



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

RICARDO DE OLIVEIRA DART

SENSORIAMENTO REMOTO APLICADO À PREDIÇÃO DE CLASSES DE SOLOS EM FLORESTA
TROPICAL SECA: comparação entre tipos, fontes e épocas de aquisição

RIO DE JANEIRO
FEVEREIRO DE 2017

Ricardo de Oliveira Dart

SENSORIAMENTO REMOTO APLICADO À PREDIÇÃO DE
CLASSES DE SOLOS EM FLORESTA TROPICAL SECA:
comparação entre tipos, fontes e épocas de aquisição

1 volume

Dissertação de Mestrado apresentada ao Programa de
Pós-Graduação em Geografia, Universidade Federal do
Rio de Janeiro, como requisito parcial à obtenção do
título de Mestre em Geografia

Orientador: Nelson Ferreira Fernandes

Coorientador: Gustavo de Mattos Vasques

Rio de Janeiro

Fevereiro de 2017

Dart, Ricardo de Oliveira

Sensoriamento remoto aplicado à predição de classes de solos em floresta tropical seca: comparação entre tipos, fontes e épocas de aquisição, 2017, 64 f.; il,

Dissertação (Mestrado em Geografia) - Universidade Federal do Rio de Janeiro – UFRJ / Instituto de Geociências / Programa de Pós-Graduação em Geografia – PPGG, Rio de Janeiro, 2017.

Orientador: Nelson Ferreira Fernandes

Coorientador: Gustavo de Mattos Vasques

1 Mapeamento Digital de Solos. 2. Árvore de Classificação. 3. Modelo Digital de Elevação. 4. Imagens de Satélite. 5. Imagens Gamarradiométricas – Dissertação.

I. Fernandes, Nelson Ferreira (Orient.), II. Universidade Federal do Rio de Janeiro Programa de Pós-Graduação em Geografia, III. Sensoriamento remoto aplicado à predição de classes de solos em floresta tropical seca: comparação entre tipos, fontes e épocas de aquisição.

RICARDO DE OLIVEIRA DART

**SENSORIAMENTO REMOTO APLICADO À PREDIÇÃO DE CLASSES DE SOLOS EM FLORESTA
TROPICAL SECA: comparação entre tipos, fontes e épocas de aquisição**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Geografia, Universidade Federal do Rio de Janeiro, como requisito parcial à obtenção do título de Mestre em Geografia

Aprovado por:

Orientador, Nelson Ferreira Fernandes, PhD, UFRJ

Coorientador, Gustavo de Mattos Vasques, PhD, Embrapa

Carla Bernadete Madureira Cruz, Dra., UFRJ

Maurício Rizzato Coelho, Dr., Embrapa

Rio de Janeiro
Fevereiro de 2017

Dedicatória

*Para minha esposa Mariana, minhas filhas Sofia e Luísa,
meu pai Fernando e minha mãe Zênia.*

Agradecimentos

Gostaria de agradecer a minha esposa Mariana, que eu amo e me proporcionou as maiores alegrias da minha vida, Sofia e Luísa, e me incentivou e apoiou a concluir o curso de mestrado. Aos meus pais (Zênia e Fernando) que eu amo muito e sempre estão por perto para me ajudar e sempre se sacrificaram para me proporcionar o melhor. Às minhas queridas sogra e cunhada Conceição Costa e Andrea Costa que sempre me ajudam cuidando de minha filha, Sofia, para que eu possa me dedicar ao trabalho.

Em especial ao meu coorientador e colega Gustavo Vasques por todo o apoio desde a elaboração da proposta de projeto, sempre disponível para ajudar e com ótimas revisões e comentários, ao meu orientador Nelson Fernandes por aceitar me orientar mesmo que em uma linha diferente da sua principal linha de atuação, pela confiança e pela liberdade que me deu para desenvolver meu trabalho.

Gostaria de agradecer ao apoio financeiro do projeto MP3 da Embrapa “Mapeamento Digital de Classes e Atributos de Solos: um estudo de caso no Bioma Caatinga, Região Norte do Estado de Minas Gerais” (nº 03.10.06.013.00.00), coordenado pelo pesquisador Maurício Rizzato Coelho, a quem agradeço imensamente pelas ótimas experiências vividas em campo e nos trabalhos de escritório e por todo apoio e ensinamento (desde 2009 trabalhamos juntos no Parque Estadual da Mata Seca, PEMS).

Ao suporte financeiro para as campanhas de campo advindos do projeto IAI “Functional links between aboveground changes and belowground activity with land use in the Americas: Soil biodiversity and food security” (CRNII_014), coordenado pelo professor Ricardo Luiz Louro Berbara da UFRRJ, ao professor Mário Marcos do Espírito Santo (UNIMONTES) pelo apoio logístico e fornecimento de informações sobre a área de estudo, ao pesquisador Edgar Shinzato (Serviço Geológico do Brasil) pelo fornecimento da imagem gamarradiométrica, ao apoio fornecido pelo gerente José Luiz do PEMS (IEF-MG) e sua equipe de ajudantes de campo (Dorinho, Boi, Buiú, Santo e outros cujos nomes/apelidos me falham a memória), que nos guiaram por todas as trilhas possíveis, zelaram por nossa segurança durante as campanhas de campo e abriram as trincheiras para descrição e coleta de solo.

À minha ex-chefe e colega Maria de Lourdes Mendonça Santos pelos ensinamentos e por ter me apresentado em 2006 ao tema Mapeamento Digital de Solos, o que me possibilitou trabalhar em uma área empolgante e com diversos desafios, na qual sinto muito prazer em trabalhar.

À Rachel Bardy, que foi meu primeiro contato com o universo da Embrapa, em 2005, via orientação de estágio acadêmico. Desde então eu não quis mais sair desta maravilhosa empresa, que desde 2009 tenho prazer de servir como funcionário e à qual agradeço imensamente pelo apoio ao meu desenvolvimento profissional.

Ao meu querido chefe Mario Aglio, por todo apoio e incentivo para poder concluir minha jornada. Aos colegas da Embrapa Solos José Silva, Elaine Fidalgo, Marysol Schuler, Margareth Simões, Jesus Mansilla, Ronaldo Oliveira, Ricardo Arcanjo, Wenceslau Teixeira e Daniel Vidal pelo incentivo, apoio e amizade.

Epígrafe

*“Vem, vamos embora, que esperar não é saber,
Quem sabe faz a hora, não espera acontecer”*

(GERALDO VANDRÉ)

RESUMO

DART, Ricardo de Oliveira. **Sensoriamento remoto aplicado à predição de classes de solos em Floresta Tropical Seca**: comparação entre tipos, fontes e épocas de aquisição. Rio de Janeiro, 2017, Dissertação (Mestrado em Geografia) – Instituto de Geociências, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2017

Modelos Digitais de Elevação (MDE) e imagens de satélite são frequentemente utilizados como covariáveis ambientais para predição de classes de solos em procedimentos de Mapeamento Digital de Solos (MDS). Para avaliar a utilização de imagens de sensoriamento remoto como covariáveis de entrada para a predição de classes de solo, aplicou-se o algoritmo de árvore de classificação (AC) para predizer classes de solo ao nível de grande grupo, de acordo Sistema Brasileiro de Classificação de Solos, em uma área de floresta tropical seca (FTS) com 102 km² no norte do Estado de Minas Gerais, Brasil. Foram elaborados 17 modelos de AC, comparando três tipos de variáveis relacionadas aos fatores de formação do solo, quais sejam: (i) covariáveis topográficas (*topo*) obtidas a partir de três fontes de MDE (SRTM-90m, SRTM-30m e Ikonos); (ii) covariáveis obtidas a partir de imagens de satélite multiespectrais (*sat*) de duas fontes (Landsat 8 e RapidEye) em dois períodos (seco e úmido), representando o fator de formação “organismos”; e (iii) covariáveis de material de origem (*parent*) obtidas de imagens aéreas gamarradiométricas e magnetométricas. Comparou-se modelos usando covariáveis preditoras específicas de cada fator de formação, bem como, três resoluções espaciais (10-m, 30-m e 90-m) nas covariáveis *topo* e *sat* e apenas a resolução espacial de 90-m para as covariáveis *parent* para fins de predição de classes de solo. Para treinar e validar os modelos, utilizou-se, respectivamente, 296 e 128 observações de solos classificadas em nível de grande grupo, compondo 13 classes. O procedimento de AC apresentou bons resultados para classificar solos em FTS quando comparado a pesquisas similares, com acurácia global (AG) de até 91,6% na etapa de treinamento e até 53,1% na etapa de validação. Os resultados mostram que as covariáveis ambientais gratuitas com menor resolução espacial (90-m) podem produzir boas ou até melhores predições de classes do que as covariáveis ambientais pagas e com maior resolução espacial (10-m). Dentre os três tipos de covariáveis ambientais testadas para estimar classes de solos, os modelos com covariáveis *topo* apresentaram melhor

desempenho, seguido por covariáveis *parent* e pelas covariáveis *sat*. Os modelos *topo* e *parent* conseguiram estimar todas as 13 classes de solos. Entre os modelos *sat*, os que utilizaram imagens no período seco obtiveram resultados mais acurados do que os com imagens no período úmido para estimar classes de solos em FTS. Em conclusão, identificou-se, dentre as covariáveis ambientais avaliadas, quais fatores de formação do solo e quais fontes, resoluções espaciais e períodos de aquisição de imagens de sensores remotos são mais adequados para prever classes de solos em FTS no Brasil.

Palavras-chave: Mapeamento digital de solos, árvore de classificação, modelo digital de elevação, imagens de satélite, imagens gamarradiométricas

ABSTRACT

DART, Ricardo Oliveira. **Sensoriamento remoto aplicado à predição de classes de solos em Floresta Tropical Seca**: comparação entre tipos, fontes e épocas de aquisição. Rio de Janeiro, 2017, Dissertação (Mestrado em Geografia) – Instituto de Geociências, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2017

Digital Elevation Models (DEM) and satellite images are frequently used as environmental covariates for the prediction of soil classes in Digital Soil Mapping (DSM) procedures. To evaluate the use of remote sensing imagery as input covariates for the prediction of soil classes, classification tree (CT) was applied to predict soil classes at the great group level, according to the Brazilian Soil Classification System, in a tropical dry forest (TDF) area with 102 km² in the north of Minas Gerais state, Brazil. Seventeen CT models were derived, comparing as covariates three types of variables related to soil formation factors, including: (i) topographic covariates (*topo*) obtained from three DEM sources (SRTM-90m, SRTM-30m and Ikonos); (ii) covariates obtained from multispectral satellite images (*sat*) from two sources (Landsat 8 and RapidEye) and two periods (dry and wet), representing the “organisms” soil formation factor; and (iii) covariates of parental material (*parent*) obtained from gamma radiometric and magnetometric aerial images. Models using covariates representing each soil formation factor, respectively, as well as three spatial resolutions (10-m, 30-m and 90-m) of these covariates, were compared to predict soil classes. To train and validate the models, 296 and 128 soil class observations at great group level were used, respectively. Compared to similar studies, the CT procedure presented good results to classify soils in TDF, as the overall accuracy (OA) reached up to 91.6%, in the training stage, and up to 53.1% in the validation stage. The results show that freely available environmental covariates with coarser spatial resolution (90-m) may produce good or even better soil class predictions than paid environmental covariates with higher spatial resolution (10-m). Among the three types of environmental covariates tested to estimate soil classes, models with *topo* covariates showed better performance, followed by *parent* covariates and then *sat* covariates. The *topo* and *parent* models were able to estimate all 13 soil classes. Among the *sat* models, those that used images from the dry period obtained more accurate results than those using images from the wet period to estimate soil classes in TDF. In conclusion, the study identified, among the

environmental covariates tested, which soil formation factors, and which sources, spatial resolutions and acquisition periods of remote sensing images better predict soil classes in TDF in Brazil.

