Data. **Science**, Washington, DC, v. 324, n. 5923, p. 81–85, 2009.

TOMASELLA, J.; HODNETT, M.G. Estimating soil water retention characteristics from limited data in Brazilian. **Soil Science**, New Brunswick, v. 163, p. 190–202, 1998.

WEBSTER, R. Mathematical treatment of soil information. In: INTERNATIONAL CONGRESS OF SOIL SCIENCE, 11., 1978, Edmonton, Canada. **Proceedings...** Edmonton, Canada: University of Alberta, 1978. v. 3, p. 161–190.

## 5. Mapping deep plant-derived soil carbon storage in soils of the Rio Negro basin

### Abstract

Despite its importance as a carbon reservoir at global scale, the Amazonia region still needs detailed surveys evolving the quantification and mapping of soil organic carbon (SOC) stock at regional and continental scales. The precise estimation of carbon stock is essential for the assessment of carbon sequestration capacity, greenhouse gas emissions and national carbon balance inventories, which depends on the understanding of SOC storage capacity. Some effort has been made in the last decade towards the estimation of SOC stock of the Amazon region. However, most of the research developed in this vast region, takes into account surface and sub-superficial soil horizons up to 0.3m. Moreover, these studies are not systematically organized, which makes it difficult the generalization of the SOC stock to smaller map scales. Therefore, in the present research, we applied legacy data, field sample data and remote sensing imagery to quantify and map the SOC stock in deep soil horizons in the region of Rio Negro basin. Ordinary kriging (OK) and regression kriging (RK) were employed to generate the SOC stock maps at Rio Negro basin map scale, at 1m and 3m soil depth. Both kriging methods generated similar results, with a better performance for RK at 1m soil depth and for OK at 3m soil depth. Nevertheless, the mapping of SOC stock in Rio Negro basin by RK, allowed for a more detailed spatialization of SOC stock distribution according to an ancillary database. Therefore, the RK resulting maps were used to estimate the SOC stock in the study area. According to RK map, the Rio Negro basin had an absolute SOC stock of about 5.75 Pg at 1m soil depth and 10.12 Pg at 3m soil, which is about twice the value found on the first soil meter. However, if we take into account an average stock of 80 kg C m<sup>-2</sup> found in Podzols at 6m soil depth, these soils have an absolute SOC stock of 9.19 Pg, within the area of the Rio Negro basin.

### 5.1. Introduction

The evaluation of the soil organic carbon (SOC) stock at regional and national map scales depends on the availability of systematic well distributed soil samples, which might be not available in most of the tropical forests of the world (BATJES, 1996; 1997). In extensive natural regions (eg.: Amazon Forest), the only source of soil information comprises legacy databases (eg.: data derived from conventional soil surveys) provided by governmental agencies or academic researches (CERRI et al., 1999; BERNOUX et al., 2002; CERRI et al., 2007; IBGE, 2008; MONTES et al., 2007; 2011; PEREIRA et al., 2015). Given this, the systematization of available information in a harmonized database is important in order to combine multisource information that can be used to estimate SOC stock based on soil profile data. The estimation of SOC stock depends mostly on the adoption of multisource legacy database due to the lack of well distributed field sample data in Amazon Forest. Moreover, the application of pedotransfer functions (PTF) is necessary to standardize multisource information and to predict soil attributes based on soil sample data (MONTES et al., 2011).

Numerous studies have attempted to assess the potential of the soils to store and retain carbon as organic matter at the first soil meter (POST et al., 1982; KIMBLE, 1990; BURINGH, 1994; ESWARAN et al., 1993; BATJES, 1996; CERRI et al., 1999; BERNOUX et al., 2002). However, less attention is given to the role of soil carbon pools stored in deep soil horizons below 1m soil depth, on modelling future scenarios of global climate change. As pointed out by Batjes (1996), regions of tropical Ferralsols and Acrisols might have an increase of about 50% on their SOC stock if the second soil meter is taken into account. According to the guide of good practice for land use - IPCC (DRÖSLER et al., 2013), it is important to measure the soil carbon pool at soil depth of at least 0.30m. Therefore, depending on the soil type and the vegetation cover, the depth for SOC content measurement is variable, considering the SOC pool more likely to be mineralized due to changes on the natural soil cover or soil hydrologic regime (MONTES et al., 2011). In areas of Amazon ferralitic soils (Ferralsols and Acrisols) an exponential decay in SOC stock with depth, was attested (BERNOUX et al., 1998; BERNOUX et al., 2002), which implies on significant diminution on SOC content in the second soil meter. Given this, the SOC stock estimates in those soils are always limited to superficial and sub-superficial horizons.

In Amazon Podzols, the vertical distribution of SOC stock has a different pattern when compared to neighbouring soils (Ferralsols and Acrisols). According to recent research developed in this region (MONTES et al., 2011), these soils have a great capacity of storing carbon in surface horizons (A/O horizon) and in deep thick spodic horizons (Bh). The Bh horizon has an average thickness of about  $1.8 \pm 0.81$ m and starts at a soil depth ranging from 0.9 to 2.3m (PEREIRA et al., 2015b<sup>4</sup>). Therefore, the acquisition of soil samples in Podzols is laborious and demands for innovative techniques (MONTES et al., 2011). Accordingly, detailed information regarding the SOC storage capacity of Amazon Podzols, is an open debate (BATJES, 2002; EASTER et al., 2007; IBGE, 2008; EMBRAPA, 2014). Thus, new approaches are necessary in order to map the occurrence of Podzols and its SOC storage capacity in Amazon region.

The SOC storage capacity of Amazon soils depends on a range of factors such as soil types, land cover, annual input of vegetation biomass, soil moisture, topography, lithology and annual rainfall. These factors can be related to the lateral variation of SOC stock through the application of spatially depended multiple regression analysis (VAN-MEIRVENNE et al.,

---

<sup>4</sup> PEREIRA, O.J.R.; MONTES, C.R.; LUCAS, Y.; MELFI, A.J. Evaluation of pedotransfer equations to predict depth soil carbon stock in tropical podzols compared to other soils of Brazilian Amazon Forest. Digital Soil Morphometrics: Srings, v. 1, 2015. *No prelo.*

1996; POST et al., 2001). Therefore, thematic maps and remote sensing images associated to field sample data are useful in order to map the SOC stock at different map scales, from local to continental maps (MISHRA et al., 2010). The approaches to obtain maps of SOC storage varies from the measure-and-multiply method (the study area is separated into different strata and the SOC measurements within each stratum are multiplied by the area of that strata) to more sophisticated approaches that takes into account the spatial variability of SOC and the spatial correlation with ancillary environmental variables (MINASNY et al., 2006; SIMBAHAN et al., 2006).

Many studies have explored the capability of geostatistics, artificial neural networks, and multiple regression techniques to map SOC stock in local and regional map scales (MINASNY et al., 2006; SIMBAHAN et al., 2006; MEERSMANS et al., 2008). However, a fewer attention is given to extensive tropical regions, where these techniques remains unexplored. Given this, the main goal of this research was to map the SOC stock of Amazon soils located in the region of the Rio Negro basin, based on different interpolation methods, in order to generate maps of deep SOC stock. The prediction of SOC stock was carried out within 1m and 3m soil depth at Rio Negro basin map scale, using profile depth functions, ordinary kriging and regression kriging.

## **5.2. Methodology**

### **5.2.1. Study Area**

The study was developed in the region of Rio Negro basin, within its Brazilian portion, comprising an area of 603,661 km<sup>2</sup> (Figure 5.1), which represents around 16% of the area of the Amazon basin. In this region prevailing soils are Ferralsols, Acrisols and Podzols and their distribution is usually related to the parent material (lithology). The Rio Negro basin comprises a vast region of low level plateaus called “*Pediplano Rio Branco-Rio*” (BRASIL, 1977), in which the only emerging reliefs are granitic inselbergs, quartzitic crests and fields of rolling convex hills, usually associated to Ferralsols and Acrisols. Podzols are commonly found in poorly drained depressions of the central parts of the plateaus (BRAVARD; RIGHI, 1990; NASCIMENTO et al., 2004).

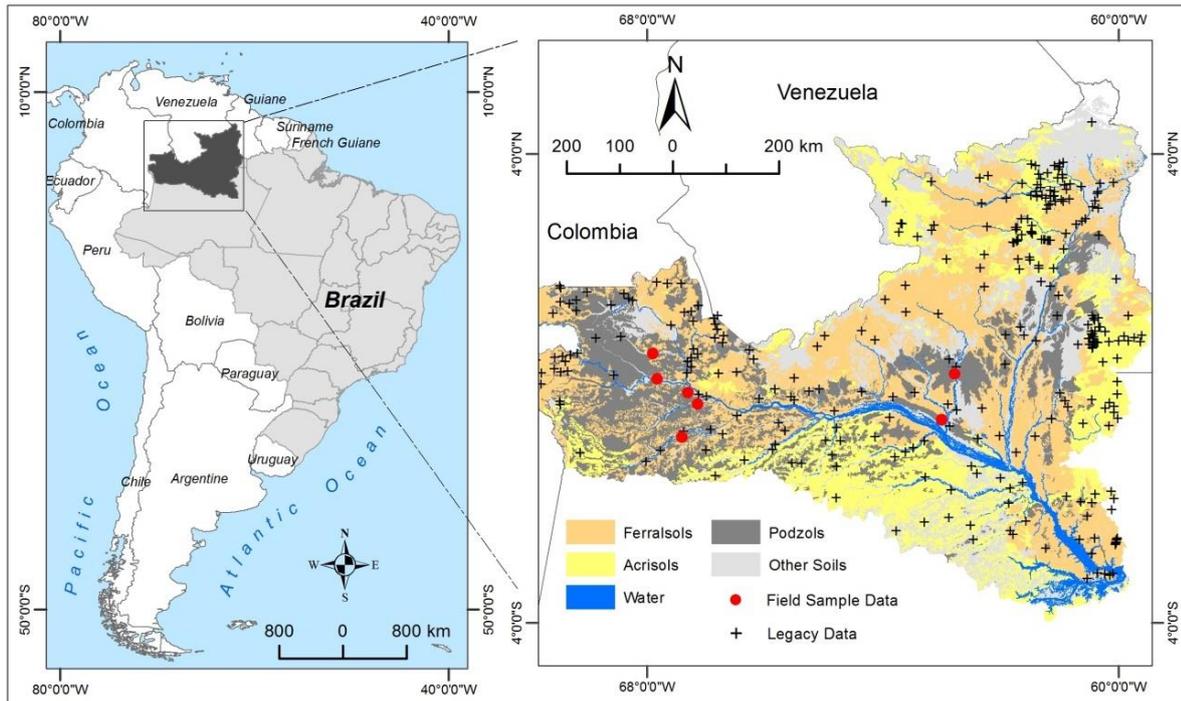


Figure 5.1 – Map of the studied area, showing the major soil orders of Rio Negro basin at the original map scale of 1:250,000 (IBGE, 2008). The legacy data (Dataset 1) was provided by IBGE (2008) and EMBRAPA (2014). Field sample data (Dataset 2) represents the samples obtained in the frame of this research in Podzol regions.

Numerous studies have attempted to describe the occurrence of Podzol-type soils in Rio Negro basin and their correlation with lithology and topography (LUCAS et al., 1984; DUBROEUCQ; VOLKOFF, 1998; NASCIMENTO et al., 2004; MONTES et al., 2007; BUENO, 2009; MONTES et al., 2011). According to these researches, the soils of Rio Negro basin are characterized by the Ferralsol/Podzol soil system, which occurs on a single landform unit and on a single parent material (DUBROEUCQ et al., 1991), with close genetic connection between Ferralsols and Podzols and without any lithogenic discontinuity (LUCAS et al., 1984; LUCAS et al., 1987).

### 5.2.2. *Field Sample Data*

The field sample data is divided in two datasets comprising the samples collected in areas of Podzols, selected in the frame of this research (Figure 5.1) and the database provided by IBGE (2008) and Embrapa (2014), which is randomly distributed in the region of Rio Negro basin (Figure 5.1). The legacy data provided by IBGE and Embrapa refers to soil samples collected by numerous researches, which have adopted different methods in order to estimate values of SOC content. Given this, we adopted the results obtained by Pereira et al.

(2015b<sup>5</sup>) at the soil profile scale. The values of SOC stock were estimated and validated according to PTF functions applied in soil profiles of Ferralsols, Acrisols, Gleysols, Arenosols, Nitisols, Planosols and Plinthosols. These soils have shown a clear exponential decay in SOC content from surface to deep soil horizons up to 3m soil depth. Therefore, the samples that had their SOC stock modelled by PTF functions are here referred as Dataset 1 and the ones comprising Podzol profiles are referred as Dataset 2.

#### 5.2.2.1. *Dataset 1*

The Dataset 1 comprises soil samples provided by IBGE (2008) and accounts to 1442 sampled soil horizons in 324 profiles randomly distributed in the region of Rio Negro basin (Figure 5.1). The soil database of IBGE (2008) was compiled between 1972 and 2008, with most of the data collected between 1972 and 1984 in the frame of the RADAMBRASIL project. The database is systematically organized by soil horizons according to the Brazilian soil classification system (EMBRAPA, 2014). The values of SOC stock were obtained by exponential depth functions. Values of soil bulk density are not available in IBGE (2008) database. Given this, the soil bulk density values were obtained by symbolic regression analysis, according to PTF model proposed by Pereira et al. (2015b<sup>5</sup>). We used soil samples with values of soil bulk density to train and to validate the PTF model, which comprises a set of 893 soil samples distributed in the region of the Amazonas state, provided by Embrapa (2014). A detailed description of the SOC stock estimation methods and the respective prediction errors are discussed by Pereira et al. (2015b<sup>5</sup>).

#### 5.2.2.2. *Dataset 2*

The SOC stock in Podzols was estimated in 18 soil profiles. The vertical distribution of SOC stock along Podzol profiles is different when compared to adjacent soils, which demands for detailed sampling of soil horizons in order to represent the real distribution of organic carbon along the soil horizons. Thus, we collected 393 soil samples, within 18 Podzols profiles. The values of SOC content were obtained by the dry combustion technique using a Shimadzu TOC-5000 apparatus. Values of soil bulk density were measured in undisturbed samples by the Kopeck ring method (BLAKE; HARTGE et al., 1986). The SOC stock was estimated at 1m and 3m soil depth. Additionally, we estimated the SOC stock in Podzol areas at 6m soil depth, taking into account the average thickness of spodic horizons.

---

<sup>5</sup> PEREIRA, O.J.R.; MONTES, C.R.; LUCAS, Y.; MELFI, A.J. Evaluation of pedotransfer equations to predict depth soil carbon stock in tropical podzols compared to other soils of Brazilian Amazon Forest. Digital Soil Morphometrics: Srings, v. 1, 2015. *No prelo*.

### 5.2.3. Ancillary data

The ancillary data refers to information indirectly related to the lateral variation of SOC stock at Rio Negro basin. The ancillary dataset is composed by georeferenced remote sensing images (USGS, 2015), digital maps at the map scale of 1:250,000 (IBGE, 2008), vegetation biomass (SAATCHI et al., 2007) and climate data (CPRM, 2015), which refers to mean annual temperature and mean annual precipitation, measured for the last three decades and spatialized in a continuous surface by ordinary kriging (CPRM, 2015). Table 5.1 summarizes the ancillary data used in the present research.

Table 5.1 - List of ancillary data used to predict the distribution of soil organic carbon stock.

Ancillary data	Variable Type	Description	Range of values	Map Scale or Resolution	Source
Soil Orders	Categorical	Soils organized by order	9 Classes	1:250000	IBGE (2008)
Geology	Categorical	Parent Material (age)	10 Classes	1:250000	IBGE (2008)
Soil Cover	Categorical	Natural Vegetation	13 Classes	1:250000	IBGE (2008)
Geomorphology	Categorical	Type of Relief Dissection	4 Classes	1:250000	IBGE (2008)
Elevation	Continuous	Land Surface Elevation SRTM* (m)	5 to 2968 m	90 x 90m	USGS (2015)
Slope Gradient	Continuous	Maximum rate of change between the cells and neighbours	0 to 90°	90 x 90m	USGS (2015)
Catchment Area	Continuous	Area of catchments	61454 to 1.09 10 <sup>8</sup> m <sup>2</sup>	90 x 90	USGS (2015)
Wetness Index	Continuous	SAGA* Wetness Index (Böhner et. al, 2006).	8.2 to 16.7	90 x 90m	USGS (2015)
Biomass	Continuous	Vegetation Biomass in Megagrams per Hectare	0 to 11 Mg ha <sup>-1</sup>	500x500m	Saatchi et al. (2007)
NDVI*	Continuous	Vegetation Index obtained from MODIS* products	-1 to 1	250x250m	USGS (2015)
Precipitation	Continuous	Annual Mean Rainfall from 1977 to 2006	800 to 3700 mm	1:2500000	CPRM (2015)

\* SRTM: Shuttle Radar Topographic Mission; SAGA: System for Automated Geoscientific Analyses; NDVI: Normalized Difference Vegetation Index, MODIS: Moderate-Resolution Imaging Spectroradiometer.

We used categorical and continuous data in order to proceed with the spatial correlation with SOC stock values (kg C m<sup>-2</sup>), as show in Table 5.1. The maps of soil classes, soil cover, geology and geomorphology are provided by the Brazilian Institute of Geography and Statistics (IBGE, 2008) in vector format (shapefile polygons). The elevation data was

derived from Shuttle Radar Topographic Mission (SRTM) images, comprising 13 scenes at 90m spatial resolution (USGS, 2015). The SRTM images were combined through mosaicking procedure and filtered to remove noises and voids, which resulted in a DEM (Digital Elevation Model) for the entire region of the Rio Negro basin. After mosaicking the SRTM scenes, the watershed boundaries were delimited from the resulting DEM using automatic hydrologic analysis tool in ArcGIS 10.3 (ESRI, 2014), thus the Rio Negro basin mask was automatically generated and used to delimit the continuous and categorical data (Table 5.1) within the limits of the Brazilian portion of the Rio Negro basin. After obtaining the Rio Negro basin limits, the ancillary data (Table 5.1) was subset to the area of the basin and converted to grid format. The resulting grids were resampled by bicubic interpolation to the nominal spatial resolution of 250-m, in order to standardize the ancillary database.

The continuous and categorical data were brought to the common projection of Mercator, Datum: WGS 1984. The spatial accuracy of the final data was verified based on the IBGE (2008) thematic maps and its associated georeferenced control points.

#### ***5.2.4. Mapping the SOC stock***

Variations in soil properties can be quantitatively modelled by correlation with continuous environmental attributes. Given this, we adopted variables described by the SCORPAN principle, which comprises climate, organisms, relief, parent material and geological time (BURROUGH, 1993; MCKENZIE; RYAN, 1999). These variables were derived from the dataset presented in Table 5.1. The incorporation of environmental variables to predict the spatial distribution of different soil properties, including SOC stock, is referred as Digital Soil Mapping (DSM) and its usage has been extensively discussed in several researches (ODEH et al., 1994; GRUNWALD, 2006; GREVE et al., 2007; MINASNY et al., 2008). Therefore, we explored DSM techniques in order to predict the spatial distribution of SOC stock in the Rio Negro basin region.

The spatial prediction of SOC stock was carried out at two soil depths based on ordinary kriging (OK) and stepwise multiple linear regression kriging (RK), according to method presented on the flowchart of Figure 5.2. The RK function is composed by a deterministic model based on stepwise linear multiple regression (SLMR) and the spatially correlated residuals of the regression (unexplained variation). The general principle of RK includes regression, and kriging of the residuals from the regression (QUINLAN, 1993), where outputs from these two steps are added to obtain the final prediction. The kriging methods were performed according to a calibration dataset extracted from the original

pointing data. Calibration and validation datasets were obtained by *Geostatistical Analysis* tool in ArcGIS 10.3 (ESRI, 2014) using the *Create Subset* function. Thus, the original data was subset into calibration (75%) and validation (25%) data. The calibration dataset was then used to generate the kriging predicted maps, according to methods described below.

#### 5.2.4.1. Ordinary Kriging

Ordinary kriging is the simplest form of kriging. In OK, the regionalized variable is assumed to be stationary. Where the variable of interest ( $Z$ ) is modelled by ordinary kriging at location  $X_i$  as:

$$Z(X_i) = m + e(X_i) \quad \text{Eq. 5.1}$$

Where  $m$  represents a regional mean and is assumed constant and unknown across the field.  $e(X_i)$  is a spatially correlated random component estimated from the variogram model. Therefore, the equation of ordinary kriging is given as:

$$\hat{Z}(X_0) = \sum_{i=1}^n \lambda_i + Z(X_i) \quad \text{Eq. 5.2}$$

Where  $\hat{Z}$  is the estimated value at point  $X_0$  and  $\lambda_i$  is the optimal weight assigned to all sample points. The weights  $\lambda_i$  assigned to the sample points sum to 1: i.e.:  $\sum_{i=1}^n \lambda_i = 1$ . Thus, ordinary kriging is assumed to be optimal because kriging equations are used to minimize the kriging variance at each point to be predicted (WEBSTER; OLIVER, 2007).

#### 5.2.4.2. Regression Kriging

The applied RK model is a hybrid method that combines either a SLMR regression model with ordinary or simple kriging of the regressed residuals (ODEH et al., 1995). Thus, RK is a mixed predictor which considers long range structure and local structure. It models the trend and its associated residuals, separately. RK is summarized as:

$$\hat{Z}(S_0) = \hat{m}(S_0) + \hat{e}(S_0) \quad \text{Eq. 5.3}$$

$$\hat{Z}(S_0) = \sum_{k=0}^p \hat{\beta}_k q_k(S_0) + \sum_{i=1}^n \lambda_i(S_0) e(S_i) \quad \text{Eq. 5.4}$$

Where,  $\hat{Z}(S_0)$  is the value of target variable at unvisited location  $S_0$ .  $\hat{m}(S_0)$  is the drift value or fitted deterministic part (trend) at location  $S_0$  and  $\hat{e}(S_0)$  is the value of the residuals taken according to SLMR resulting models at location  $S_0$ .  $\hat{\beta}_k$  are the estimated coefficient of the deterministic part,  $\lambda_i$  are the kriging weights determined by the spatial dependence structure of the residuals,  $e(S_i)$  is the residual at location  $S_i$ ,  $q_k$  is the predictor variable at location  $S_0$  and  $p$  is the number of predictors.

The SLMR resulting equation defines the independent variables (ancillary database) used to predict the dependent variable (SOC stock). In the kriging process, this model is described by the first term of the Equation 5.4, which defines the regressed surface. Each independent variable leads to a coefficient value describing its predictive strength and whether it has a positive or negative relationship. Therefore, the regression predicted surface is defined as the estimation of the target variable values at pixel's location based on the ancillary raster dataset. The SMRL models were obtained at 1m and 3m soil depth in STATISTICA software (STATSOFT Inc., 2006) and written in Python code, compatible with ArcGIS 10.3 *Spatial Analyst* extension, to acquire the regressed surface in the abovementioned soil depths.

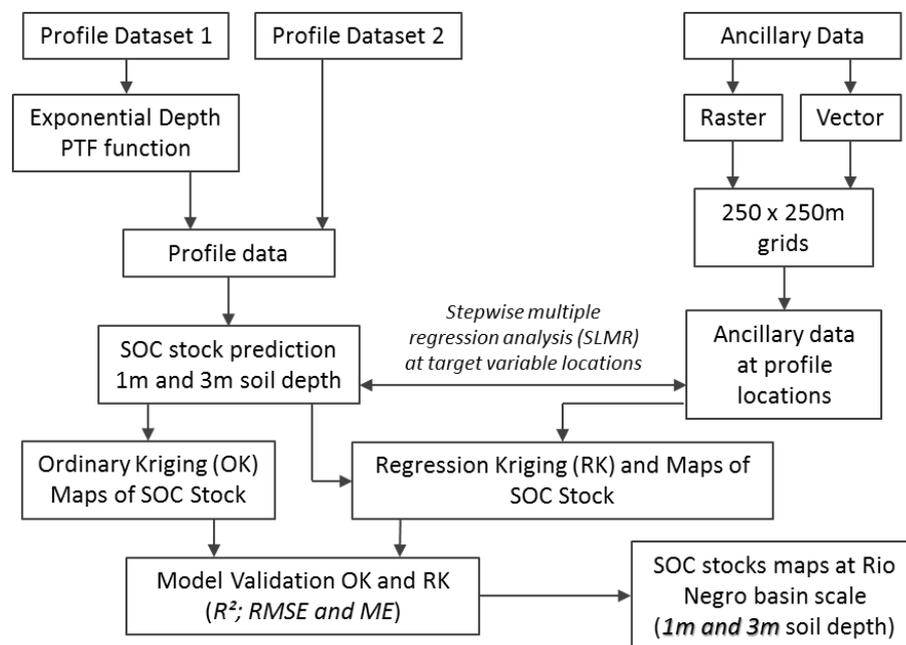


Figure 5.2 – Flowchart of the overall SOC stock prediction method.

The RK was carried out in two steps. The first one comprises the SLMR analysis, which was applied to generate the linear correlated model to predict SOC stock values based on ancillary data. According to Equation 5.4, we can run the RK by predicting the values of

the target variable ( $S_0$ ) at unvisited locations based on the residuals of the ancillary variables ( $\hat{e}(S_0)$ ). Therefore, the second step evolves the derivation of the regression residuals as well as the analysis of the residuals' variance for spatial autocorrelation and fitting of the variogram model. The final step evolves the visualization and validation of the results using control points. The results of OK and RK were compared based on the validation dataset extracted from the original target variable.