Keywords: Digital soil mapping, classification tree, digital elevation model, satellite images, gamma radiometric images

Lista de Equações

Equação 1. Normalized difference vegetation index (NDVI).....	26
Equação 2. Ratio vegetation index (RVI).....	26
Equação 3. Soil-adjusted vegetation index (SAVI).....	26
Equação 4. Acurácia global (AG).....	32
Equação 5. Acurácia do produtor (AP).....	32
Equação 6. Acurácia do usuário (AU).....	32
Equação 7. Índice KAPPA (K).....	32

Lista de Figuras

Figura 1. Abordagem metodológica do Mapeamento Digital de Solos, seguindo o modelo conceitual s.c.o.r.p.a.n.	4
Figura 2. Exemplo de árvore de classificação (AC) para estimar classes de solos.....	9
Figura 3. Dois exemplos de covariáveis gamarradiométricas utilizadas. A) Concentração de K (%); B) Concentração de Th (ppm).	12
Figura 4. Localização das Florestas Tropicais Secas no Brasil (IBGE, 2006).....	14
Figura 5. Área de estudo: A) Parque Estadual da Mata Seca (PEMS) no norte do Estado de Minas Gerais; B) Mapa de solos convencional do PEMS (COELHO et al., 2013) com legenda agrupada em nível de grande grupo, mostrando os domínios geológicos do PEMS e a localização das observações de solos; C) Imagem Landsat 8 OLI (02/07/2014, composição RGB 6,5,4) mostrando o traçado do perfil topográfico cruzando os três domínios geológicos do PEMS; D) Perfil topográfico através dos três domínios geológicos do PEMS; E) Proporção do conjunto de observações de treinamento (T) e validação (V).	15
Figura 6. Fluxograma do procedimento metodológico usado para obter e validar os 17 modelos de predição e mapas de classe de solos. A seta circular vermelha indica que as etapas de “Modelagem” foram executadas para todos os 17 modelos (Tabela 3).....	17
Figura 7. Exemplo do modelo digital de elevação obtido de imagens Ikonos (MDE-I), reamostrado para as três resoluções espaciais utilizadas, sendo: A) 10-m; B) 30m; e C) 90-m.	20
Figura 8. Mapas de solos obtidos pelos modelos usando imagens RapidEye (RE) nos períodos seco (<i>dry</i>) e úmido (<i>wet</i>).....	43
Figura 9. Mapas de solos obtidos pelos modelos usando imagens Landsat 8 (L8) nos períodos seco (<i>dry</i>) e úmido (<i>wet</i>) de diferentes fontes e resoluções espaciais.	44
Figura 10. Mapas de solos obtidos pelos modelos usando covariáveis topográficas (topo) de diferentes fontes e resoluções espaciais.....	46
Figura 11. Mapa de solos obtido pelo modelo usando variáveis gamarradiométricas e magnetométricas (parent).....	47
Figura 12. Detalhe do mapa de classes de solo produzido a partir do modelo M90m_D_s3, mostrando topossequência de solos na área de depósitos sedimentares do Rio São Francisco. A) Extrato do mapa de solos obtido pelo modelo M90m_D_s3; B) Localização da topossequência de solos; e C) Perfil topográfico mostrando a topossequência de solos preditos pelo modelo M90m_D_s3, com destaque para a identificação da classe GXve na posição mais baixa da paisagem (~446 m).	52
Figura 13. Detalhe do mapa de classes de solo produzido pelo modelo M90m_D_s3, mostrando topossequência de solos na área de transição do Grupo Urucuia para o Grupo Bambuí. A) Mapa de solos obtido pelo modelo M90m_D_s3; B) Localização da topossequência de solos; C) Perfil topográfico mostrando a topossequência de solos preditos pelo modelo M90m_D_s3.	54

Lista de Tabelas

Tabela 1. Classes de solos agrupadas em nível de grande grupo, de acordo com o Sistema Brasileiro de Classificação de Solos, nos conjuntos de treinamento (T) e validação (V).	19
Tabela 2. Conjuntos de covariáveis ambientais utilizadas para predizer classes de solos.	24
Tabela 3. Modelos, fontes e covariáveis ambientais utilizadas para predizer classes de solos.....	29
Tabela 4. Faixas de concordância em função do índice KAPPA (K), adaptadas de Landis e Koch (1977).	33
Tabela 5. Resultados dos modelos ordenados do maior para o menor índice KAPPA (K) calculado no conjunto de dados de validação.	35
Tabela 6. Tipo de covariável utilizada em cada modelo, número de covariáveis e cinco covariáveis mais utilizadas mostrando a frequência em que aparecem na árvore de classificação, e covariáveis que não foram utilizadas nos modelos.	38
Tabela 7. Acurácias do usuário (AU) e do produtor (AP) nos modelos sat, para o conjunto de dados de validação.....	49
Tabela 8. Acurácias do usuário (AU) e do produtor (AP) nos modelos topo e parent, para o conjunto de dados de validação.	50
Tabela 9. Matriz de confusão do modelo M90m_D_s3, na etapa de treinamento.	56
Tabela 10. Matriz de confusão do modelo M90m_D_s3, na etapa de validação.	56

Lista de Abreviaturas e Siglas

\hat{K}	Índice KAPPA
1DV	1ª derivada vertical do campo magnético
<i>a</i>	Tempo na fórmula <i>s.c.o.r.p.a.n</i>
AC	Árvore de classificação
AG	Acurácia global
AP	Acurácia do produtor
ASP	<i>Aspect</i>
AU	Acurácia do usuário
Aw	Clima tropical, segundo classificação de Köppen
BSw	Clima semiárido quente, segundo classificação de Köppen
<i>c</i>	Clima na fórmula <i>s.c.o.r.p.a.n</i>
CAR	<i>Catchment area</i>
cLHS	<i>Conditioned Latin hypercube sampling</i>
CNBL	<i>Channel network base level</i>
CT	<i>Classification tree</i>
CT	Taxa de exposição da contagem total
CVI	<i>Convergence index</i>
CXbe	CAMBISSOLOS HÁPLICOS Tb Eutróficos
CXve	CAMBISSOLOS HÁPLICOS Ta Eutróficos
CY	CAMBISSOLOS FLÚVICOS
DEM	Digital elevation model
DIFIN	<i>Diffuse insolation</i>
DRIN	<i>Direct insolation</i>
DSM	Digital soil mapping
Embrapa	Empresa Brasileira de Pesquisa Agropecuária
FTS	Floresta tropical seca
GM	GLEISSOLOS MELÂNICOS
GX	GLEISSOLOS HÁPLICOS

GXve	GLEISSOLOS HÁPLICOS Ta Eutróficos
GXve	GLEISSOLOS HÁPLICOS Ta Eutróficos
K	Potássio
L8	Landsat 8
LAd	LATOSSOLOS AMARELOS Distróficos
LSF	<i>LS-factor</i>
LVAd	LATOSSOLOS VERMELHO-AMARELOS Distróficos
LVAe	LATOSSOLOS VERMELHO-AMARELOS Eutróficos
LVd	LATOSSOLOS VERMELHOS Distróficos
LVe	LATOSSOLOS VERMELHOS Eutróficos
Mag	Campo magnético total
MBI	<i>Mass balance index</i>
MDE	Modelo digital de elevação
MDS	Mapeamento digital de solos
MMA	Ministério do Meio Ambiente
MRRTF	<i>Multiresolution index of ridge top flatness</i>
MRVBF	<i>Multiresolution index of valley bottom flatness</i>
MSP	<i>Mid slope position</i>
MXk	CHERNOSSOLOS HÁPLICOS Carbonáticos
MXo	CHERNOSSOLOS HÁPLICOS Órticos
<i>n</i>	Espaço na fórmula <i>s.c.o.r.p.a.n</i>
NDVI	<i>Normalized difference vegetation index</i>
<i>o</i>	Organismos na fórmula <i>s.c.o.r.p.a.n</i>
OLI	<i>Operational Land Imager</i>
<i>p</i>	Material de origem na fórmula <i>s.c.o.r.p.a.n</i>
<i>Parent</i>	Variável relativa ao material de origem
PEMS	Parque Estadual da Mata Seca
PFCV	<i>Profile curvature</i>
PLCV	<i>Plan curvature</i>

<i>r</i>	Relevo na fórmula <i>s.c.o.r.p.a.n</i>
RE	<i>RapidEye</i>
RSP	<i>Relative slope position</i>
RVI	<i>Ratio vegetation index</i>
RY	NEOSSOLOS FLÚVICOS
<i>s</i>	Solo na fórmula <i>s.c.o.r.p.a.n</i>
<i>sat</i>	Imagens de satélite
SAVI	<i>Soil-adjusted vegetation index</i>
SHP	<i>Slope height</i>
SiBCS	Sistema Brasileiro de Classificação de Solos
Sinal	Sinal analítico do campo magnético total
SLG	<i>Slope length</i>
SLO	<i>Slope</i>
SRTM	<i>Shuttle Radar Topography Mission</i>
T	Treinamento
TDF	Tropical dry forest
Th	Tório
ThK	Razão tório/potássio
TOIN	<i>Total insolation</i>
<i>Topo</i>	Variáveis topográficas
TPI	<i>Topographic position index</i>
TRI	<i>Terrain ruggedness index</i>
TST	<i>Terrain Surface Texture</i>
TWI	<i>Topographic wetness index</i>
U	Urânio
UK	Razão urânio/potássio
UTh	Razão urânio/tório
UTM	Universal Transversa de Mercator
V	Validação

VDCN	<i>Vertical distance to channel network</i>
VDP	<i>Valley depth</i>
VX	VERTISSOLOS HÁPLICOS
WGS	<i>World Geodetic System</i>