### 5.2.5. Evaluation of Predicted SOC stock maps

The SOC stock in the soil profiles was predicted at 1m and 3m soil depths by exponential depth functions in Ferralsols, Acrisols, Gleysols, Arenosols, Nitisols, Planosols and Plinthosols and by measured values in Podzol profiles. The SOC stock maps were generated by kriging interpolation and the estimated values were validated according to the validation dataset extracted from the original data. The model performance in predicting SOC stock maps was evaluated on 25% of the pointing data. The following three indices were calculated:

$$R^2 = \frac{\sum_{i=1}^n (cali - \overline{vali})^2}{\sum_{i=1}^n (vali_i - \overline{vali})^2} \quad \text{Eq. 5.5}$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (vali_i - cali_i)^2 \quad \text{Eq. 5.6}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (vali_i - cali_i)^2} \quad \text{Eq. 5.7}$$

Where *vali* and *cali* are observed and predicted (validation and calibration) SOC stock values from  $n$  number of observations at  $i_{th}$  locations, *MSE* is the mean squared error, and *RMSE* is the root mean squared error.

## 5.3. Results

### 5.3.1. Descriptive Statistics

The SOC stock varies considerably at the two soil depths (1m and 3m). The values vary from 1.72 to 93.32 kg C m<sup>-2</sup> at 1m soil depth. We observed the same pattern at 3m soil depth, with maximum and minimum values ranging from 2.56 to 98.13 kg C m<sup>-2</sup>. However it is important to highlight that the distribution of SOC stock at 1m soil depth is clearly skewed

towards lower values when compared to 3m soil depth, that has a distribution closer to normal (Figure 5.3a and 5.3b).

The coefficient of kurtosis, which measures the peakedness of a distribution, was high in both measured soil depths (32 kg C m<sup>-2</sup> and 12 kg C m<sup>-2</sup>, at 1m and 3m soil depths, respectively), indicating that the distribution of the SOC stock values has more peaked values, when compared to a normal distribution. Given this, we employed a lognormal transformation, which made the data approximately normal, with means and medians being about the same, and reduced the skewness from original values (Figure 5.3c and 5.3d). After a lognormal transformation, the SOC stock values for both soil depths showed a distribution closer to normal, which is confirmed by the QQ (Quantile-Quantile) plots in Figure 5.4. The QQplot provides another option to verify the normality of the SOC stock data.

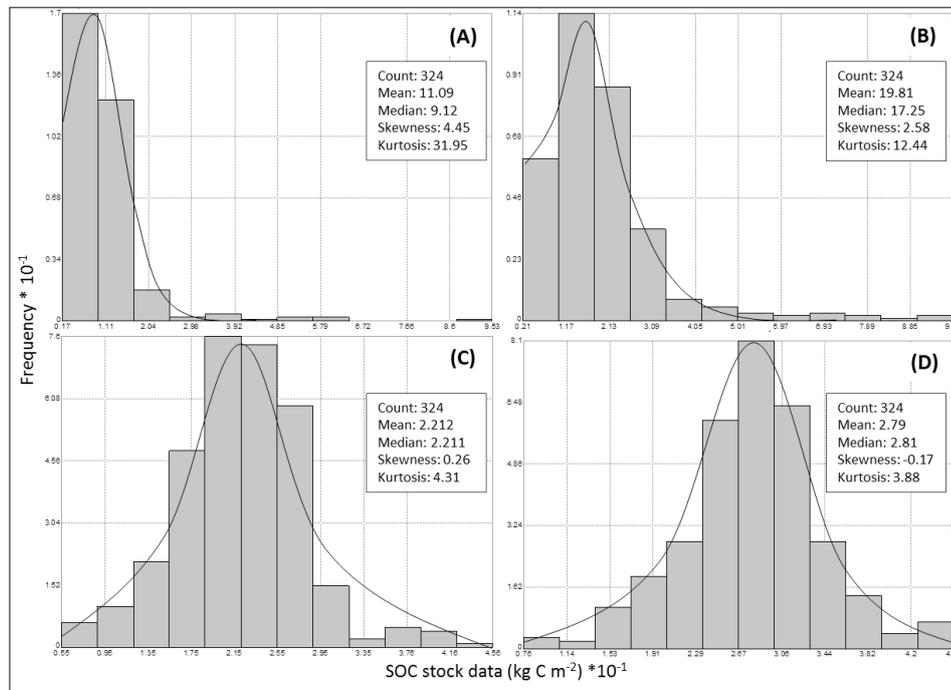


Figure 5.3 – Histograms of soil organic carbon (SOC) stock for Rio Negro basin: (a) measured SOC data at 1m soil depth; (b) measured SOC data at 3m soil depth; (c) logarithmically transformed (LnSOC) SOC data at 1m soil depth; (d) logarithmically transformed (LnSOC) SOC data at 3m soil depth.

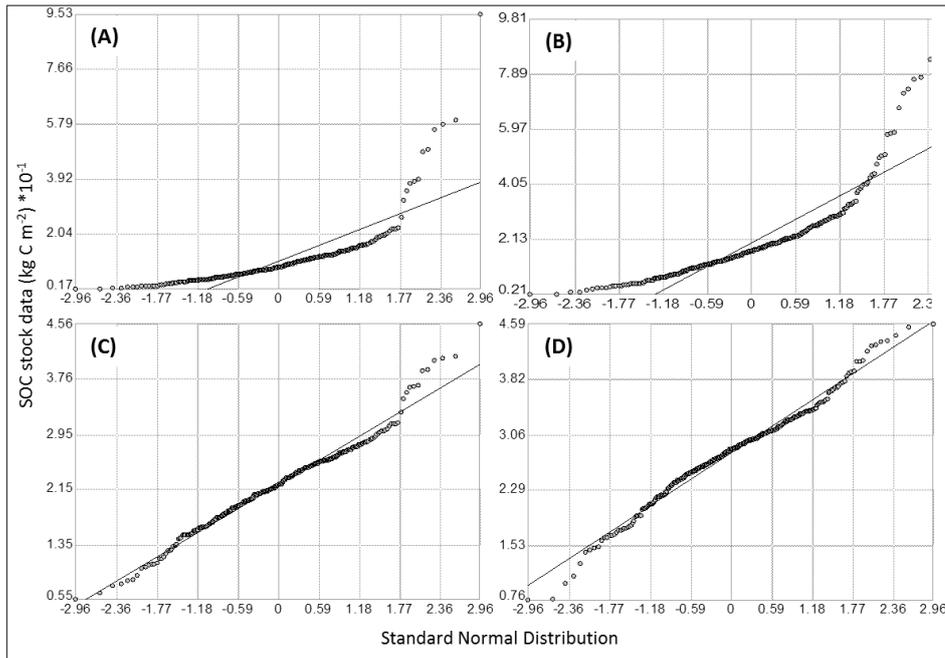


Figure 5.4 – Normal QQPlot of soil organic carbon (SOC) stock for Rio Negro basin: (a) measured SOC data at 1m soil depth; (b) measured SOC data at 3m soil depth; (c) logarithmically transformed (LnSOC) SOC data at 1m soil depth; (d) logarithmically transformed (LnSOC) SOC data at 3m soil depth.

In both soil depths (Figures 5.4a and 5.4b) we observed a significant deviation from the standard distribution, associated to higher SOC stock values. These values are related to Podzol profiles where the SOC stock increases significantly in soil depths ranging from 0.8 to more than 3m (MONTES et al., 2011). The increasing in SOC stock occurs abruptly from areas of ferralic soils (Ferralsols and Acrisols) to Podzol, where the presence of Bh horizons has been attested. Given this, the Dataset 1 has similar SOC stock values at 1 and 3m soil depth. However, in Podzol (Dataset 2) the increasing in SOC stock is clearly higher when compared to other soils, from 1 to 3m soil depth.

The segmentation of SOC stock values according to soil orders in Rio Negro basin, leads to a high standard deviation due to the extension of the studied area with a great variety of natural environments. A proper way to estimate the SOC stock in Amazon soils would be thought the application of geostatistical methods, allowing the spatialization of the SOC stock according to the different natural environments observed in this region.

### 5.3.2. Mapping the SOC stock in Rio Negro basin

#### 5.3.2.1. Ordinary Kriging

We tested a series of variogram fitting models in order to select the one that best fits to the experimental model. The variogram adjustment was verified according to the cross validation of fitted against measured values. Therefore, we considered the best  $r^2$ , MSE and RMSE as error predictors, as well as the spatial autocorrelation, to select the proper variogram (Table 5.2 and Figure 5.5). The exponential model had the best adjustment to the measured values in both soil depths, as shown in Figure 5.5 and Table 5.2. A significant difference was observed on the experimental semivariograms of the two soil depths. The most remarkable difference between the two models (Figure 5.5) is related to a higher spatial autocorrelation for SOC stock at 1m soil depth.

If we take into account an isotropic model, the spatial autocorrelation between *lag* pairs is higher at 1m soil depth. The distance at which the variogram reached its sill was 584km at 1m soil depth and 55km at 3m soil depth, according to an exponential model. The nugget ( $C_0$ ) and sill ( $C_0 + C$ ) were 0.10 and 0.17 at 1m soil depth. However, at 3m soil depth, the nugget and sill were 0.12 and 0.18, respectively, indicating a lower spatial autocorrelation (Figure 5.5). We found that the nugget, which indicates the small-scale variation of a regionalized variable, represented 57% of the sill at 1m soil depth and 66% of the sill at 3m soil depth. Given this, the spatial autocorrelation at 1m soil depth is higher when compared to 3m soil depth (Figure 5.5). We believe that the areas of Podzols had a higher influence on the modelled semivariograms at 3m soil depth. Accordingly, the abrupt lateral change of SOC content in the spodic horizons of Podzols might causes the loss of spatial autocorrelation, due to the lateral increasing in SOC stock from ferralitic soils (mostly Acrisols and Ferrasols) to Podzols, which is not evident at 1m soil depth.

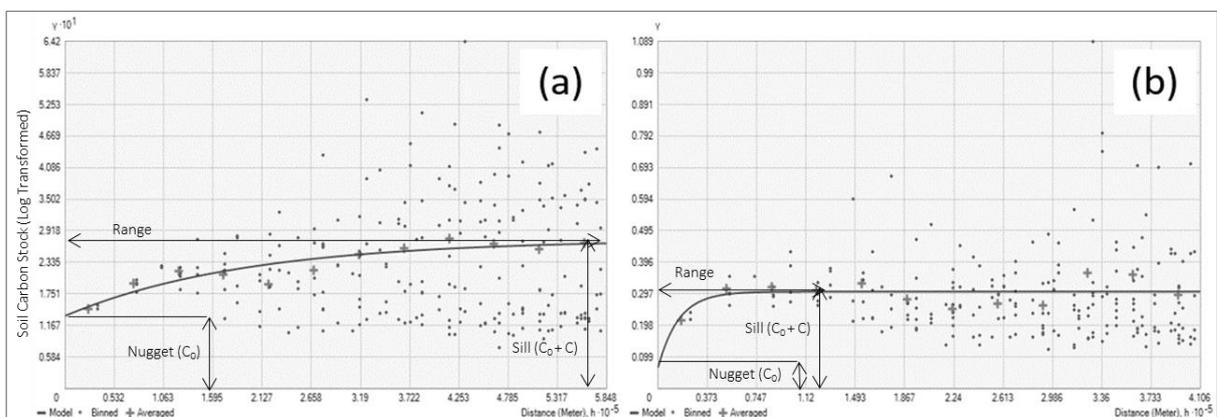


Figure 5.5 – Experimental and modelled variograms of soil carbon stock at 1m (a) and 3m (b) soil depth for OK.

In addition, the error predictors shown in Table 5.2 were used to assess the performance of the modelled variograms in the two evaluated soil depths. The Cross-validation was evaluated between measured and predicted values, according to OK interpolation. The MSE and RMSE values are smaller at 3m soil depth, when compared to 1m soil depth. The correlation coefficient also indicates a better prediction at 3m soil depth (Table 5.2). The OK predicted surfaces, explained 66% and 84% of the measured spatial variability of SOC stock in Rio Negro basin at 1m and 3m soil depth, respectively. However, it is necessary to highlight the low spatial autocorrelation for the OK procedure (Figure 5.5) at 3m soil depth, which can increase the uncertainties of the predicted values at unvisited locations. Thus, a denser validation dataset is necessary in order to evaluate the quality of the predicting maps, especially in areas of Podzols where there is a low availability of soil samples.

Table 5.2 – Prediction error parameters for Ordinary Kriging. The number of samples refers to the training dataset (85%). MSE and RMSE values are expressed in kg C m<sup>-2</sup>.

Prediction Errors (exponential model)	1m soil depth	3m soil depth
Samples	275	275
R <sup>2</sup>	0.66	0.84
Mean Squared Error (MSE)	6.22	2.36
Root-Mean-Square-Error (RMSE)	2.49	1.96

It is important to highlight that a more realistic representation of the spatial distribution of SOC stock in Rio Negro basin depends on the availability of measured SOC values at soil horizons down to 3m soil depth and in a denser pointing cloud, to improve the spatial autocorrelation. The continental extension of the Amazon basin and the inaccessibility of the sampling areas makes it difficult the acquisition of soil samples in a denser pointing cloud. Given this, the legacy data (IBGE, 2008) presented in Dataset 1 represents the most refined soil database currently available in this region. Such problem is solved by evaluating the correlation between SOC stock values and ancillary data at sampled locations, through the application of RK methods.