Sumário

1. INTRODUÇÃO	1
1.1. HIPÓTESES.....	3
1.2. OBJETIVOS	3
2. REVISÃO BIBLIOGRÁFICA	4
2.1. MAPEAMENTO DIGITAL DE SOLOS	4
2.2. ÁRVORES DE CLASSIFICAÇÃO	9
2.3. GAMARRADIOMETRIA	11
3. ÁREA DE ESTUDO	13
3.1. CONTEXTUALIZAÇÃO	13
3.2. LOCALIZAÇÃO E CARACTERÍSTICAS	15
4. METODOLOGIA	17
4.1. AMOSTRAS DE SOLOS	18
4.2. COVARIÁVEIS AMBIENTAIS	20
4.3. MODELOS E MAPAS	27
4.4. VALIDAÇÃO	31
5. RESULTADOS E DISCUSSÃO	34
5.1. AVALIAÇÃO DOS MODELOS NA ETAPA DE TREINAMENTO	34
5.2. IMPORTÂNCIA DAS VARIÁVEIS AMBIENTAIS, FONTES, RESOLUÇÃO ESPACIAL E PERÍODOS DE AQUISIÇÃO DAS IMAGENS	36
5.3. MODELOS DE PREDIÇÃO E RELAÇÕES SOLO-PAISAGEM.....	41
5.4. VALIDAÇÃO DOS MODELOS E MAPAS	47
5.5. RELAÇÕES SOLO-PAISAGEM DO MELHOR MODELO	51
6. CONCLUSÕES	57
7. REFERÊNCIAS BIBLIOGRÁFICAS	58

1. INTRODUÇÃO

Atualmente existe grande disponibilidade de covariáveis ambientais representando os fatores de formação do solo, incluindo mapas históricos, imagens de sensores orbitais e Modelos Digitais de Elevação (MDE), oferecendo boas oportunidades para desenvolver novas pesquisas em Mapeamento Digital de Solos (MDS). O MDS consiste em um conjunto de métodos quantitativos e estatísticos que possibilitam descrever e prever a variação do solo e de seus atributos na paisagem, eliminando ou reduzindo a subjetividade dos mapeamentos de solos convencionais. No MDS as classes ou atributos do solo são previstas em função de dados (covariáveis) ambientais que representam os fatores de formação do solo (*s.c.o.r.p.a.n*), incluindo dados de solos (MCBRATNEY et al., 2003; LAGACHERIE; MCBRATNEY, 2007).

No entanto, como e em que grau a qualidade dos dados de entrada afeta na acurácia dos mapas finais de solo precisam ser melhor compreendidos. Pesquisas anteriores mostraram que dados com menor resolução espacial possibilitam obter mapas de solo de qualidade igual ou superior aos dados com maior detalhe (SAMUEL-ROSA et al., 2015; CAVAZZI et al., 2013; GRINAND et al., 2008), provavelmente, com menor custo. Embora pareça contraditório, é possível que dados com resoluções espaciais mais detalhadas adicionem ruído em vez de informações úteis para melhorar as estimativas (CAVAZZI et al., 2013). No entanto, esse comportamento não se repete em todas as situações (por exemplo, VASQUES et al., 2012; KIM et al., 2014). Ao contrário, ele depende das características da paisagem, incluindo a variabilidade de solos e fatores ambientais, a densidade e o período das observações de solo, escala nominal ou resolução espacial, período de obtenção das covariáveis, resolução espectral, acurácia dos dados de entrada e covariáveis, incluindo dados de sensores remotos, entre outros.

Poucas pesquisas em MDS compararam fontes de dados de entrada para estimar atributos de solos, seja testando a influência no tamanho do grid de MDE e atributos do terreno (BEHRENS et al., 2014; VASQUES et al., 2012), fontes de sensores orbitais (KIM et al., 2014) ou fontes de variáveis ambientais (SAMUEL-ROSA et al., 2015). Maynard e Levi (2017) avaliaram a utilização de índices de vegetação por diferença normalizada (NDVI) de imagens Landsat TM comparando o efeito da temporalidade na estimativa de classes de textura do solo. Da mesma forma, pouco tem sido feito para avaliar a influência das fontes de dados de entrada e sua qualidade associada para estimar classes de solos. Basicamente, dois tipos de

abordagem têm sido testados, uma baseada em observações de perfis modais de solos (KIM et al., 2012; BRUNGARD et al., 2015) e outra via desagregação de polígonos de mapas de solos convencionais (CAVAZZI et al., 2013).

As pesquisas acima mencionadas, com exceção de Maynard e Levi (2017), compararam fontes de covariáveis ambientais em conjunto, dificultando a interpretação da influência da fonte de um determinado fator de formação do solo individualmente. Além disso, poucos estudos similares em ambientes tropicais, especialmente em florestas tropicais secas (FTS), foram feitos até agora, a exemplo de Dart et al. (2010, 2015), que estimaram classes de solo a nível de subordem no Parque Estadual da Mata Seca (PEMS) no norte de Minas Gerais usando árvores de classificação (AC); e Vasques et al. (2016), que na mesma área realizaram a predição de 11 atributos do solo comparando regressão krigagem e krigagem.

Em suma, não foi observado nenhum trabalho que avalie a influência da fonte, período de aquisição, e resolução espacial das covariáveis ambientais nas relações solo-paisagem, comparando diferentes tipos de covariáveis ambientais e fontes de dados separadamente. Deste modo, o presente estudo visa preencher essa lacuna de conhecimento através de um estudo de caso desenvolvido em área com relevo predominantemente plano com FTS no Brasil utilizando a técnica de árvores de classificação C5 (QUINLAN, 1993).

1.1. HIPÓTESES

As hipóteses da dissertação foram: A) modelo digital de elevação com maior resolução espacial é melhor para estimar classes de solos; B) imagens orbitais com maior resolução espectral são melhores para estimar classes de solos; C) imagens orbitais em período seco possibilitam melhores estimativas de classes de solos do que imagens de períodos úmidos em Florestas Tropicais Secas; D) imagens gamarradiométricas são melhores para estimar solos do que imagens multiespectrais de sensores orbitais (com comprimentos de onda do visível ao infravermelho médio).

1.2. OBJETIVOS

Os objetivos do presente trabalho foram: A) avaliar três tipos de covariáveis ambientais com diferentes resoluções espaciais e fontes de aquisição, a fim de estimar classes de solos em nível de grande grupo de acordo com o Sistema Brasileiro de Classificação de Solos (SiBCS; EMBRAPA, 2006) em uma área com Floresta Tropical Seca no Brasil, utilizando algoritmo de AC; B) validar os resultados utilizando amostras independentes; e C) comparar os mapas resultantes dos modelos de predição via AC com o mapa de solos convencional da mesma área de estudo (COELHO et al., 2013).

2. REVISÃO BIBLIOGRÁFICA

2.1. MAPEAMENTO DIGITAL DE SOLOS

O MDS consiste em um conjunto de técnicas que permitem mapear atributos e classes de solos utilizando métodos quantitativos e técnicas estatísticas para descrever e prever a variação do solo na paisagem, eliminando ou reduzindo a subjetividade dos mapeamentos de solos convencionais. No MDS as classes ou atributos do solo são previstas em função de covariáveis ambientais que representam os fatores de formação do solo (*s.c.o.r.p.a.n.*), incluindo dados de solos (MCBRATNEY et al., 2003; LAGACHERIE; MCBRATNEY, 2007). A **Figura 1** representa a abordagem metodológica do MDS, seguindo o modelo conceitual *s.c.o.r.p.a.n.*

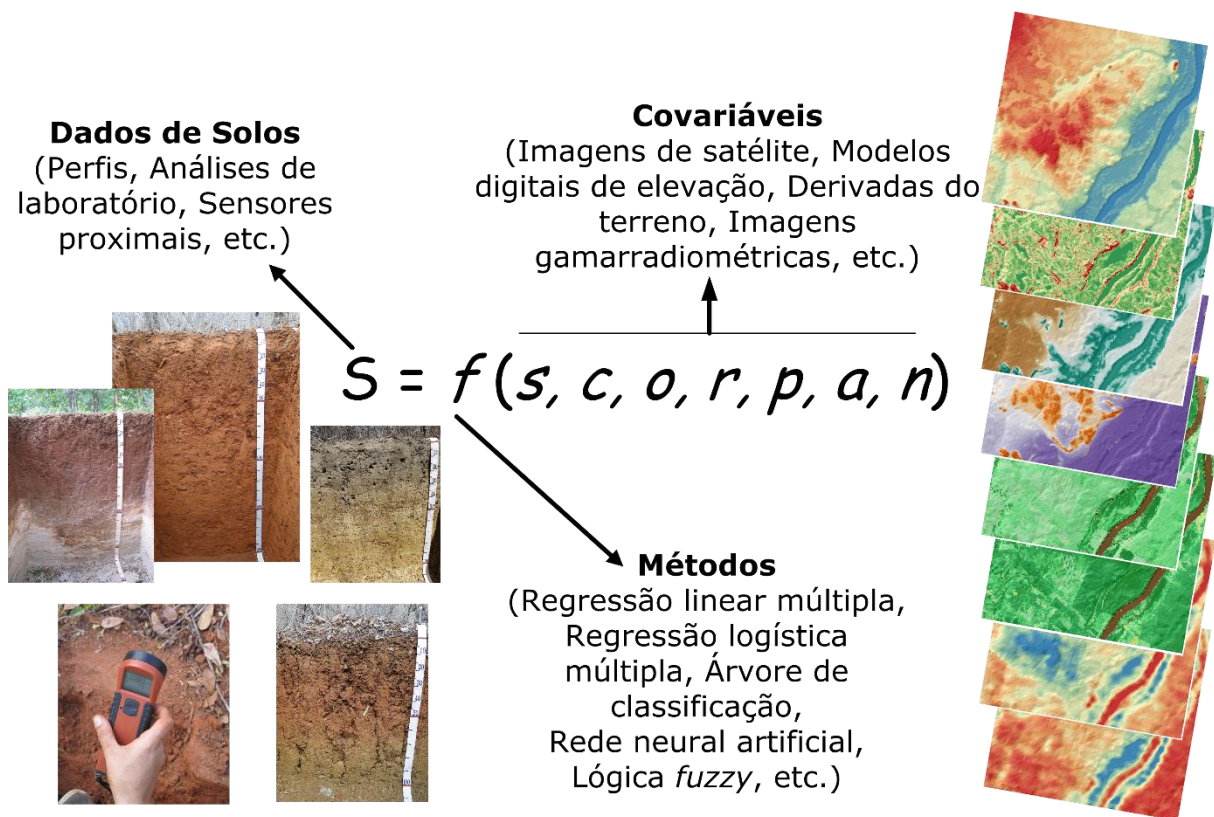


Figura 1. Abordagem metodológica do Mapeamento Digital de Solos, seguindo o modelo conceitual *s.c.o.r.p.a.n.*

A partir do conhecimento dos fatores de formação da equação de Jenny, Dokuchaev e Zakharov (FLORINSKY, 2012), McBratney et al. (2003) formularam o modelo *s.c.o.r.p.a.n.* O objetivo foi descrever quantitativamente as relações empíricas entre os atributos ou classes de solo e os fatores de formação do solo *s.c.o.r.p.a.n.*, onde: *s* - solo; *c* - clima; *o* - organismos, incluindo a atividade humana; *r* - relevo; *p* - material de origem; *a* - tempo; *n* - espaço. Estes fatores são geralmente representados por camadas matriciais ou vetoriais em um Sistema de

Informação Geográfica, constituindo um conjunto de covariáveis ambientais, que podem ser contínuas ou categóricas. Os métodos estatísticos para prever classes ou atributos podem ser: regressões multilíneas, árvores de classificação, redes neurais, florestas aleatórias, entre outros (MCBRATNEY et al., 2003).