#### 5.3.2.2. *Regression Kriging (RK)*

Usually RK is performed using geostatistical packages coupled together with the statistical software *R* (PEBESMA, 2004) by the *gstat* extension, which allows the multiple regression analysis, the derivation of the regressed surface and the fitting of the variogram of the residuals. However, we decided to use ArcGIS, since it makes it possible the dynamic optimized selection of variograms and it provides a wide range of options to process raster and vector data. Moreover, programming in Python language is easier and intuitive, which

makes it easy the implantation of the SMLR model to generate regressed surfaces. Two regression models were obtained according to the evaluated soil depths. Different ancillary variables were considered at the two soil depths, taking into account *P* values for variable inclusion and removal of 0.05 and 0.1, respectively. The SMLR analysis returned models with moderate to good prediction power, according to a forward automatic selection (Tables 5.3 and 5.4).

Table 5.3 – Summary of the SMRL variables selection at 1m soil depth (Regression Kriging).

No. of variables	Variables	MSE	R <sup>2</sup>	Akaike's AIC
1	Vegetation	0,17	0,35	-502,07
2	Geology/Vegetation	0,15	0,47	-539,96
3	Geology/Geomorphology/Vegetation	0,14	0,49	-543,39
4	Geology/Geomorphology/Soil/Vegetation	0,14	0,50	-539,47
5	Rainfall / Geology/Geomorphology/Soil/Vegetation	0,14	0,51	-539,40
<b>6</b>	<b>Catch. Area/Rainfall /Geology/Geomorphology/Soil/Vegetation*</b>	<b>0,14</b>	<b>0,52</b>	<b>-538,62</b>
7	Catch. Area /NDVI/Rainfall/Geology/Geomorphology/Soil/Vegetation	0,14	0,51	-537,19

\* Selected model at 1m soil depth.

Table 5.4 – Summary of the SMRL variables selection at 3m soil depth (RK).

No. of variables	Variables	MSE	R <sup>2</sup>	Akaike's AIC
1	Vegetation	0,25	0,27	-402,25
2	Geology/Vegetation	0,22	0,38	-428,37
3	Geology/Soil/Vegetation	0,20	0,45	-449,27
4	Geology/Geomorphology/Soil/Vegetation	0,19	0,47	-457,88
<b>5</b>	<b>Biomass/Geology/Geomorphology/Soil/Vegetation*</b>	<b>0,19</b>	<b>0,48</b>	<b>-457,78</b>
6	Biomass/Rainfall/Geology/Geomorphology/Soil/Vegetation	0,19	0,48	-456,46

\* Selected model at 3m soil depth.

The MSE, adjusted R<sup>2</sup> and Akaike's (AIC) criterion, indicate the regressed model that best explains the target variable. We selected the models 6 and 5 (Tables 5.3 and 5.4) to generate the regressed surfaces, at 1m and 3m soil depth, respectively. The relative usage of each variable (Figure 5.6 and Figure 5.7) on the general regression models was useful on verifying the environmental conditions related to the lateral variation of SOC stock in Rio Negro basin. The standardized usage coefficient helps on comparing the relative weights of the continuous and categorical variables to build the model. The higher the absolute value of the coefficient, the more important the weight of the corresponding variable. As we can see in Figures 5.6 and 5.7, the usage of continuous variables was low in both soil depths with a higher usage related to Rainfall data at 1m soil depth. Catchment area and biomass were used in the models, however with lower weights. Therefore, the variance of the target variable was explained, mostly, by the categorical data (Figures 5.6 and 5.7).

The categorical information regarding vegetation classes and soil orders had the highest usage in the models (Figures 5.6 and 5.7). The areas of high dense forests and the geological units representing recent parent materials (Pliocene and Holocene epochs) also had an important usage in the regression models. The usage of soil orders was higher at 3m soil depth with high standard coefficients related to areas of Ferralsols, Acrisols and Podzols, which sums 74% of the total area of the Rio Negro basin.

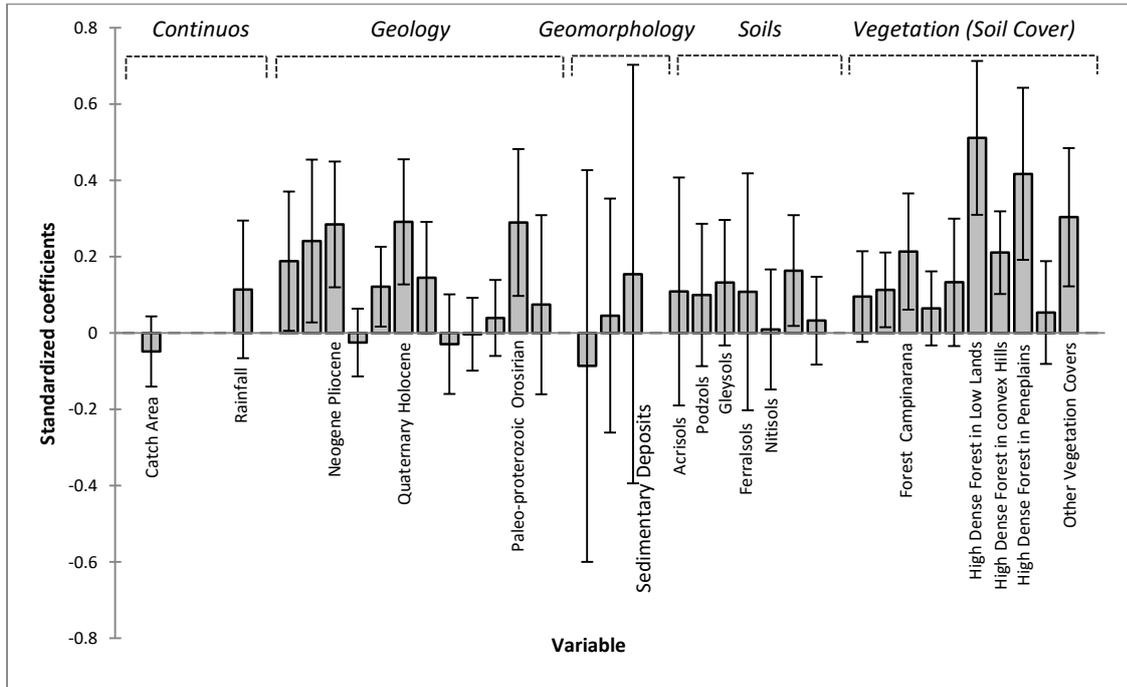


Figure 5.6 – Standardized coefficients chart at 1m soil depth, highlighting the most significant classes.

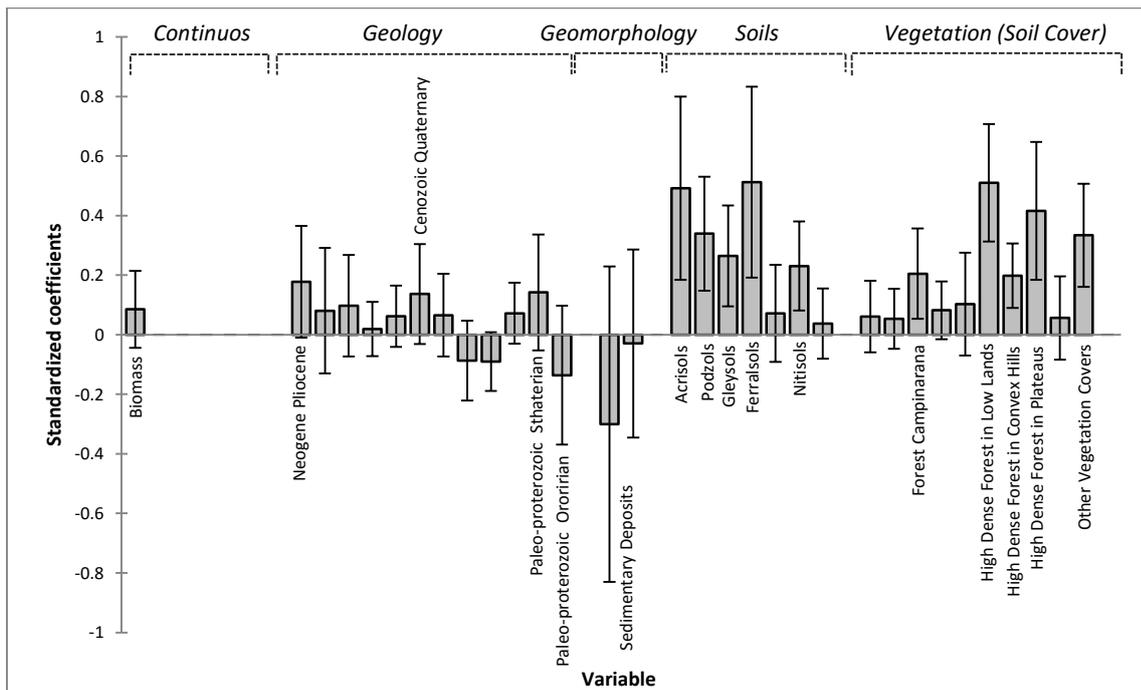


Figure 5.7 – Standardized coefficients chart at 3m soil depth, highlighting the most significant classes.

The distribution of vegetation classes had the highest standardized coefficients at 1m soil depth, whereas soils (Acrisols and Ferralsols) had a strong influence on the resulting regression model at 3m soil depth. Two aspects are related to the high influence of vegetation classes' distribution on the lateral variation of SOC stock in Rio Negro basin. According to Anderson et al. (1981) high dense forest occurs in areas of ferralitic soils (Acrisols and Ferralsols), while sclerophyllous vegetation physiognomies (Campinarana formations) are associated to hydromorphic soils, mostly in Podzol areas. Besides, the soil mapping method adopted by IBGE (2008), takes into account the indirect correlation between soils and vegetation cover, which explains the strong influence of the vegetation classes on the lateral variation of SOC stock. Moreover, it is important to highlight the weight of the geology map on the resulting models at 1m soil depth, where we observed high coefficients related to recent parent material (Neogene and Quaternary deposits). Therefore, the spatial distribution of environmental variables in the categorical images (IBGE, 2008) was decisive to build the regressed surface at 1m and 3m soil depth.

Regarding the percentage of use in the resulting models, the areas of Acrisols, Ferralsols (soil map) and sedimentary deposits (geomorphology map) had a frequency usage of above 35% in all regressed models. On the other hand, the continuous data had a usage below 5% due to their low correlation with SOC stock values. The highest correlation among the continuous data was found between rainfall and SOC stock with an adjusted  $r^2$  of 0.26. Therefore, it is not possible to make any assumption about the lateral distribution of SOC stock based only on the continuous dataset. Given this, the regressed model in the two evaluated soil depths was mostly based in categorical data provided by IBGE (2008) maps.

At 1m soil depth the experimental variogram of the residuals of the regression, had the best adjustment to an exponential model (ordinary kriging of correlated residuals), with a nugget ( $C_0$ ) of 0.80 and a range of about 167 km (Figure 5.8), which might indicate a higher spatial autocorrelation at longer distances when compared to the OK kriging method (Figure 5.5a). At 3m soil depth the variogram was adjusted to an exponential model, with a nugget ( $C_0$ ) of 0.101 and a range of 84.83km. Thus, it is important to highlight a moderate dependence structure at the two evaluated soil depths, with nugget representing 47% and 48% of the sill ( $C_0 + C$ ) at 1m and 3m soil depth, respectively. The nugget/sill ratio of the residuals of the regressions reveals that around 48% of the SOC stock variability consists

of unexplainable or random variations, for the two evaluated soil depths. Accordingly, new efforts are necessary to refine the current number of soil samples in the region of Rio Negro basin, which would increase the quality of the kriging procedure. Nevertheless, the spatial autocorrelation was higher at the two evaluated soil depths for RK when compared to OK.

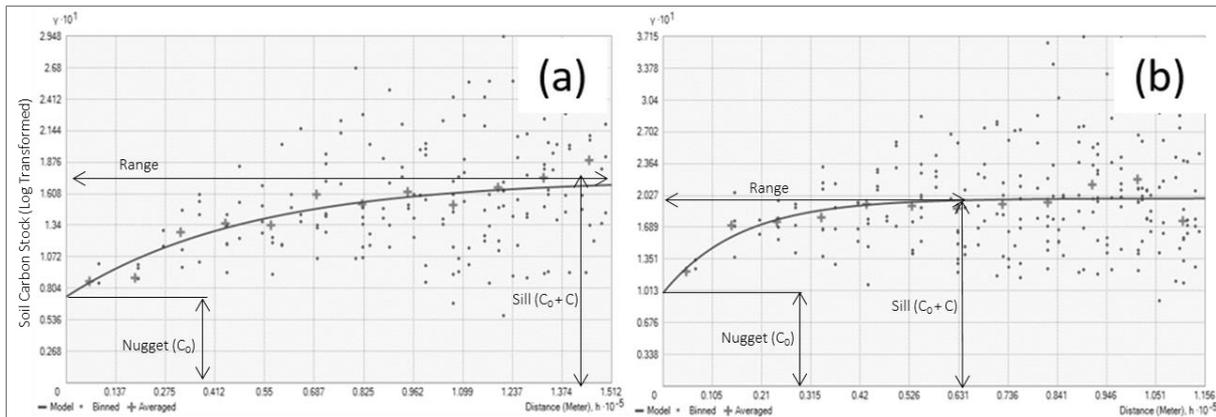


Figure 5.8 – Experimental and modelled variograms of soil carbon stock at 1m (a) and 3m (b) soil depth for RK.

The predicted values of the regressed surfaces without kriging were compared with the values obtained by RK (Figure 5.9). As we can see in Figure 5.9, the regression kriging resulted in good prediction capability at 1m soil depth and a moderate capability at 3m soil depth. The resulting RK predicted map at 1m soil depth, explained 78% of the variability of SOC stock in Rio Negro basin (Figure 5.9c), whereas it explained 54% at 3m soil depth, with a higher deviation related to high values of SOC stock (Figure 5.9d).

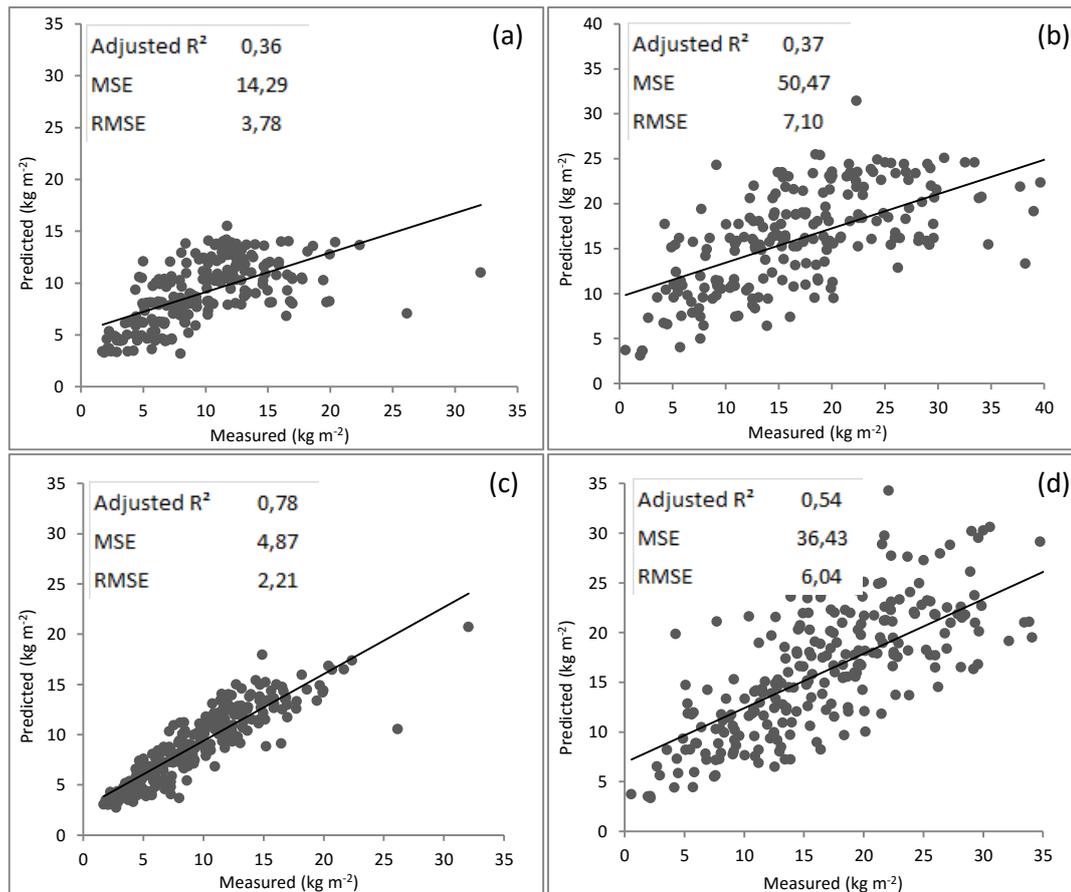


Figure 5.9 – Measured against predicted values of SOC stock: (a) Regressed surface at 1m soil depth; (b) regressed surface at 3m soil depth; (c) RK at 1m soil depth; (d) RK at 3m soil depth. The values were converted back to SOC stock in  $\text{kg C m}^{-2}$  (exponential of Log-SOC).

The RK predicted map had a better performance at 1m soil depth (MSE: 4.86; RMSE: 2.21) when compared to OK map (MSE: 6.22; RMSE: 2.49). However, the difference between OK and RK predicted maps increases at 3m soil depth. The MSE and RMSE values for OK were 2.36 and 1.96, respectively, and increased to 36.43 and 6.04 for RK predicted map. The low correlation between environmental attributes and SOC stock at 3m soil depth, as well as the abrupt increasing in SOC content, are pointed out as the major reason for the moderated predictability capacity of the RK model. The lateral increasing in SOC stock from ferralitic soils (Acrisols and Ferralsols) to Podzols is the major reason for such behaviour.

The RK predicted SOC stock maps at 1m and 3m soil depths, have shown a significant increase in SOC stock in the region of hydromorphic soils (Figures 5.10 and 5.11). These areas are located at east and northwest of the Rio Negro basin and occur associated to Podzols and Gleysols (Figure 5.1). The RK maps have shown a similar pattern on the lateral distribution of SOC stock at the two evaluated soil depths. However the SOC stock at 3m soil depth is about twice the one stored at 1m, with averaging values ranging from

15 to 19 kg C m<sup>-2</sup> in areas of ferralitic soils and 21 to 60 kg C m<sup>-2</sup> in the region of hydromorphic soils (soil orders are shown in Figure 5.1). The areas of hydromorphic soils of Rio Negro basin have a high capacity of storing carbon in superficial organic horizons (A/O horizons of Gleysols and Podzols) and in the spodic horizon of Podzols, which explains the high amount of carbon stored in hydromorphic soils at the two evaluated soil depths.

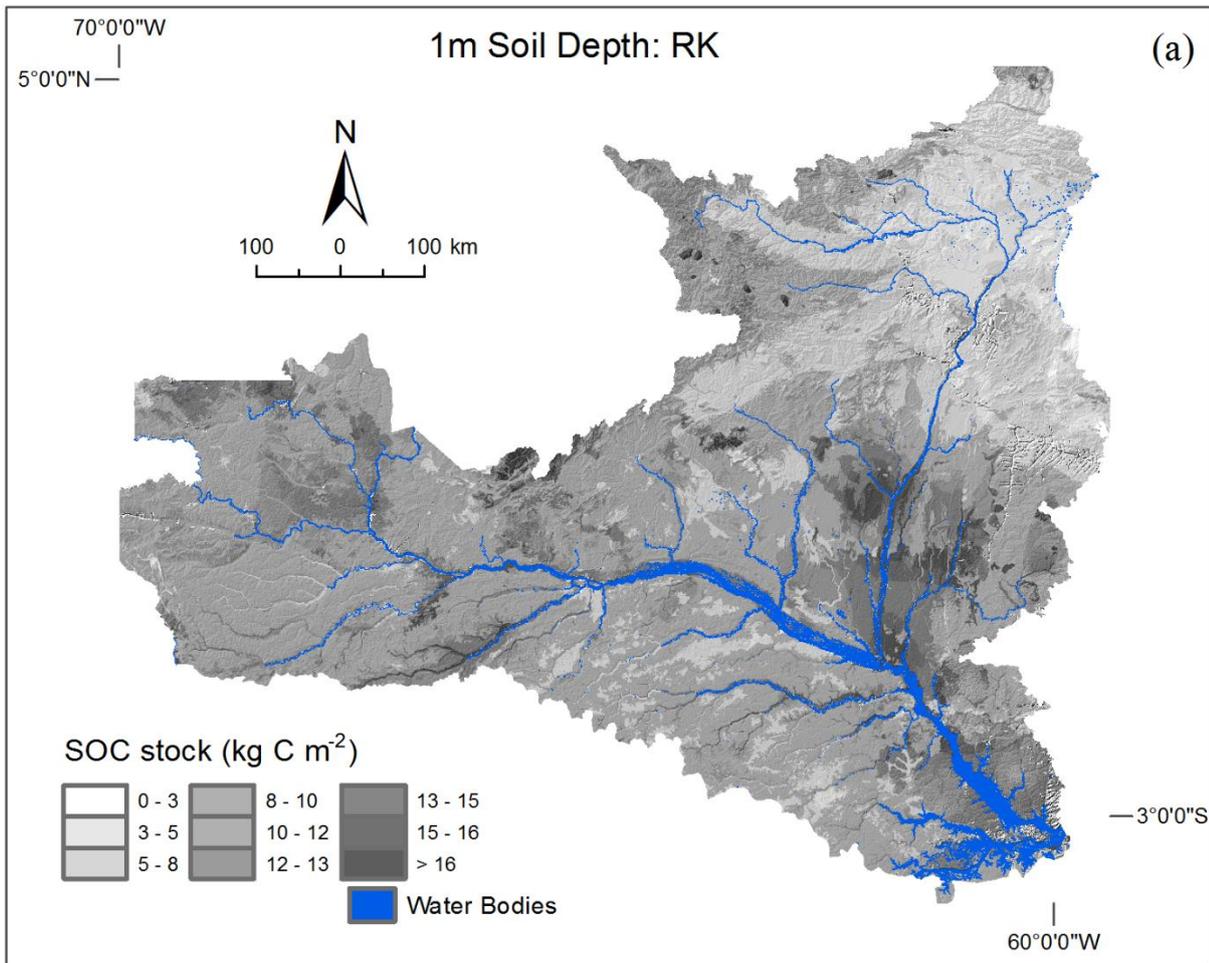


Figure 5.10 – Predicted SOC stock map according to RK procedure at 1m soil depth.

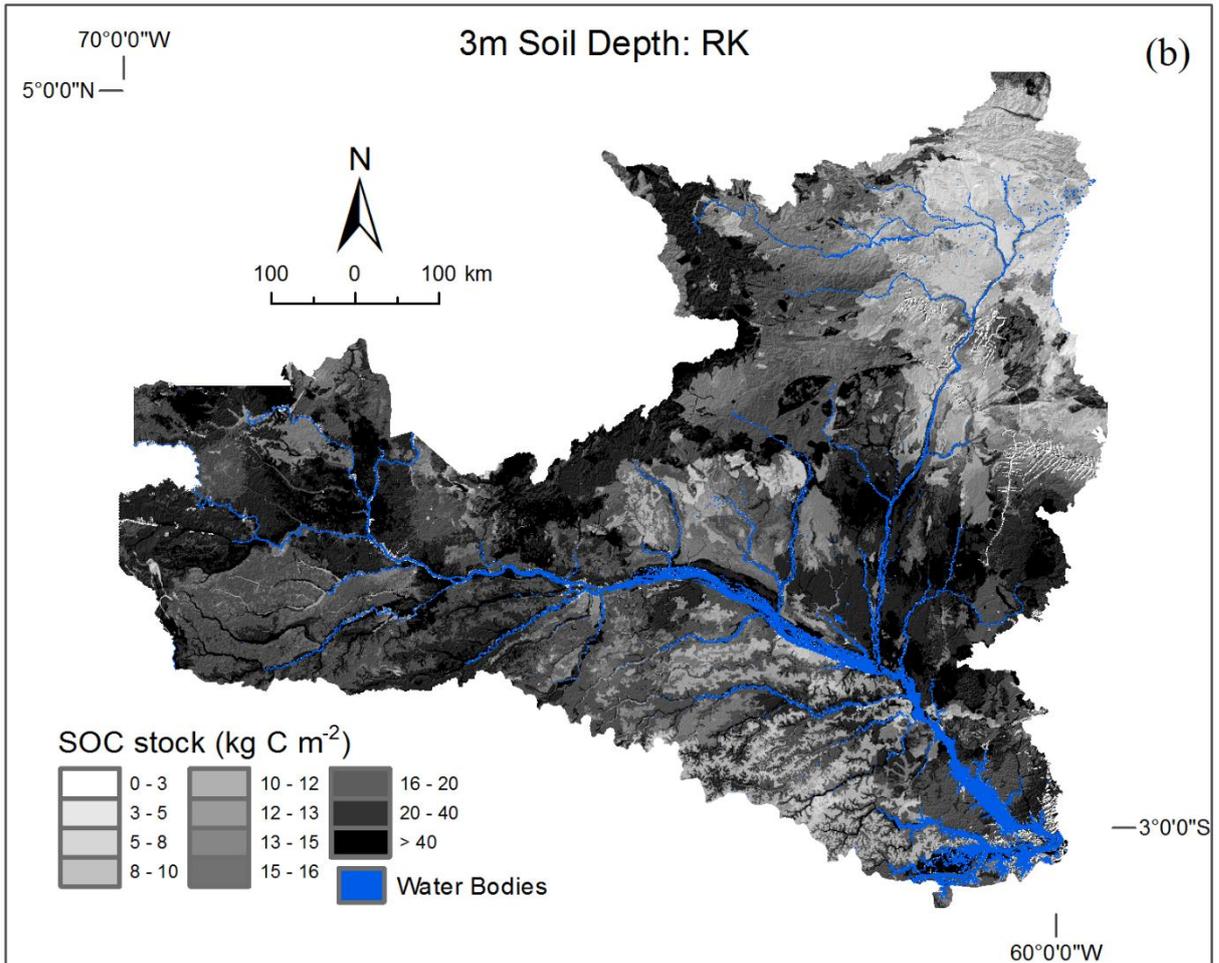


Figure 5.11 – Predicted SOC stock map according to RK procedure at 3m soil depth.