Poucas pesquisas em MDS compararam fontes de dados de entrada para estimar **atributos** de solos. Vasques et al. (2012) testaram a influência da resolução espacial dos dados de entrada (ou seja, covariáveis), entre outros fatores, na qualidade de modelos preditores de carbono do solo no estado da Flórida, EUA, e observaram que resoluções espaciais mais detalhadas conduzem a melhores estimativas. Sendo que, neste estudo, a resolução espacial mais detalhada foi 30-m. Analogamente, em uma região pantanosa na Flórida, de aproximadamente 42.000 ha, Kim et al. (2014) utilizaram três tipos de imagens de satélites com diferentes resoluções espaciais a fim de identificar os fatores que controlam a distribuição espacial do fósforo e nitrogênio do solo, além de esclarecer os efeitos das diferentes resoluções espaciais das imagens de satélite nos modelos de inferência, através de procedimento de mineração de dados *random forest* (RF; floresta aleatória). Os autores não observaram distinção com relação às estimativas de erros na predição de nitrogênio e fósforo entre as diferentes resoluções espaciais dos sensores orbitais. No entanto, os padrões de distribuição de fósforo e nitrogênio foram melhor expressos nas resoluções mais detalhadas (Spot 10-m) do que nas menores (Landsat ETM 30-m e MODIS 500-m). Maynard e Johnson (2014) investigaram a influência de fatores relacionados à escala espacial (resolução e tamanho de janela de vizinhança) de atributos do terreno derivados a partir de MDE obtido por sensor LiDAR (*Light Detection and Ranging*) em modelos solo-paisagem. Eles usaram dados de uma bacia hidrográfica florestada de aproximadamente 2300 ha nas montanhas do Oregon, EUA, e testaram resoluções espaciais entre 1 e 50-m e extensões de vizinhança de 3 a 350-m para estimar argila, soma de bases e carbono do solo em duas profundidades (0-20 e 20-50 cm). Observaram forte influência dos fatores de escala espacial nos modelos de predição dos atributos do solo, sendo que os atributos do terreno derivados em escalas menos detalhadas (*coarser scale*) (por exemplo, com extensão de vizinhança de 150 a 250-m) apresentaram forte correlação com os atributos do solo. Além disso, houve apenas pequena diferença na acurácia de predição dos atributos do terreno entre resoluções detalhadas (1-5-

m) e moderadas (10-20-m) dos MDE's, mostrando que não se justifica a utilização do LiDAR, devido aos altos custos e requisitos computacionais e pouca ou nenhuma melhoria no modelo.

Samuel-Rosa et al. (2015) compararam o uso de covariáveis representando diferentes fatores de formação do solo, com diferentes níveis de detalhe, para prever argila, carbono orgânico e capacidade de troca catiônica do solo em um campo experimental com cerca de 2000 ha no sul do Brasil. Concluíram que, em geral, o aumento do detalhamento das covariáveis ambientais apenas resultou em modesta melhora na qualidade das previsões, o que talvez não justifique os gastos adicionais com covariáveis ambientais mais detalhadas. Os autores sugerem que, ao invés de gastos adicionais com covariáveis ambientais, seria preferível investir em coleta de mais observações de solos, por exemplo. Behrens et al. (2014) utilizaram 15 atributos descritivos do MDE (mínimo, máximo, média, desvio padrão, e outros), derivados em múltiplas escalas, para estimar atributos do solo (argila e silte) em duas áreas com distintos processos de formação do solo, uma em Piracicaba, Brasil (30.000 ha), e outra em Rhine-Hesse, Alemanha (115.000 ha). Concluíram que a análise de terreno multiescalar pode ser usada para revelar processos complexos de formação do solo e, assim, interpretá-los mesmo quando o relevo não é o principal condutor da formação do solo. Isto deve-se ao fato de que atributos do terreno, no caso os 15 atributos descritivos do MDE extraídos em várias escalas, podem servir como indicadores ou serem função (reflexo) das condições climáticas ou parentais quando o terreno é considerado não apenas em uma observação pontual, na escala da encosta e em escalas regionais. Assim, em situações, onde apenas um MDE está disponível, ele pode resolver o problema da ausência de dados de covariáveis ambientais, fornecendo dados sobre a assinatura geomórfica e relações solo-paisagem. Maynard e Levi (2017) avaliaram a utilização de NDVI de imagens Landsat TM comparando o efeito da temporalidade na estimativa de classes de textura, utilizando perfis modais e o algoritmo *support vector machine* (SVM), em uma área de aproximadamente 6000 ha no semiárido do Arizona, EUA. Observaram que modelos de previsão utilizando NDVI de imagens hipertemporais (627 imagens) possibilitaram prever classes de textura com maior acurácia do que os modelos que utilizaram imagens multitemporais (28 imagens), bitemporais e monotemporal, resultando em um padrão de distribuição espacial que se aproxima dos padrões observados no mapa de solos disponível e o supera quanto ao delineamento das classes texturais do solo.

Da mesma forma, existem poucos trabalhos que avaliaram a influência das fontes de dados de entrada e sua qualidade associada para estimar **classes** de solos. Basicamente, dois tipos de abordagem têm sido testadas, uma baseada em observações de perfis modais de solos (KIM et al., 2012; BRUNGARD et al., 2015;) e outra via desagregação de polígonos de mapas de solos convencionais (GRINAND et al., 2008; BEHRENS et al., 2010; CAVAZZI et al., 2013). Brungard et al. (2015), por exemplo, avaliou 11 algoritmos de modelagem estatística e conjuntos de covariáveis ambientais com diferentes resoluções espaciais a fim de prever classes de solos através de perfis modais em nível de subgrupo taxonômico em três áreas distintas (Utah, Novo México e Wyoming) no semiárido dos EUA. Observaram que RF foi o algoritmo mais bem sucedido na predição de classes de solo, utilizando diferentes fontes de covariáveis em diferentes escalas, sendo que mais de 50% destas foram variáveis hidrológicas relacionadas à umidade do solo potencial (índice de umidade, área de contribuição e inclinação da captação). Behrens et al. (2010) avaliaram atributos do terreno em várias escalas aplicando filtros de média com diferentes tamanhos de janelas (de 3 a 31 pixels) e seleção de atributos a fim de identificar características relevantes dos fatores de formação do solo, utilizando procedimento de desagregação de polígonos e árvores de classificação através do algoritmo *CRUISE*, no sudoeste da Alemanha. Observaram que cada classe de solo é melhor predita usando uma combinação de atributos em diferentes escalas, constatando que a utilização de atributos do terreno em várias escalas é uma importante ferramenta para MDS, pois permite aumentar a acurácia das predições. Cavazzi et al. (2013) avaliaram a influência da escala, ajustando o tamanho do pixel (20 a 260-m) e tamanho da janela (de 3x3 a 21x21 pixels) usada para derivar atributos do terreno a partir do MDE, perfazendo 143 combinações, na predição de classes do solo utilizando procedimento de desagregação de polígonos através de redes neurais e RF, em três áreas na Irlanda. Observaram dois padrões: áreas homogêneas e planas tendem a obter melhores estimativas nas resoluções mais baixas (acima de 140-m) e em todos os tamanhos de janelas testados, enquanto áreas morfologicamente variadas (com mudanças abruptas de declive, por exemplo) obtiveram melhor desempenho nas resoluções espaciais mais detalhadas (30-m) e com pequenos tamanhos de janela. Kim et al. (2012) compararam os efeitos da resolução espacial (10, 30 e 250-m) de três sensores orbitais a fim de prever classes de solos através de perfis modais em nível de série taxonômica, utilizando árvores de classificação através do algoritmo *CART*, na Flórida. Observaram que nos modelos

que utilizaram somente covariáveis ambientais (MDE, imagens de satélite, gamarradiometria) o melhor resultado foi obtido a partir de imagens com menor resolução espacial (Spot 10-m). Já nos modelos usando covariáveis ambientais e profundidade do solo como preditores, o resultado mais acurado foi aquele usando imagens com menor resolução espacial (Modis 250-m). Além disso, os autores constataram que nesta área a variação espacial das séries de solos pode ser explicada, em conjunto, pelo material de origem, variáveis do terreno e índices de vegetação obtidos através de sensores orbitais. Heung et al. (2017) compararam a acurácia das estimativas de classes de solos em nível de grande grupo taxonômico obtidas através de procedimentos de desagregação de polígonos *versus* perfis modais, utilizando nove procedimentos de mineração de dados, na província de British Columbia, Canadá. Observaram que a qualidade das predições foi consistentemente maior nos modelos derivados de polígonos, não importando o procedimento de mineração de dados utilizado, provavelmente devido à maior densidade de dados de treinamento que puderam ser obtidos a partir dos polígonos, o que melhorou a representação das classes de solos.

2.2. ÁRVORES DE CLASSIFICAÇÃO

Segundo McBratney et al. (2003), os métodos estatísticos para prever classes ou atributos podem ser: regressões multilíneas, árvores de classificação, redes neurais, florestas aleatórias, entre outros. O método árvore de classificação (AC) foi escolhido por ser amplamente utilizado em trabalhos de estimativa de classes de solos (BEHRENS et al., 2010; SCHMIDT et al., 2008; HANSEN et al., 2009; JAFARI et al., 2014; MORAN; BUI, 2002; BUI; MORAN, 2003; MINASNY; MCBRATNEY, 2007; MENDONÇA-SANTOS et al., 2008; TAGHIZADEH-MEHRJARDI et al., 2015), inclusive obtendo bons resultados comparado com outros métodos de mineração de dados (BRUNGARD et al., 2015; TAGHIZADEH-MEHRJARDI et al., 2015; HEUNG et al., 2016).

Métodos de AC se fundamentam na partição recursiva de conjuntos homogêneos (HASTIE et al., 2013), que são divididos baseados em sentenças do tipo “se-então” (*if-then*) geradas a partir de um conjunto de variáveis preditoras (BREIMAN et al., 1984). Cada árvore de classificação possui nós, cada um representando uma covariável ambiental, ramos, que representam o valor das covariáveis ambientais, e folhas, cada uma representando uma classe de solo predita (**Figura 2**; TAGHIZADEH-MEHRJARDI et al., 2015). O método de AC é flexível para lidar com covariáveis preditoras contínuas ou categóricas e não requer pressupostos de distribuição normal nas covariáveis preditoras contínuas, entre outros (HASTIE et al., 2013).