### 5.3.2.3. Kriging Models Validation and SOC Stock in Rio Negro Basin

The best prediction was found at 1m soil depth for the RK method with an adjusted  $R^2$  of 0.85, according to validation dataset (Table 5.5). The obtained  $R^2$  at the same soil depth for the OK method was 0.66. Nevertheless, the validation dataset is more correlated with values obtained by OK at 3m soil depth, when compared to RK, with respective  $R^2$  of 0.76 and 0.61 (Table 5.5). The close correlation between RK and OK SOC stock maps and the better spatial autocorrelation for the RK models, justify the use of RK predicted maps to quantify the SOC stock in Rio Negro basin due to a higher level of detail that can be found in RK predicted maps when compared to OK maps.

Table 5.5 – Model performance to predict soil carbon stock ( $\text{kg C m}^{-2}$ ) based on validation dataset.

Parameters	1m Soil Depth		3m Soil Depth	
	OK	RK	OK	RK
Adjusted $R^2$	0.66	0.85	0.76	0.61
MSE	6.23	3.86	25.36	32.04
RMSE	2.50	1.96	5.04	5.66

The average amount of SOC stored in Rio Negro basin soils according to RK predicted maps was  $9.4 \pm 4 \text{ kg C m}^{-2}$  at 1m soil depth and  $17.3 \pm 7.8 \text{ kg C m}^{-2}$  at 3m soil depth. If we consider the SOC stock divided by soil order, the areas of Podzols have the highest storage capacity with a relative stock of about  $12.2 \pm 5.1 \text{ kg C m}^{-2}$  at 1m soil depth and  $24.9 \pm 10.8 \text{ kg C m}^{-2}$  at 3m soil depth, whereas in Ferralsols, the SOC stock at 1m and 3m soil depths are  $11.3 \pm 5.3 \text{ kg C m}^{-2}$  and  $18.2 \pm 7.7 \text{ kg C m}^{-2}$ , respectively. In Acrisols the SOC stock is  $10.5 \pm 4.8 \text{ kg C m}^{-2}$  at 1m soil depth and  $17.2 \pm 8.3$  at 3m soil depth. Understanding the spatial variability of SOC stock at these soil units is important due to their spatial extension (74% of the soils of Rio Negro basin). However, the areas of alluvial soils (Gleysols) have also shown a high SOC storage capacity in superficial soil horizons, with an average stock of about  $15.2 \pm 6.4 \text{ kg C m}^{-2}$  at 1m soil depth and  $21.2 \pm 9.8 \text{ kg C m}^{-2}$  at 3m soil depth, which highlights the importance of these soils on evaluating the carbon sink capacity of Amazon soil.

The soils of Rio Negro basin stores about 5.75 Pg of SOC at 1m soil depth and 10.12 Pg at 3m soil (twice the value found on the first soil meter: Figure 5.12), according to RK maps. However, we observed that some areas might have an increment of about  $40 \text{ kg C m}^{-2}$ , from 1m to 3m soil depth, as shown in Figure 5.12. The values obtained by OK procedure are similar being about 5.94 Pg at 1m soil and 11.38 Pg at 3m soil depth. Previous research developed in this region (MONTES et al., 2011) have found higher values of SOC stock in Podzol areas (13.6 Pg). However, these studies have considered the occurrence of SOC content in deeper soil horizons. The soil samples provided by IBGE (2008) are limited to the first soil meter, which makes it difficult the prediction of SOC stock in soil depths below 3m, due to the low availability of soil samples to calibrate the PTF functions. Nevertheless, we observed an important increment in SOC stock from 1m to 3m soil depth, even considering just the first 3m soil depth (Figure 5.12).

The most important increment in SOC stock from 1m to 3m soil depth occurs in areas of highlands at northwest (Figure 5.12) of the basin, in the areas of Podzols located at east of the Rio Negro basin and in the region of giant Podzols of the upper Rio Negro Basin (Figure 5.12). Different from the other regions, the areas of giant Podzols might have a great amount of carbon stored in depths below 3 m. In some cases, a great amount of organic carbon can be found in depths up to 11m, which highlights the importance of these soils on the estimation of SOC stock at national and continental scale.

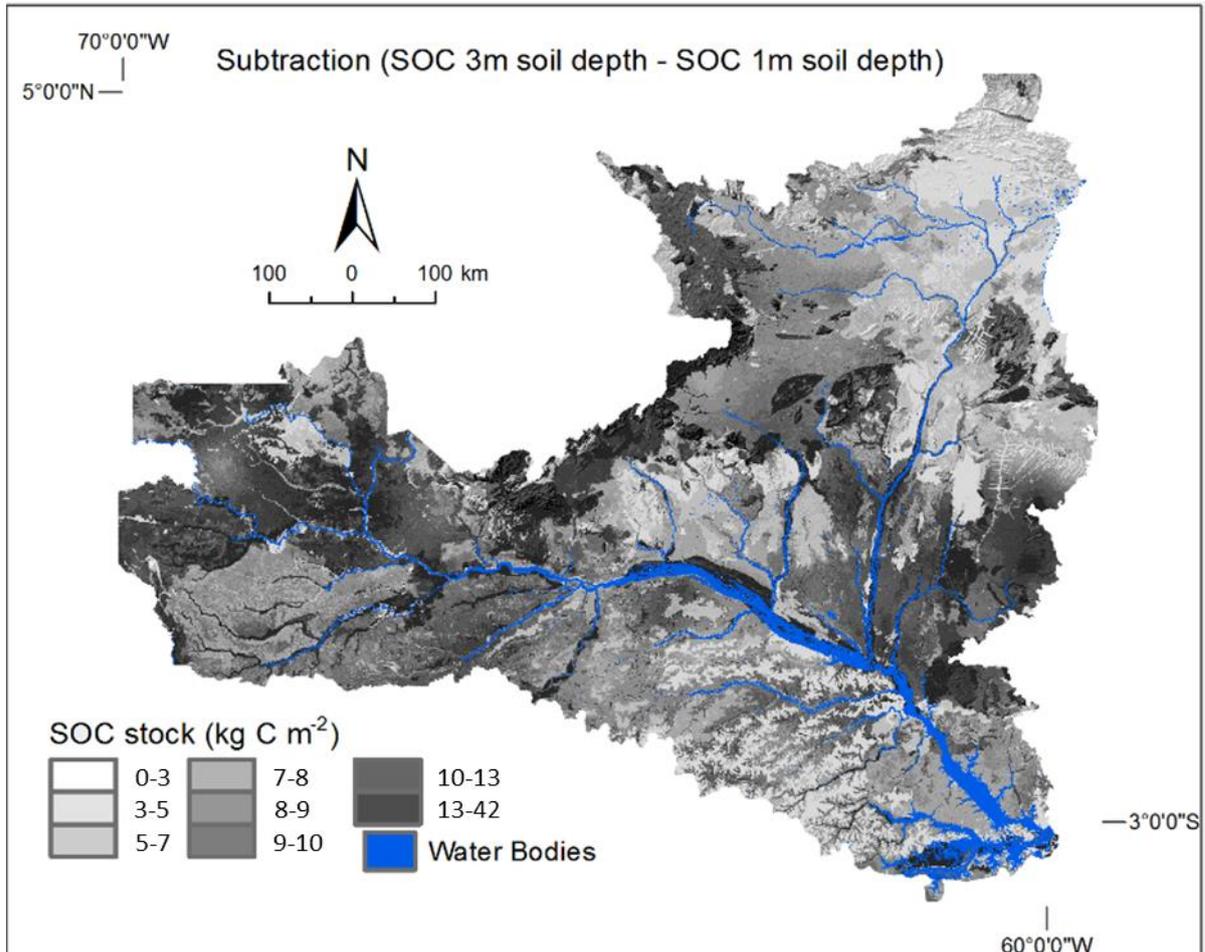


Figure 5.12 – SOC Stock map obtained by subtracting the 3m soil depth map (Figure 5.11) from the 1m soil depth map (Figure 5.10).

If we take into account an average stock of about  $80 \pm 17.2$  kg C m<sup>-2</sup> for Podzols at 6m soil depth, considering the sampled areas in the frame of this study, the Podzols of Rio Negro basin have an absolute SOC stock of 9.19 Pg. Thus, the carbon stored in Podzols at 6m soil depth is close to the one stored in the entire region of Rio Negro basin at 3m soil depth, due to the typical increasing of SOC content in deep thick spodic horizons, whereas the adjacent soils (Ferralsols and Acrisols) have shown a clear exponential decay in SOC stock from surface to deeper soil horizons.

#### 5.4. Conclusions

The OK predicted maps are well correlated to validation dataset at 3m soil depth; however, it resulted in low correlation at 1m soil depth. On the other hand, the RK maps had the opposite behavior with good correlation at 1m soil depth and moderate correlation at 3m soil depth. Despite the lower correlation at 3m soil depth, the maps obtained by RK have the advantage of representing the lateral variation of the SOC stock in Rio Negro basin with a higher level of detail, when compared to OK predicted maps. Besides, the modeled

semivariograms had a higher spatial autocorrelation according to RK procedure. Therefore, RK is pointed out as an efficient technique to map deep SOC storage in Amazon soils, by combining PTF models, regression analysis and kriging of the residuals of the regression. Nevertheless, we observed a moderate spatial autocorrelation for the application of spatially dependent interpolators (kriging), which is explained, mostly, by lack of soil samples. Thus, future researches in this region aiming to apply kriging methods to map SOC stock, depends on the availability of a denser pointing cloud (soil samples).

Few researches have explored the usage of RK techniques to map SOC stock in extensive tropical regions, mostly due to the lack of soil samples in such regions. Despite the limitations of the IBGE (2008) dataset, it provides more than 1000 soil samples randomly distributed in the region of Rio Negro basin. If we extrapolate to the entire region of the Brazilian Amazon forest, samples of 3785 soil profiles are provided by IBGE (2008) and its usage to map the SOC stock remains undone. Given this, the methods adopted in this research could be used to map deep SOC stock in the region of the Amazon forest, with a high level of detail.

With regards to Amazon Podzols, the proper evaluation of its SOC storage capacity is crucial to update current maps of SOC stock in Amazon soils. In the present research, we found that all Amazon soils have the capacity of storing an important amount of organic carbon on them superficial and sub-superficial, soil horizons. However, we observed a clear decay in SOC content from superficial to deep soil horizons in all soils, excluding Podzols. In Podzols, the soil horizons below 1m soil depth have an increase in SOC stock when compared to the one stored above 1m soil depth. Accordingly, the current measurements based on soil superficial horizons up to 0.3m, are insufficient to represent the real amount of SOC stored in Podzols. In these soils, we concluded that the SOC stock should be measured at soil depths varying from 1 to more than 3m. Moreover, new studies are necessary to evaluate the sensibility of those stocks to mineralization, taking into account different future scenarios of global climate change.

## References

- ANDERSON, B.A. White-sand vegetation of Brazilian Amazonia. **Biotropica**, Washington, DC, v. 13, n. 3, p. 199-210, 1981.
- BATJES, N.H. Total carbon and nitrogen in the soils of the world. **European Journal of Soil Science**, Oxford, v. 47, p. 151-163, 1996.
- BATJES, N.H.; SOMBROEK, W.G. Possibilities for carbon sequestration in tropical and subtropical soils. **Global Change Biology**, Oxford, v. 3, p. 161-173, 1997.

BERNOUX, M.; ARROUAYS, D.; CERRI, C.; VOLKOFF, B.; JOLIVET, C. Bulk densities of Brazilian Amazon soils related to other soil properties. **Soil Science Society of America Journal**, Madison, v. 162, p. 743–749, 1998.

BERNOUX, M.; CARVALHO, M.D.S.; VOLKOFF, B.; CERRI, C.C. Brazil's soil carbon stocks. **Soil Science Society of America Journal**, Madison, v. 66, p. 888–896, 2002.

BLAKE, G.R.; HARTGE, K.H. Bulk Density. In: KLUTE, A. (Ed.). **Methods of Soil Analysis**. Part 1. Madison: SSSA, 1986. p. 363-376.

BRASIL. Ministério das Minas e Energia. **Projeto RADAMBRASIL. Folha SA.19-Içá-AM: Geomorfologia**. Rio de Janeiro, 1977. p. 125-180. (Levantamento dos Recursos Naturais, v. 14).

BRAVARD, S.; RIGHI, D. Geochemical differences in an oxisolspodosol toposequence of Amazonia (Brazil), **Geoderma**, Amsterdam, v. 44, p. 29–42, 1989.

BUENO, G.T. **Appauvrissement et podzolisation des latérites du bassin du Rio Negro et genèse des Podzols dans le haut bassin amazonien**. 2009. 193 p. Thesis (Ph.D.) - Institut de Physique du Globe de Paris (IPGP), Paris, 2009.

BURINGH, P. Organic carbon in the soils of the world. In: WOODWELL, G. (Ed.). **The role of terrestrial vegetation in the global carbon cycle: measurement by remote sensing**. Chichester: Wiley, 1994. p. 91-109. (SCOPE, v. 23).