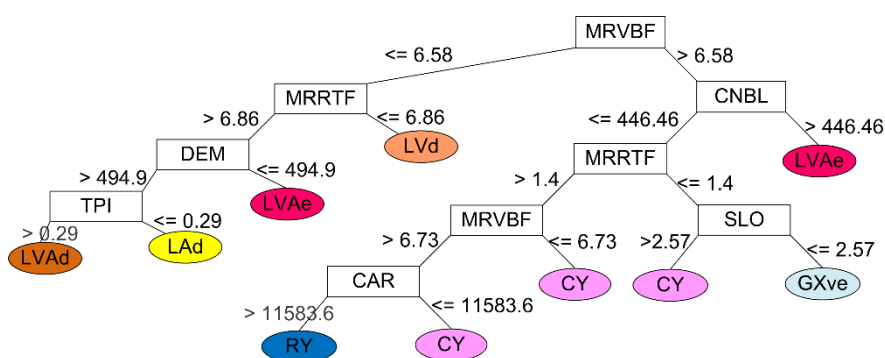


Figura 2. Exemplo de árvore de classificação (AC) para estimar classes de solos.

Métodos de AC frequentemente utilizados na ciência do solo incluem *Classification and Regression Tree* (CART; BREIMAN et al., 1984), o C5 (QUINLAN, 1993) e a *Random Forest* (RF; BREIMAN, 2001). O algoritmo CART possibilita classificar, ou seja, prever categorias (por exemplo, classes de solos) e prever variáveis contínuas através de árvores de regressão, tais

como atributos do solo. O algoritmo C5 é similar ao CART, sendo que possibilita apenas a predição de categorias (por exemplo, classes de solos). Segundo Hastie et al. (2013), a diferença mais marcante destes é a possibilidade do C5 derivar um conjunto de regras simplificadas, que podem ser mais atrativas ao usuário.

Shamy e Dheeba (2016) compararam cinco algoritmos de AC (QUEST, GUIDE, CRUISE, C5 e CART) através de revisão de literatura para avaliar pontos fortes e fracos dos algoritmos. Entre os algoritmos avaliados, as maiores acurácias nas predições foram obtidas através do algoritmo GUIDE e as piores acurácias através do CART. Além disso, entre os algoritmos comparados, o CART foi o que resultou em um menor número de sentenças “se-então” (*if-then*) geradas a partir de um conjunto de variáveis preditoras, enquanto o C5 foi o que teve mais sentenças “se-então” e sempre apresentou processamento computacional mais rápido que os demais algoritmos avaliados. Os algoritmos CART e C5 são considerados modelos de aprendizado simples e o RF é considerado como um conjunto de modelos de aprendizado (*ensemble-model learners*; TAGHIZADEH-MEHRJARDI et al., 2015; HEUNG et al., 2016; HEUNG et al., 2017). O algoritmo RF é conceitualmente similar aos algoritmos de AC, no entanto múltiplas árvores são treinadas e os resultados são baseados na agregação dos resultados de árvores simples (BREIMAN, 2001).

O algoritmo C5 (QUINLAN, 1993) foi utilizado em alguns trabalhos para predizer classes de solos (MORAN; BUI, 2002; BUI; MORAN, 2003; MINASNY; MCBRATNEY, 2007; MENDONÇA-SANTOS et al., 2008; TAGHIZADEH-MEHRJARDI et al., 2015; DART et al., 2010; DART et al., 2015). Inclusive, apresentou melhor desempenho quando comparado a outros métodos de mineração de dados para estimar família de solos em ambiente semiárido (TAGHIZADEH-MEHRJARDI et al., 2015).

2.3. GAMARRADIOMETRIA

A espectroscopia de raios gama aerotransportada é uma técnica de sensoriamento remoto passiva que mede a emissão natural de raios gama por solos e rochas em profundidades de aproximadamente 30 cm (WILFORD; MINTY, 2007). Segundo Wilford (2012), essa profundidade pode variar de acordo com a densidade dos solos. Rawlins et al. (2009) observaram, em solos orgânicos com baixa densidade na Irlanda, a capacidade dessa técnica de medir emissões de profundidades de até um metro. Quase toda a radiação gama detectada na superfície da Terra resulta da decomposição radioativa de três elementos químicos: potássio (^{40}K), tório (^{232}Th) e urânio (^{232}U).

Os levantamentos com espectroscopia de raios gama inicialmente foram utilizados para exploração mineral e estudos sobre intemperismo das rochas e mais recentemente têm sido realizados na ciência do solo para compreender as relações solo-paisagem e mapear solos (WILFORD; MINTY, 2007). Segundo Wilford (2012), a gamarradiometria possibilita inferir sobre propriedades do solo e regolito. Por isso, tem sido aplicada em ciência do solo, principalmente em estudos relacionados aos fatores de formação do solo, como o material de origem. Segundo Wilford e Minty (2007), a concentração de potássio (K) diminui com o aumento do intemperismo. Contudo, a concentração dos elementos urânio (U) e tório (Th) aumenta em solos altamente intemperizados. A **Figura 3** apresenta dois exemplos de covariáveis gamarradiométricas utilizadas no atual trabalho.

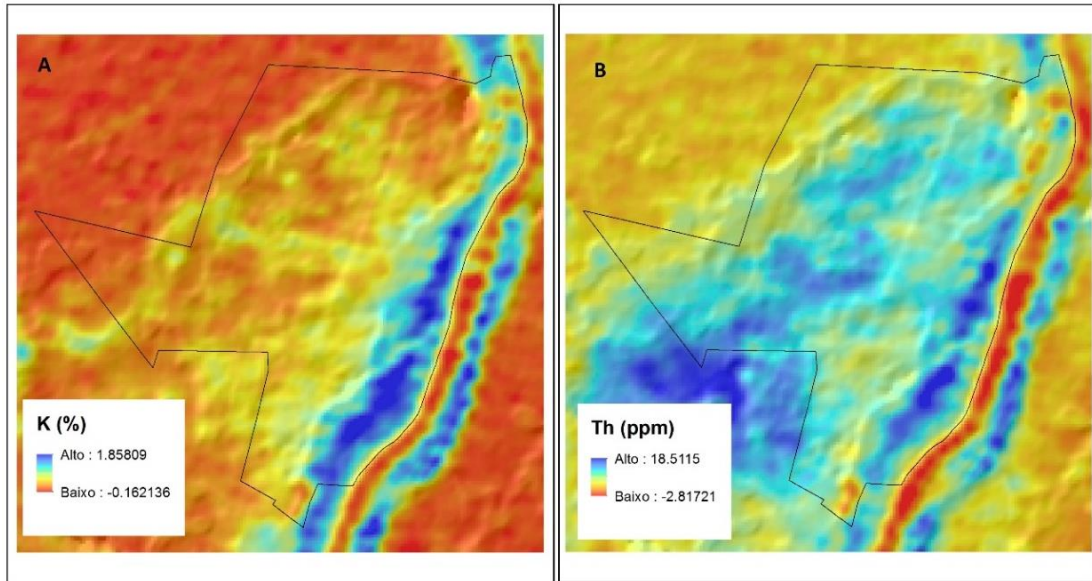


Figura 3. Dois exemplos de covariáveis gamarradiométricas utilizadas. A) Concentração de K (%); B) Concentração de Th (ppm).

3. ÁREA DE ESTUDO

3.1. CONTEXTUALIZAÇÃO

O presente estudo deu continuidade ao projeto MDS-Caatinga (“Mapeamento digital de classes e atributos de solo: um estudo de caso no Bioma Caatinga, região norte de Minas Gerais”), coordenado pelo pesquisador da Embrapa Maurício Rizzato Coelho. O projeto MDS-Caatinga disponibilizou grande número de dados, sendo 424 observações de solos (classificadas e com resultados analíticos), um mapa de solos semidetalhado (escala 1:30.000) e um MDE de alta resolução espacial (1-m, reamostrado para 10-m), o que possibilitou testar procedimentos de MDS para mapear classes de solos pelo presente trabalho. O referido projeto está inserido em uma rede de pesquisa internacional, denominada “*Tropi-dry: Human, ecological and biophysical dimensions of tropical dry forests*”, que envolve pesquisadores de 11 países (Argentina, Bolívia, Brasil, Canadá, Chile, Costa Rica, Cuba, Equador, EUA, México e Venezuela) e realiza estudos comparativos de uso do solo nas regiões de FTS em diversos ecossistemas das Américas. A rede de pesquisas é financiada pelo *Inter-American Institute of Global Change Research* (IAI) desde 2007.

Segundo Murphy e Lugo (1986), 42% das florestas intertropicais no mundo são consideradas FTS, sendo que mais da metade (54,2%) destes remanescentes ocorrem na América do Sul (MILES et al., 2006). O termo FTS tem sido empregado de forma ampla para definir florestas estacionais decíduais, que perdem mais de 50% das folhas de suas árvores durante uma parte do ano (ESPÍRITO-SANTO et al., 2008) e apresentam temperatura média anual de 25°C e precipitação anual média entre 700 e 2000 mm, com pelo menos três meses secos (precipitação ≤ 100 mm/mês) por ano (SÁNCHEZ-AZOFEIFA et al., 2005).

Apesar de sua riqueza de recursos naturais, as FTS são os ecossistemas mais ameaçados e desconhecidos do planeta. Segundo Sánchez-Azofeita et al. (2003), na América Latina e Caribe as FTS são convertidas em agricultura ou pasto a taxas superiores do que a de florestas úmidas, sendo a extração de lenha para a produção de carvão vegetal importante causa de desmatamento local (MURPHY; LUGO, 1986). Apesar disso, do total de 2300 artigos publicados sobre ambientes tropicais desde 1945, apenas 14% destes referiam-se a florestas secas (SÁNCHEZ-AZOFEIFA et al., 2005).

No Brasil, as FTS são classificadas como florestas estacionais decíduais, correspondem a 2,5% do território nacional (**Figura 4**), são interioranas e determinadas pelo clima seco e

sazonal. Essas florestas estão distribuídas por todo o país, de norte a sul, com padrão “espalhado” devido ao último período de glaciação. As FTS do norte de Minas Gerais estão em uma ampla faixa de transição entre os biomas Cerrado, Caatinga e Mata Atlântica. Estas possuem espécies de Cerrado e Caatinga devido à proximidade geográfica desses biomas, mas não o suficiente para que estas florestas sejam pertencentes à Caatinga ou ao Cerrado, exclusivamente. No Brasil, as florestas estacionais decíduais são protegidas pela Lei da Mata Atlântica (Lei Federal nº 11.428 de 2006), mas por pressão de setores agropecuários mineiros foi feita uma alteração na Lei Estadual (MG) nº 17.353 de 2008, que aproxima as FTS ao bioma Caatinga, o que permitirá a supressão de 70% da área atualmente protegida na região Norte de Minas Gerais.

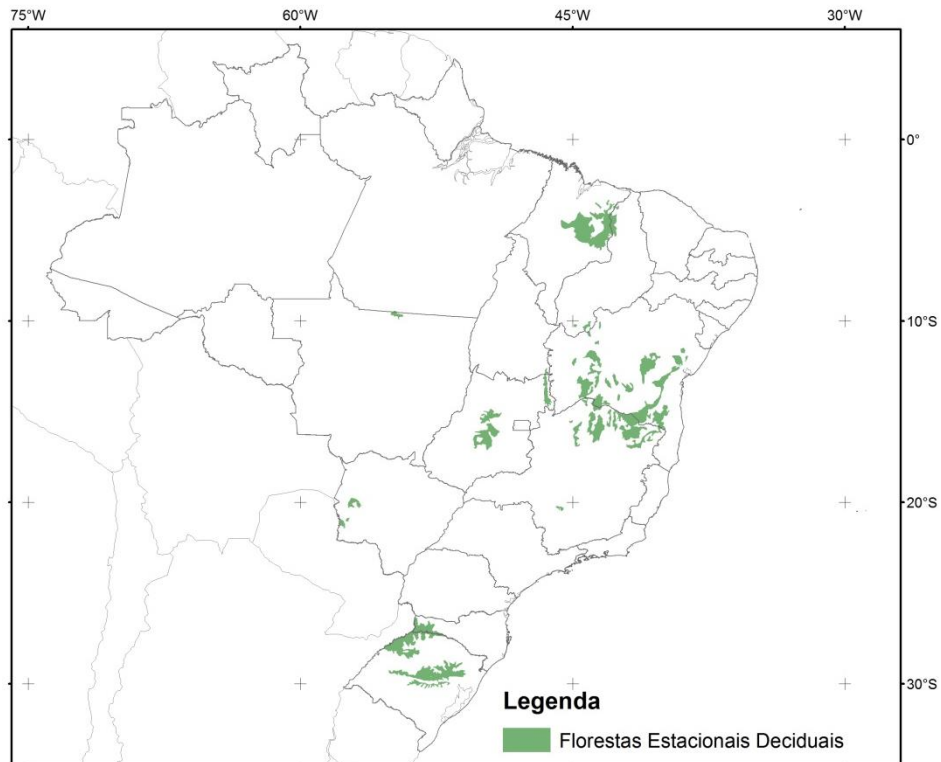


Figura 4. Localização das Florestas Tropicais Secas no Brasil (IBGE, 2006).

3.2. LOCALIZAÇÃO E CARACTERÍSTICAS

O presente estudo foi realizado no Parque Estadual da Mata Seca (PEMS), situado entre as latitudes 14°56'37" e 14°48'52" Sul e longitudes 43°55'49" e 44°03'41" Oeste, no município de Manga no norte do estado de Minas Gerais, com área de aproximadamente 102 km² (Figura 5A e B). O PEMS foi criado pelo Decreto Estadual (MG) n° 41.479 de 2000, com a finalidade de proteger a fauna, a flora e as nascentes dos rios e córregos da região, além de criar condições ao desenvolvimento de pesquisas e estudos. Porém, antes da criação do Parque, houve extração de madeira para lenha e carvão vegetal (ainda existe presença de fornos de carvão), agricultura com dois pivôs centrais desativados em 2009, cuja presença ainda pode ser observada através de imagens orbitais, pecuária até o final de 2009, além da presença de áreas de extração de calcário e brita. Mesmo com a criação do Parque, ainda existe pressão de caça e pesca.

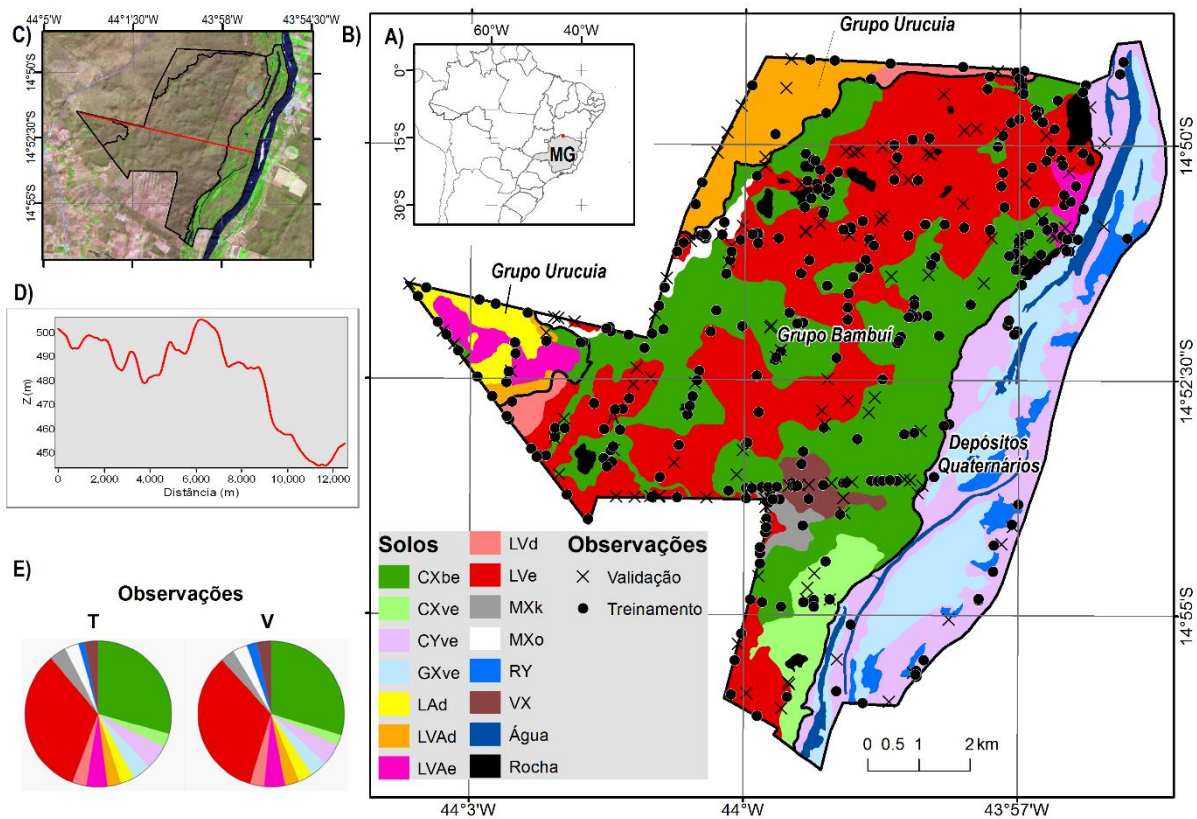


Figura 5. Área de estudo: A) Parque Estadual da Mata Seca (PEMS) no norte do Estado de Minas Gerais; B) Mapa de solos convencional do PEMS (COELHO et al., 2013) com legenda agrupada em nível de grande grupo, mostrando os domínios geológicos do PEMS e a localização das observações de solos; C) Imagem Landsat 8 OLI (02/07/2014, composição RGB 6,5,4) mostrando o traçado do perfil topográfico cruzando os três domínios geológicos do PEMS; D) Perfil topográfico através dos três domínios geológicos do PEMS; E) Proporção do conjunto de observações de treinamento (T) e validação (V).

O clima é semiárido, classificado como transicional entre as tipologias clima tropical (Aw) e semiárido quente (BSw), segundo a classificação de Köppen, com temperatura média anual de 24°C e precipitação média anual de 828 mm, com estação seca de maio a setembro (COELHO et al., 2013). A vegetação predominante é caducifólia, além de florestas perenifólias nos depósitos quaternários do Rio São Francisco (IEF, 2007**Erro! Fonte de referência não encontrada.**). A vegetação é sazonalmente regulada pela precipitação, com grande parte das árvores decíduas perdendo suas folhas (90-95%) durante a estação seca (PEZZINI et al., 2014). O relevo do Parque, de acordo as classes de declividade do SiBCS, é plano (64%) e suave ondulado (31%), e as altitudes variam de 440 a 500 m acima do nível do mar (**Figura 5D**).

As principais classes de solos que ocorrem no PEMS, segundo o mapa de solos convencional de Coelho et al. (2013) (**Figura 5B**), são: GLEISSOLOS HÁPLICOS Ta Eutróficos (GXve), NEOSSOLOS FLÚVICOS (RY) e CAMBISSOLOS FLÚVICOS (CY) nos depósitos quaternários do Rio São Francisco, com floresta tropical pluvial perenifólia; na parte oeste do parque sob vegetação denominada de “Carrasco”, de pequeno porte que ocorre nos platôs secos, ocorrem LATOSSOLOS AMARELOS Distróficos (LAd) e LATOSSOLOS VERMELHO-AMARELOS Distróficos (LVAd) de textura média, relacionados aos arenitos do grupo Urucuia; entre os platôs do grupo Urucuia e os depósitos quaternários existe uma área com litologia do Grupo Bambuí (rochas pelítico-carbonáticas), com presença de rochas calcárias, onde ocorrem vegetação dos tipos Caatinga Arbórea Densa e Floresta Estacional Decidual Densa de porte alto (Mata Seca) e os solos eutróficos do Parque, sendo LATOSSOLOS VERMELHO-AMARELOS ou VERMELHOS Eutróficos (LVAe e LVe), CHERNOSSOLOS HÁPLICOS Carbonáticos ou Órticos (MXk e MXo) e CAMBISSOLOS HÁPLICOS Tb Eutróficos (CXbe), havendo também a presença de VERTISSOLOS HÁPLICOS (VX) e GLEISSOLOS HÁPLICOS ou MELÂNICOS (GX e GM).

4. METODOLOGIA

Os procedimentos metodológicos adotados no trabalho são apresentados de forma sintética no fluxograma da **Figura 6**, consistindo em oito etapas. As etapas 1 e 2, descritas no item 4.1 (Amostras de solos), dizem respeito à preparação das observações de solos, generalização das classes a serem preditas em nível de grande grupo taxonômica (SiBCS) e separação do conjunto de observações para treinamento (T) e validação (V) (**Tabela 1**). As etapas 3 e 4, descritas no item 4.2 (Covariáveis ambientais), tratam da preparação e organização das covariáveis ambientais (**Tabela 2**). A etapa 5, descrita no item 4.3 (Modelos e mapas), refere-se à extração das covariáveis para as observações de solos (pontos). As etapas 6, 7 e 8 estão descritas nos itens 4.3 (Modelos e mapas) e 4.4 (Validação) e compreendem o treinamento dos 17 modelos (M) usando AC C5, a validação dos modelos e a produção dos mapas finais de classes de solos preditas, destacando-se que os procedimentos foram aplicados aos 17 modelos elaborados (**Tabela 3**). Cabe ressaltar que este fluxo está partindo do pressuposto de que já foram definidas a área de estudo, descritos os perfis de solos, coletadas e analisadas as amostras de solos, e que todas as covariáveis necessárias estão disponíveis para o desenvolvimento das etapas posteriores.