BURROUGH, P.A. Soil variability: a late 20<sup>th</sup> century view. **Soils and Fertilizers**, Harpenden, v 56, p. 529-562, 1993.

CERRI, C.C.; BERNOUX, M.; ARROUAYS, D.; FEIGL, B.J.; PICCOLO, M.C. Carbon stocks in soils of the Brazilian Amazon. In: LAL, R.; KIMBLE, J.; FOLLET, R.; STEWART, B.A. **Global climate change and tropical ecosystems**. Boca Raton: CRC Press, 1999. p. 33-50.

CERRI, C.E.P.; EASTER, M.; PAUSTIAN, K.; KILLIAN, K.; COLEMAN, K.; BERNOUX, M.; FALLOON, P.; POWLSON, D.S.; BATJES, N.H.; MILNE, E.; CERRI, C.C. Predicted Soil Organic Carbon Stocks and Changes in the Brazilian Amazon between 2000 and 2030. **Agriculture Ecosystems & Environment**, Amsterdam, v. 122, n. 1, p. 58-72, 2007.

COMPANHIA DE PESQUISA DE RECURSOS MINERAIS - CPRM. Brazilian Geological Service. **Geospatial Database** (shape). Rio de Janeiro, RJ. Available at: <<http://geobank.cprm.gov.br/>>. Accessed in: Jan. 10, 2015.

DUBROEUCQ, D.; VOLKOFF, B. From oxisols to spodosols and histosols: evolution of the soil mantles in the Rio Negro Basin (Amazonia).

**Catena**, Amsterdam, v. 32, p. 245–280, 1998.

DUBROEUCQ, D.; VOLKOFF, B.; FAURE, P. Les couvertures pédologiques à podzols du bassin du haut Rio Negro (Amazonie). **Etude et Gestion des Sols**, Orléans, v. 6, p. 131–153, 1999.

DUBROUECQ, D.; VOLKOFF, B.; PEDRO, G. La couverture pédologique du Bouclier du Nord de l' Amazonie (bassin du Haut Negro). Séquence évolutive des sols et son rôle dans l' aplanissement généralisé des zones tropicales perhumides. **Comptes Rendus de l'Académie des Sciences**, Paris, v. 312, n. 2, p. 663-667, 1991.

DRÖSLER, M.; VERCHOT, L.V.; FREIBAUER, A.; PAN, G.; EVANS, C.D.; BOURBONNIERE, R.A.; ALM, J.P.; PAGE, S.; AGUS, F.; HERGOUALC'H. K.; COUWENBERG, J.; JAUHAINEN, J.; SABIHAM, S.; WANG, C.; SRIVASTAVA, N.; BORGEAU-CHAVEZ, L.; HOOIJER, A.; MINKKINEN, K.; FRENCH, N.; STRAND, T.; SIRIN, A.; MICKLER, R.; TANSEY, K.; LARKIN, N. Drained Inland Organic Soil. In: HIRAISHI, T.; KRUG, T.; TANABE, K.; SRIVASTAVA, N.; BAASANSUREN, J.; FUKUDA, M.; TROXLER, T.G. (Ed.). **2013 Supplement to the 2006 IPCC guidelines for national greenhouse gas inventories**. Wetlands, IPCC, Switzerland, 2014. EASTER, M.; PAUSTIAN, K.; KILLIAN, K.; WILLIAMS, S.; FRENG, T.; AL-ADAMAT, R.; BATJES, N.H.; BERNOUX, M.; BHATTACHARYYA, T.; CERRI, C.C.; CERRI, C.E.P.; COLEMAN, K.; FALLOON, P.; FELLER, C.; GICHERU, P.; KAMONI, P.; MILNE, E.; PAL, D.K.; POWLSON, D.S.; RAWAJHIF, Z. The GEFSOC soil carbon modeling system: A tool for conducting regional-scale soil carbon inventories and assessing the impacts of land change on soil carbon. **Agriculture Ecosystems & Environment**, Amsterdam, v. 122, p. 13–25, 2007.

EMBRAPA. **Digital Soils System Information** (Brazilian Soils Database). Rio de Janeiro: Embrapa Solos, 2014. Available at: [http://www.bdsolos.cnptia.embrapa.br/consulta\\_publica.html](http://www.bdsolos.cnptia.embrapa.br/consulta_publica.html). Accessed in: Oct. 25, 2014.

ENVIRONMENTAL SYSTEMS RESEARCH INSTITUTE - ESRI. **ArcGIS Desktop**: Release 10.3. Redlands, CA, 2014.

ESWARAN, H.; VAN DEN BERG, E.; REICH, P. Organic carbon in soils of the world. **Soil Science Society of American Journal**, Madison, v. 57, p. 192–194, 1993.

GRUNWALD, S.; MCSWEENEY, K.; ROONEY, D.J.; LOWERY, B. Soil layer models created with profile cone penetrometer data. **Geoderma**, Amsterdam, v. 103, p. 181-201, 2001.

GREVE, M.H.; GREVE, M.B.; BØCHER, P.K.; BALSTROM, T.; MADSEN, H.B. Generating a Danish raster-based topsoil property map combining choropleth maps and point information. **Geografisk Tidsskrift-Danish Journal of Geography**, London, v. 107, p. 1–12, 2007.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. Geoscience Division - DGC. Coordenação de Recursos Naturais e Estudos Ambientais - CREN. **Georefered Maps of Natural Resources**: Scale 1:250:000. Digital Format: shp. Rio de Janeiro, 2008. Available at: [ftp://geoftp.ibge.gov.br/mapas/banco\\_dados\\_georeferenciado\\_recursos\\_naturais](ftp://geoftp.ibge.gov.br/mapas/banco_dados_georeferenciado_recursos_naturais). Accessed in: Aug. 15, 2014.

KIMBLE, J.M.; HEATH, L.S.; BIRDSEY, R.; LAL, R. **The Potential of US. Forest Soils to Sequester Carbon and Mitigate the Greenhouse Effect**. Boca Raton: CRC Press, 1990. 429 p.

LUCAS, Y.; CHAUVEL, A.; BOULET, R.; RANZANI, G.; SCATOLINI, F. Transição latossolos-podzols sobre a formação Barreiras na região de Manaus, Amazônia. **Revista Brasileira de Ciências do Solo**, Viçosa, v. 8, p. 325-335, 1984.

LUCAS, Y.; BOULET, R.; CHAUVEL, A.; VEILLON, L. Systèmes sols ferrallitiques podzols en région amazonienne. In: RIGHI, D.; CHAUVEL, A (Ed.). **Podzols et Podzolization**. Plaisier et Paris: AFES et INRA, 1987. p. 53-65, 1987.

LUCAS, Y.; NAHON, D.; CORNU, S.; EYROLLE, F. Genèse et fonctionnement des sols en milieu équatorial. **Comptes Rendus de l'Académie des Sciences**, Paris, v. 322, p. 1-16, 1996.

MCKENZIE, N.J.; RYAN, P.J. Spatial prediction of soil properties using environmental correlation. **Geoderma**, Amsterdam, v. 89, p. 67-94, 1999.

MEERSMANS, J.; DE RIDDER, F.; CANTERS, F.; DE BAETS, S.; VAN MOLLE, M. A multiple regression approach to assess the spatial distribution of soil organic carbon (SOC) at the regional scale (Flanders, Belgium). **Geoderma**, Amsterdam, v. 143, p. 1–13, 2008.

MINASNY, B.; MCBRATNEY, A.B.; MENDONÇA-SANTOS, M.L.; ODEH, I.O.A.; GUYON, B. Prediction and digital mapping of soil carbon storage in the Lower Namoi Valley. **Australian Journal of Soil Research**, Melbourne, v. 44, p. 233–244, 2006.

MINASNY, B.; MCBRATNEY, A.B.; SALVADOR-BLANES, S. Quantitative models for pedogenesis: A review, **Geoderma**, Amsterdam, v. 144, p. 140–157, 2008.

MISHRA, U.; LAL, R.; LIU, D.; VAN MEIRVENNE, M. Predicting the spatial variation of soil organic carbon pool at a regional scale, **Soil Science Society of America Journal**, Madison, v. 74; p. 906–914, 2010.

MONTES, C.R.; LUCAS, Y.; MELFI, A.J.; ISHIDA, D.A. Systèmes sols ferrallitiques–podzols et genèse des kaolins. **Comptes Rendus Geoscience**, Paris, v. 339, p. 50–56, 2007.

MONTES, C.R.; LUCAS, Y.; PEREIRA, O.J.R.; ACHARD, R.; GRIMALDI, M.; MELFI, A. J. Deep plant-derived carbon storage in Amazonian podzols. **Biogeosciences**, Göttingen, Germany, v. 8, p. 113-120, 2011.

NASCIMENTO, N.R.; BUENO, G.T.; FRITSCH, E.; HERBILLON, A.J.; ALLARD, T.; MELFI, A.; ASTOLFO, R.; BOUCHER, H.; LI, Y. Podzolization as a deferralization process: a study of an Acrisol- Podzol sequence derived from Palaeozoic sandstones in the northern upper Amazon Basin. **European Journal of Soil Science**, Oxford, v. 55, p. 523–538, 2004.

NASCIMENTO, N.R.; FRITSCH, E.; BUENO, G.T.; BARDY, M.; GRIMALDI, C.; MELFI, A.J. Podzolization as a deferralization process: dynamics and chemistry of ground and surface waters in an Acrisol–Podzol sequence of the upper Amazon Basin. **European Journal of Soil Science**, Oxford, v. 59, p. 911-924, 2008.

ODEH, I.O.A.; MCBRATNEY, A.B.; CHITTLEBOROUGH, D.J. Further results on prediction of soil properties from terrain attributes - Heterotopic cokriging and Regression-kriging. **Geoderma**, Amsterdam, v. 67, p. 215-226, 1995.

PEBESMA, E.J. Multivariate geostatistics in R: the gstat package. **Computers and Geosciences**, Amsterdam, v. 30, p. 683-691, 2004.

PEDRO, G. Podzols et Podzolisation: un problème pédologique fort ancien, mas toujours d'actualité. In: **PODZOLS ET PODZOLISATION. COMPTES RENDUS DE LA TABLE RONDE INTERNATIONALE**, 1986, Poitiers, France. Paris, FRA: INRA, 1987. p. 1-10.

PEREIRA, O.J.R.; MONTES, C.R.; LUCAS, Y.; SANTIN, R.C.; MELFI, A.J. A multisensor approach for mapping plant-derived carbon storage in Amazonian podzols. **International Journal of Remote Sensing**, London, v. 36, n. 8, p. 2076-2092, 2015.

POST, W.M.; EMANUEL, W.R.; ZINKE, P.J.; STANGENBERGER, A.G. Soil carbon pools and world life zones. **Nature**, London, v. 298, p. 156-159, 1982.

POST, W.M.; IZAURRALDE, R.C.; MANN, L.K.; BLISS, N. Monitoring and verifying changes of organic carbon in soil. **Climatic Change**, Heidelberg, v. 51, p. 73-99, 2001.

QUINLAN, J.R. **C4. 5: Programs for machine learning**. San Mateo, CA: Morgan Kaufmann, 1993.

SAATCHI, S.S.; HOUGHTON, R.A.; DOS SANTOS ALVALÁ, R.C., SOARES, J.V.; YU, Y. Distribution of Aboveground Live Biomass in the Amazon Basin. **Global Change Biology**, Oxford, v. 13, p. 816-837, 2007.

SIMBAHAN, G.C.; DOBERMANN, A.; GOOVAERTS, P.; PING, J.; HADDIX, M.L. Fine-resolution mapping of soil organic carbon based on multivariate secondary data. **Geoderma**, Amsterdam, v. 132, p. 471-489, 2006.

STATSOFT INC. **STATISTICA** (Data Analysis Software System). Version 8. Tulsa, 2006. Available at: <[www.statsoft.com](http://www.statsoft.com)>.

USGS. Earth Resources Observation and Science. **EROS Center**. Reston, VA, 2010. Available at: <<http://eros.usgs.gov/>>. Accessed in: Mar. 10, 2014.

VAN-MEIRVENNE, M.; PANNIER, J.; HOFMAN, G.; LOUWAGIE, G. Regional characterization of the long-term change in soil organic carbon under intensive agriculture. **Soil Use and Management**, Hoboken, v. 12, p. 86-94, 1996.

WEBSTER, R.; OLIVER, M A. **Geostatistics for environmental scientists**. 2. ed. Chichester: Wiley, 2007. 330 p.

## 6. GENERAL REMARKS

In this research we applied remote sensing imagery and field sample data in order to quantify and map the soil carbon stock (SOC) in soils of the Rio Negro basin, with special attention to Podzols, due to their high capacity of storing organic carbon as soil organic matter in deep thick spodic (Bh) horizons. Previous research developed in the same region, has found that Podzols might store  $86.8 \pm 7.1 \text{ kg C m}^{-2}$ , which is about three times bigger than the SOC stock found in adjacent soils (mostly Acrisols and Ferralsols). However, the number of soil samples, the extrapolation methods and the ancillary data used in previous studies were insufficient to make precise estimates of SOC stock in the whole area of Rio Negro basin, due to the impossibility of estimating associated error and the lack of information concerning soil depths below 1m. Therefore, a systematic study was employed evolving the quantification of SOC stock at local and regional map scales, as well as at the scale of the Rio Negro basin.