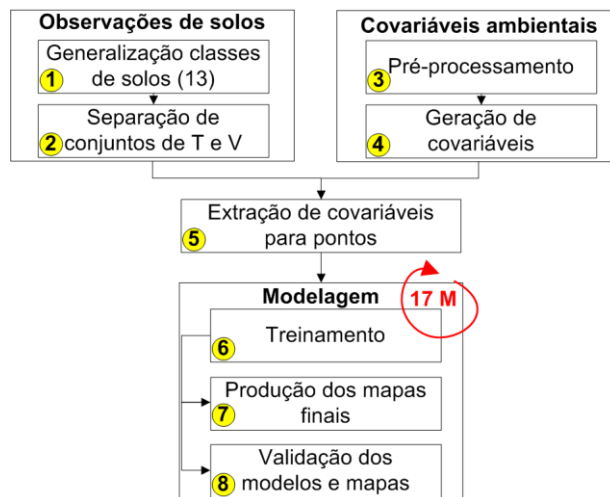


Figura 6. Fluxograma do procedimento metodológico usado para obter e validar os 17 modelos de predição e mapas de classe de solos. A seta circular vermelha indica que as etapas de “Modelagem” foram executadas para todos os 17 modelos (**Tabela 3**).

4.1. AMOSTRAS DE SOLOS

As observações de campo consistiram na coleta e descrição de perfis de solos em trincheiras (totalizando 44 perfis com profundidades de até 2 m), amostras obtidas por tradagens e minitrincheiras (95 tradagens e 83 minitrincheiras com profundidades de até 1 m), e observações extras (97) contendo a classificação expedita do solo, sem coleta de amostras. Os pontos amostrais (perfis, tradagens, minitrincheiras e observações extras) foram alocados por meio do procedimento de caminhamento livre, totalizando 319 observações, sendo o objetivo destas auxiliar na delimitação e definição das unidades de mapeamento do mapa de solos convencional (COELHO et al., 2013). Um conjunto de 41 observações foram alocadas através de procedimento *conditioned Latin hypercube sampling* (cLHS; MINASNY; MCBRATNEY, 2006), que possibilita eficiente amostragem através de análise dados auxiliares (variáveis ambientais), sendo eles o NDVI e uso e cobertura da terra, ambos obtidos a partir de imagem QuickBird (04/07/2007). O último conjunto de observações foi alocado usando amostragem aleatória estratificada dentro de uma máscara ao redor de transetos cortando locais até então pouco inventariados. Os estratos foram definidos por meio de uma combinação de fatores, densidade observações e heterogeneidade ambiental, conforme detalhado em Vasques et al. (2016), porém, devido à dificuldade de acesso na área dos depósitos quaternários, esta foi excluída do procedimento de coleta estratificado. Desta forma, foram realizadas 64 observações em minitrincheiras. Esses procedimentos amostrais totalizaram **424** observações, classificadas no nível de subgrupo, de acordo com o SiBCS, as quais são mostradas na **Figura 5B**.

O conjunto de 424 observações de solos foi dividido aleatoriamente em dois conjuntos de dados, treinamento com 70% (**296** observações) para predição de classes de solos através do procedimento de AC, e um com 30% (**128** observações) para validação das predições. Para garantir a proporção de classes de solos nos conjuntos, treinamento e validação, usou-se seleção aleatória estratificada por grande grupo taxonômico (**Figura 5E; Tabela 1**) através do programa de estatística da SAS JMP® versão 12. Classes de solos (ao nível de grande grupo) com menos de três observações no conjunto de validação foram agrupadas a classes de solos similares, considerando os componentes da unidade de mapeamento do mapa de solos convencional (COELHO et al., 2013), e três classes foram agrupadas em nível de subordem (**Tabela 1**).

Tabela 1. Classes de solos agrupadas em nível de grande grupo, de acordo com o Sistema Brasileiro de Classificação de Solos, nos conjuntos de treinamento (T) e validação (V).

Símbolos	Classes de solos (grandes grupos e subordens)	Observações	
		T	V
CXbe	CAMBISSOLOS HÁPLICOS Tb Eutróficos	87	38
CXve	CAMBISSOLOS HÁPLICOS Ta Eutróficos	9	3
CY	CAMBISSOLOS FLÚVICOS	16	6
GXve	GLEISSOLOS HÁPLICOS Ta Eutróficos	12	5
LAd	LATOSSOLOS AMARELOS Distróficos	9	4
LVAAd	LATOSSOLOS VERMELHO-AMARELOS Distróficos	9	4
LVAe	LATOSSOLOS VERMELHO-AMARELOS Eutróficos	14	6
LVd	LATOSSOLOS VERMELHOS Distróficos	9	4
LVe	LATOSSOLOS VERMELHOS Eutróficos	98	43
MXk	CHERNOSSOLOS HÁPLICOS Carbonáticos	11	4
MXo	CHERNOSSOLOS HÁPLICOS Órticos	9	4
RY	NEOSSOLOS FLÚVICOS	4	3
VX	VERTISSOLOS HÁPLICOS	9	4
TOTAL		296	128

4.2. COVARIÁVEIS AMBIENTAIS

Três tipos de covariáveis ambientais, segundo o modelo *s.c.o.r.p.a.n* (MCBRATNEY et al., 2003), foram avaliados para estimar classes de solos, sendo eles os fatores de formação do solo organismos (*o*), relevo (*r*) e material de origem (*p*). Foram testados, respectivamente: imagens de satélite e índices de vegetação como fator organismos, neste trabalho denominado *sat*; MDE e derivadas do terreno como fator relevo, neste trabalho denominados *topo*; e imagens gamarradiométricas como fator material de origem, neste trabalho denominados *parent*.

Três conjuntos de covariáveis ambientais foram preparados em diferentes resoluções espaciais (**Figura 7**), denominadas: alta (*pixel* de 10-m), média (30-m) e baixa (90-m). Os dados de alta resolução espacial foram obtidos de imagens de satélite RapidEye e Ikonos. Os dados de média resolução espacial foram obtidos do MDE da *Shuttle Radar Topography Mission* (SRTM) com tamanho de *pixel* de 1 arco de segundo (MDEs1; *pixel* com cerca de 30 m) e imagens de satélite Landsat 8 (L8) *Operational Land Imager* (OLI). Os dados de baixa resolução espacial foram obtidos do MDE SRTM com 3 arcos de segundos (MDEs3; cerca de 90 m) e de aerolevanteamento gamarradiométrico (125-m). Todas as variáveis ambientais foram ajustadas para projeção Universal Transversa de Mercator (UTM), fuso 23 sul, no sistema de referência *World Geodetic System* (WGS) de 1984.

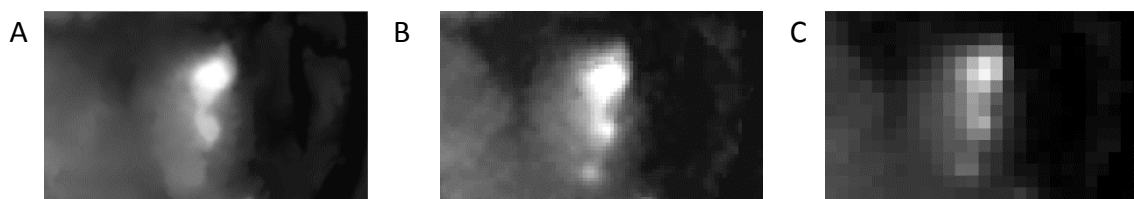


Figura 7. Exemplo do modelo digital de elevação obtido de imagens Ikonos (MDE-I), reamostrado para as três resoluções espaciais utilizadas, sendo: A) 10-m; B) 30m; e C) 90-m.

No conjunto de dados de alta resolução espacial (10-m), o MDE foi obtido a partir de um par estereoscópico ortorretificado de imagens do satélite Ikonos, com 1-m de resolução espacial (MDE-I). Após pareamento das imagens e extração do MDE, usando o *software* Erdas®, foram extraídas curvas de nível com espaçamento vertical de 5-m. Em seguida, as curvas de nível foram interpoladas para obter um MDE hidrológicamente consistente, com resolução espacial de 5-m, por meio da ferramenta *Topo to Raster* do programa ArcGIS®,

usando o algoritmo proposto por Hutchinson (1989). O MDE-I resultante foi reamostrado para 10 m através do procedimento *Resampling* no programa Saga GIS®.

Para preparar os MDE's para geração dos 17 modelos de AC, primeiramente, tanto o MDE-I como o MDEs1 e o MDEs3, foram submetidos ao preenchimento de depressões espúrias, sendo utilizada a ferramenta *Fill Sinks* (PLANCHON; DARBOUX, 2002) no programa Saga GIS®. Em seguida, o MDE-I foi reamostrado para 30-m e 90-m, e o MDEs1 foi reamostrado para 90-m, perfazendo, ao final, seis MDE's (um com 10 m, dois com 30 m e três com 90 m de resolução). Isso possibilitou comparar fontes e resoluções espaciais de MDE's, respectivamente, como covariáveis.

A partir dos MDE's gerados, foram obtidas 25 derivadas do terreno através do programa Saga GIS®, descritas abaixo.

1. *Aspect* (ASP) ou orientação de vertentes, segundo Valeriano (2008) representa a medida do ângulo horizontal da direção esperada dos fluxos, expressa em relação ao norte geográfico (com valor 0°), crescendo no sentido horário. Calculado através do método de nove parâmetros e polinômio de 2ª ordem (ZEVENBERGEN; THORNE, 1987) e expressa em graus.
2. *Catchment area* (CAR) ou área de contribuição da bacia, segundo Olaya (2009) representa a magnitude dos fluxos.
3. *Channel network base level* (CNBL) ou nível de base dos canais de drenagem.
4. *Convergence index* (CVI), índice de convergência ou divergência em relação ao escoamento superficial, similar à curvatura plana, porém mais suave, onde valores negativos representam superfícies convergentes, enquanto que os positivos as superfícies divergentes (QGIS, 2016).
5. *Diffuse insolation* (DIFIN) ou insolação difusa.
6. *Direct insolation* (DRIN) ou insolação direta.
7. *LS-factor* (LSF), componente LS da equação universal de perda de solo, sendo utilizada para estimar o efeito da topografia na erosão, calculado através do método de Moore et al. (1991).
8. *Mass balance index* (MBI), segundo Böhner e Selige (2006) é útil para especificar padrões de solos em superfícies convexas na parte superior da encosta. Calculado através do método descrito por Möller et al. (2008), onde valores negativos

representam superfícies convexas na parte superior da encosta, ou seja, áreas de deposição, como depressões e planícies de inundação, valores positivos representam superfícies de erosão, como as encostas, e valores próximos de zero um balanço entre deposição e erosão.