Nowadays, there are no systematic researches in Amazon forest concerning the deep SOC stock. Thus, most of the surveys are centred in the first soil metre, taking into account the SOC stock that is much more labile in the short term, due to the interaction of the superficial and sub-superficial soil horizons with atmosphere. Therefore, the estimates usually are limited to the first 0.3m soil depth. The lack of studies concerning deep SOC storage capacity is related to a series of factors, notably: (1) the still limited knowledge of the extent of different kinds of soil of the world, especially in the tropical regions; (2) the limited availability of soil databases; (3) the considerable spatial variation in carbon content, and soil bulk density, as well as the complexity to retrieve soil bulk density data; and ultimately, (4) the clear effects of climate, relief, parent material, vegetation and land use that are still poorly explored. Therefore, in this research we considered the aforementioned points as relevant factors on mapping SOC stock at Rio Negro basin map scale.

In the first and second chapters of the present research, we explored the usage of remote sensing images widely available in Amazon region, for mapping soil cover and soil orders by correlating remote sensing variables, notably, vegetation spectral signatures, surface temperature, topography and soils. We observed that the amount of carbon stored in soils is related to environmental aspects such as topography, vegetation type, and soil surface moisture. These aspects are the key for spatializing Podzol in Rio Negro basin, evolving map scales greater than the one available in current maps of this region. Thus, new efforts are

necessary for mapping the soils of Rio Negro basin at regional and local map scales, which would allow for a better estimate of the SOC storage capacity of this region.

In the chapters three and four, we discussed the application of pedotransfers (PTF) models and kriging techniques for mapping SOC stock in Rio Negro basin at 1m and 3m soil depths. We observed that the combination of PFT function and kriging interpolation was an efficient way to estimate the deep SOC stock of Rio Negro basin soils, which resulted in systematic SOC stock maps with an unpredicted precision. According to our maps, the Rio Negro basin had an absolute SOC stock of about 5.75 Pg at 1m soil depth and 10.12 Pg at 3m soil. Nevertheless, the amount of carbon stored in Podzols at 6m soil depth is close to the one stored in the entire region of Rio Negro basin at 3m soil depth.

Previous researches carried out in the year of 2002, have shown that the first 0.3m of the Brazilian soils, store about  $36.4 \pm 3.4$  Pg of SOC. According to our research, the area of the Rio Negro basin stores about 10 Pg of SOC at 3m soil depth, which represents 27% of the SOC stock of Brazil. However, the area of the Rio Negro basin represents 14% of the total area of the Brazilian Amazonian basin and just 6.4% of the Brazilian area. Given this, our findings highlight the importance of the SOC stock of the Rio Negro basin at national scale, with especial attention to Podzol areas. If we extrapolate our estimates to deeper soil horizons in areas of Podzols (6m soil depth), these soils would have an absolute SOC stock of 9.19 Pg, which represents 25% of the total SOC stock of Brazilian soils at 0.3m soil depth. Even though, the Podzols cover just 1.3% of the total area of Brazil, it has the capability of storing one quarter of the total soil carbon of Brazilian soils at surface and sub-superficial soil horizons. Given this, the amount of organic carbon stored in Podzols is significantly higher, than the one presented in previous systematic researches. Thus, this new estimate lead us to the following questions: How sensitive are these SOC stocks to global climate changes? Are the deep SOC stock sensible to mineralization, due to changes in soil cover?

We limited our research to the estimation of the SOC stock in Rio Negro basin, with any assumption about the sensibility of these stocks to mineralization, according to future changes in the soil hydrological regime. However, unconcluded results obtained by researches developed in this area, might suggest that future climate changes, followed by oxygenation in superficial and eluvial soil horizons could cause the intensification of microbial activity in spodic horizons (Bh) leading to the intensification of CO<sub>2</sub> emissions by Podzols. Nevertheless, nowadays, the Podzols of Rio Negro basin are protected by a continuous forest (Campinarana) and are subject to the highest rainfall rates of Brazil, being around 3000mm year<sup>-1</sup> in the high Rio Negro basin.

We observed some areas of Podzols located at the east region of the Rio Negro basin (Roraima State), where the annual rainfall is about 1900 mm year<sup>-1</sup>. In this region the average SOC stock of Podzols, according to our results, is about 8 to 12 kg C m<sup>-2</sup>, significantly lower than the values found in areas of giant Podzols located at the high Rio Negro basin, where the stocks are above 60 kg C m<sup>-2</sup>. The clear difference in SOC stock between those two areas, for the same soil type (Podzol), might be an indicative of the high sensibility of the SOC stock of Podzols to climate change, assuming a future decreasing in annual rainfall, flowed by the oxygenation in spodic horizons. Therefore, future researches focused on the assessment of the sensibility of deep SOC stocks of Podzols to mineralization in this region are essential in order to understand the effect of the mineralization of these stocks on the greenhouse gases emissions.

## REFERENCES

- BATJES, N.H. Carbon and nitrogen stocks in the soils of Central and Eastern Europe. **Soil Use and Management**, Hoboken, v. 18, n. 4, p. 324–329, 2002.
- BATJES, N.H. Total carbon and nitrogen in the soils of the world. **European Journal of Soil Science**, Oxford, v. 47, p. 151-163, 1996.
- BATJES, N.H.; SOMBROEK, W.G. Possibilities for carbon sequestration in tropical and subtropical soils. **Global Change Biology**, Oxford, v. 3, p. 161-173, 1997.
- BATJES, N. H.; DIJKSHOORN, J. A. Carbon and nitrogen stocks in the soils of the Amazon Region. **Geoderma**, Amsterdam, v. 89, p. 273–286, 1999.
- BRASIL. Ministério das Minas e Energia. **Projeto RADAMBRASIL. Folha SA.19-Içá-AM: Geomorfologia**. Rio de Janeiro, 1977. p. 125-180. (Levantamento dos Recursos Naturais, v. 14).
- BRAVARD, S.; RIGHI, D. Geochemical differences in an oxisolspodosol toposequence of Amazonia (Brazil). **Geoderma**, Amsterdam, v. 44, p. 29–42, 1989.
- BURINGH, P. Organic carbon in the soils of the world. In: WOODWELL, G. (Ed.). **The role of terrestrial vegetation in the global carbon cycle: measurement by remote sensing**. Chichester: Wiley, 1994. p. 91-109. (SCOPE, v. 23).
- CAMPBELL K.E. JR., FRAILEY C.D., ROMERO-PITTMAN L. The Pan-Amazonian Ucayali Peneplain, late Neogene sedimentation in Amazonia, and the birth of the modern Amazon River system. **Palaeogeography, Palaeoclimatology, Palaeoecology**, v. 239, p. 166-219, 2006.

CERRI, C.E.P.; EASTER, M.; PAUSTIAN, K.; KILLIAN, K.; COLEMAN, K.; BERNOUX, M.; FALLOON, P.; POWLSON, D.S.; BATJES, N.H.; MILNE, E.; CERRI, C.C. Predicted Soil Organic Carbon Stocks and Changes in the Brazilian Amazon between 2000 and 2030. **Agriculture Ecosystems & Environment**, Amsterdam, v. 122, n. 1, p. 58-72, 2007.

DUBROEUCQ, D.; VOLKOFF, B. From oxisols to spodosols and histosols: evolution of the soil mantles in the Rio Negro Basin (Amazonia). **Catena**, v. 32, p. 245–280, 1998.

DUBROEUCQ, D.; VOLKAFF, B.; FAURE, P. Les couvertures pédologiques à podzols du bassin du haut Rio Negro (Amazonie). **Etude et Gestion des Sols**, Orléans, v. 6, p. 131–153, 1999.

EASTER, M.; PAUSTIAN, K.; KILLIAN, K.; WILLIAMS, S.; FRENG, T.; AL-ADAMAT, R.; BATJES, N.H.; BERNOUX, M.; BHATTACHARYYA, T.; CERRI, C.C.; CERRI, C.E.P.; COLEMAN, K.; FALLOON, P.; FELLER, C.; GICHERU, P.; KAMONI, P.; MILNE, E.; PAL, D.K.; POWLSON, D.S.; RAWAJHIF, Z. The GEFSOC soil carbon modeling system: A tool for conducting regional-scale soil carbon inventories and assessing the impacts of land change on soil carbon. **Agriculture Ecosystems & Environment**, Amsterdam, v. 122, p. 13–25, 2007.

EMBRAPA. **Digital Soils System Information** (Brazilian Soils Database). Rio de Janeiro: Embrapa Solos, 2014. Available at: [http://www.bdsolos.cnptia.embrapa.br/consulta\\_publica.html](http://www.bdsolos.cnptia.embrapa.br/consulta_publica.html). Accessed in: Oct. 25, 2014.

ESWARAN, H.; VAN DEN BERG, E.; REICH, P. Organic carbon in soils of the world. **Soil Science Society of American Journal**, Madison, v. 57, p. 192–194, 1993.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. Geoscience Division - DGC. Coordenação de Recursos Naturais e Estudos Ambientais - CREN. **Georefered Maps of Natural Resources**: Scale 1:250:000. Digital Format: shp. Rio de Janeiro, 2008. Available at: [ftp://geofp.ibge.gov.br/mapas/banco\\_dados\\_georeferenciado\\_recursos\\_naturais](ftp://geofp.ibge.gov.br/mapas/banco_dados_georeferenciado_recursos_naturais). Accessed in: Aug. 15, 2014.

KIMBLE, J.M.; HEATH, L.S.; BIRDSEY, R.; LAL, R. **The potential of US forest soils to sequester carbon and mitigate the greenhouse effect**. Boca Raton: CRC Press, 1990. 429 p.

LUCAS, Y.; CHAUVEL, A.; BOULET, R.; RANZANI, G.; SCATOLINI, F. Transição latossolos-podzols sobre a formação Barreiras na região de Manaus, Amazônia. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 8, p. 325–335, 1984.

LUCAS, Y.; CHAUVEL, A. Soil formation in tropically weathered terrains, in: **Regolith exploration geochemistry in tropical and subtropical terrains**, edited by: C. R. M. Butt and H. Zeegers, Elsevier, Amsterdam – London – New-York – Tokyo, p. 57–77, 1992.

LUCAS, Y.; NAHON, D.; CORNU, S.; EYROLLE, F. Genèse et fonctionnement des sols en milieu équatorial. **Comptes Rendus de l'Académie des Sciences**, Paris, v. 322, p. 1-16, 1996.

LUCAS, Y. The role of the plants in controlling rates and products of weathering: importance of the biological pumping, **Annual Review of Earth and Planetary Science**, Palo Alto, v. 29, p. 135–163, 2001.

MONTES, C.R.; LUCAS, Y.; PEREIRA, O.J.R.; ACHARD, R.; GRIMALDI, M.; MELFI, A.J. Deep plant-derived carbon storage in Amazonian podzols. **Biogeosciences**, Göttingen, Germany, v. 8, p. 113-120, 2011.

NASCIMENTO, N.R.; BUENO, G.T.; FRITSCH, E.; HERBILLON, A.J.; ALLARD, T.; MELFI, A.; ASTOLFO, R.; BOUCHER, H.; LI, Y. Podzolization as a deferralitization process: a study of an Acrisol- Podzol sequence derived from Palaeozoic sandstones in the northern upper Amazon Basin. **European Journal of Soil Science**, Oxford, v. 55, p. 523–538, 2004.

NASCIMENTO, N.R.; FRITSCH, E.; BUENO, G.T.; BARDY, M.; GRIMALDI, C.; MELFI, A.J. Podzolization as a deferralitisation process: dynamics and chemistry of ground and surface waters in an Acrisol–Podzol sequence of the upper Amazon Basin. **European Journal of Soil Science**, Oxford, v. 59, p. 911-924, 2008.

PEREIRA, O.J.R.; MONTES, C.R.; LUCAS, Y.; SANTIN, R.C.; MELFI, A.J. A multisensor approach for mapping plant-derived carbon storage in Amazonian podzols. **International Journal of Remote Sensing**, London, v. 36, n. 8, p. 2076-2092, 2015.

POST, W.M.; EMANUEL, W.R.; ZINKE, P.J.; STANGENBERGER, A.G. Soil carbon pools and world life zones. **Nature**, London, v. 298, p. 156-159, 1982.

POST, W.M.; IZAURRALDE, R.C.; MANN, L.K.; BLISS, N. Monitoring and verifying changes of organic carbon in soil. **Climatic Change**, Heidelberg, v. 51, p. 73–99, 2001.

SOMBROEK, W.G.; NACHTERGAELE, F.O.; HEBEL, A. Amounts, dynamics and sequestrations of carbon in tropical and subtropical soils. **Ambio**, Heidelberg, v. 22, p. 417–426, 1993.

VAN MEIRVENNE M.; PANNIER J.; HOFMAN G.; LOUWAGIE G. Regional characterization of the long-term change in soil organic carbon under intensive agriculture. **Soil Use and Management**, Hoboken, v. 12, p. 86–94, 1996.