9. Modelo digital de elevação (MDE), representa a altitude acima do nível do mar, após procedimento de redução de depressões espúrias, expresso em metros.
10. *Multiresolution index of ridge top flatness* (MRRTF), índice multirresolução de suavidade de áreas elevadas (GALLANT; DOWILING, 2003).
11. *Multiresolution index of valley bottom flatness* (MRVBF), índice multirresolução de suavidade de áreas de deposição (GALLANT; DOWILING, 2003).
12. *Mid slope position* (MSP), segundo Dietrich e Böhner (2008) este parâmetro atribui valores 0 em posições de meia encosta, enquanto que nos locais onde ocorrem as maiores variações de altitude, da meia encosta para os vales ou para o cume, são atribuídos valores próximos de 1.
13. *Profile curvature* (PFCV) ou curvatura vertical, refere-se à forma convexa ou côncava do terreno quando analisado em perfil (VALERIANO, 2008). Também é entendida como taxa de variação da declividade (BURROUGH; MCDONNELL, 2004).
14. *Plan curvature* (PLCV) ou curvatura horizontal, refere-se ao caráter divergente ou convergente dos fluxos de matéria no terreno, analisado em projeção horizontal (VALERIANO, 2008). Também é compreendida como taxa de variação da orientação da vertente (BURROUGH; MCDONNELL, 2004).
15. *Relative slope position* (RSP) ou posição relativa da encosta, segundo Bock et al. (2007) é um parâmetro de terreno que combina dois outros, altitude acima das linhas de canais e altitude abaixo das linhas de cumeeira. Os valores resultantes variam de 0, para as posições baixas da encosta (linhas de canal), a 1, para as posições ascendentes (linhas de cumeeira).
16. *Slope height* (SHT), definido como a distância vertical a partir da base da encosta até o topo, ou linha de interseção entre dois planos de inclinação (ROCSCIENCE INC., 2016).
17. *Slope length* (SLG) ou comprimento de rampa.

18. *Slope gradient* (SLO) ou declividade, representa a taxa de variação da elevação, expressa em porcentagem.
19. *Total insolation* (TOIN) ou insolação total.
20. *Topographic position Index* (TPI) ou índice de posição topográfica, índice calculado pela diferença entre a elevação e a média da elevação (GUISAN et al., 1999).
21. *Terrain ruggedness index* (TRI), índice que representa a heterogeneidade topográfica (RILEY et al., 1999).
22. *Terrain surface texture* (TST), mede a frequência de picos e depressões em um MDE sobre um determinado tamanho de janela, representando a textura do terreno (IWAHASHI; PIKE, 2007).
23. *Topographic wetness index* (TWI) ou índice de umidade do terreno, segundo Böhner e Selige (2006) representa a superfície e subsuperfície do escoamento superficial, sendo calculada em função da área de contribuição da bacia (CAR) e declividade do terreno (SLO).
24. *Vertical distance to channel network* (VDCN) ou distância vertical até o canal de drenagem.
25. *Valley depth* (VDP) ou profundidade do vale.

Tabela 2. Conjuntos de covariáveis ambientais utilizadas para prever classes de solos.

Covariáveis ambientais	Resoluções espaciais	Abreviações	Descrições
sat, a partir de imagens de satélite multiespectrais (organismos)	5-m	REDryB1 - REDryB5	Reflectância de superfície da banda espectral do sensor RapidEye no período seco (08/2012)
		REWetB1 - REWetB5	Reflectância de superfície da banda espectral do sensor RapidEye no período úmido (05/2013)
		L8DryB1 - L8DryB7	Reflectância de superfície da banda espectral do sensor Landsat 8 OLI no período seco (07/2014)
	30-m	L8WetB1 - L8WetB7	Reflectância de superfície da banda espectral do sensor Landsat 8 OLI no período úmido (03/2014)
		NDVI_RED e REW NDVI_L8D e L8W	<i>Normalized difference vegetation index</i> (RE=RapidEye, L8=Landsat 8, D=Dry e W=Wet)
		RVI_RED e REW RVI_L8D e L8W	<i>Ratio vegetation index</i>
		SAVI_RED e REW SAVI_L8D e L8W	<i>Soil-adjusted vegetation index</i>
		ASP	<i>Aspect</i>
		CAR	<i>Catchment area</i>
CNBL	<i>Channel network base level</i>		
CVI	<i>Convergence index</i>		
MDE	Elevação (m) sem depressões espúrias		
DIFIN	<i>Diffuse insolation</i>		
DRIN	<i>Direct insolation</i>		
LSF	<i>LS-factor</i>		
MBI	<i>Mass balance index</i>		
MRRTF	<i>Multiresolution index of ridge top flatness</i>		
MRVBF	<i>Multiresolution Index of valley bottom flatness</i>		
MSP	<i>Mid slope position</i>		
PFCV	<i>Profile curvature</i>		
PLCV	<i>Plan curvature</i>		
RSP	<i>Relative slope position</i>		
SHT	<i>Slope height</i>		
SLG	<i>Slope length</i>		
SLO	<i>Slope gradient</i>		
TOIN	<i>Total insolation</i>		
TPI	<i>Topographic position index</i>		
TRI	<i>Terrain ruggedness index</i>		
TST	<i>Terrain surface texture</i>		
TWI	<i>Topographic wetness index</i>		
VDCN	<i>Vertical distance to channel network</i>		
VDP	<i>Valley depth</i>		
parent, a partir de imagens de aerolevantamento gamarradiométrico e magnetométrico (material de origem)	125-m	1DV	1ª derivada vertical do campo magnético
		CT	Taxa de exposição da contagem total, em µR/h
		K	Potássio, em %
		Mag	Campo magnético total, em nT/m
		Sinal	Sinal analítico do campo magnético total, em nT/m
		Th	Tório, em ppm
		U	Urânio, em ppm
		ThK	Razões Th/K, U/K e U/Th, respectivamente
		UK	
		UTh	

Duas imagens RapidEye (RE) foram selecionadas, uma no período seco (REdry, 05/08/2012) e outra no período úmido (REwet, 11/05/2013), a partir do sítio do Geo Catálogo do Ministério do Meio Ambiente (MMA; BRASIL, 2014). Para auxiliar na seleção de imagens representativas para os dois períodos, foi consultado o sítio do PROCLIMA (Programa de Monitoramento Climático em Tempo Real da Região Nordeste; INPE, 2015), que proporciona informação do balanço hídrico para região de atuação da Superintendência do Desenvolvimento do Nordeste (nordeste e semiárido). As imagens RE utilizadas são ortorretificadas, possuem resolução radiométrica de 16 bits, cinco bandas espectrais (1=*blue* em 440-510 nm, 2=*green* em 520-590nm, 3=*red* em 630-685 nm, 4=*red edge* em 690-730 nm, e 5=*near infrared* em 760-850 nm) e resolução espacial reamostrada de 6,5 para 5 m.

A fim de minimizar os efeitos atmosféricos nas imagens RE, realizou-se o procedimento de correção atmosférica para trabalhar com valores de reflectância de superfície terrestre. Para tal, utilizou-se o pacote FLAASH[®] no programa ENVI[®] para converter valores digitais para radiância espectral nas imagens e, posteriormente, reflectância de superfície. O posicionamento das imagens RE foi verificado usando como verdade de campo imagens Ikonos ortorretificadas, sendo que não se observou a necessidade de ajustes de geometria. As imagens RE foram reamostradas de 5-m para 10-m através do interpolador bilinear no programa Saga GIS[®] e, em seguida, foram calculados três índices de vegetação, descritos a seguir.

1. *Normalized difference vegetation index* (NDVI) ou índice de vegetação por diferença normalizada (**Equação 1**), segundo Silleos et al. (2006), o NDVI foi desenvolvido para separar a vegetação verde do fundo brilhante do solo em imagens Landsat MSS, minimizando os efeitos topográficos. O índice varia de -1 a +1, onde valores baixos ocorrem em áreas com afloramento de rocha e areia, valores médios ocorrem em áreas com vegetação esparsa, como pastagem, e valores alto em áreas com vegetação densa (USGS, 2016).
2. *Ratio vegetation index* (RVI) ou índice de vegetação por razão de bandas (**Equação 2**), segundo Silleos et al. (2006) o resultado apresenta claramente o contraste entre as bandas no vermelho (*R*) e infravermelho próximo (*NIR*), com pixels com vegetação apresentando valores elevados do índice, pela combinação de baixo valor de reflectância da banda *R* (absorção de clorofila) e

elevada reflectância no *NIR* (comportamento espectral da vegetação devido à estrutura da folha).

3. *Soil-adjusted vegetation index* (SAVI; Huete, 1988; **Equação 3**), de acordo com Silleos et al. (2006), destina-se a minimizar os efeitos de fundo do solo sobre a vegetação incorporando um fator de ajuste de solo constante (*L*) no numerador e denominador da equação do NDVI.

Equação 1. Normalized difference vegetation index (NDVI)

$$NDVI = \frac{NIR - R}{NIR + R}$$

Onde *NIR* representa a reflectância no infravermelho próximo e *R* representa a reflectância no vermelho.

Equação 2. Ratio vegetation index (RVI)

$$RVI = \frac{R}{NIR}$$

Onde *R* representa a reflectância no vermelho e *NIR* representa a reflectância no infravermelho próximo.

Equação 3. Soil-adjusted vegetation index (SAVI)

$$SAVI = \frac{(NIR - R)(1 + L)}{NIR + R + L}$$

Onde *NIR* representa a reflectância no infravermelho próximo, *R* representa a reflectância no vermelho e *L* é um fator de ajuste do solo constante (0,5).

O conjunto de dados de média resolução espacial (30-m) foi composto pelas mesmas variáveis de alta resolução (MDE-I, REWet e REDry), porém reamostradas de 10 para 30-m por meio de interpolação *bicubic spline* através do programa Saga GIS®. Além destas, fez-se uso de variáveis obtidas de imagens com resolução espacial de 30-m, sendo o MDE SRTM com 1 arco de segundo (MDEs1) e imagens L8 nos períodos seco e úmido, respectivamente. O MDEs1 foi obtido através do sítio *Earth Explorer* do Serviço Geológico Americano (USGS, 2015) e suas 25 derivadas do terreno (**Tabela 2**) foram obtidas no programa Saga GIS®.

Duas imagens L8, *path* 219 e *row* 070, foram selecionadas para os períodos seco (L8Dry, 02/07/2014) e úmido (L8Wet, 12/03/2014), através do sítio *Earth Explorer* do Serviço Geológico Americano (USGS, 2015). A seleção do período das imagens foi feita através de consulta do balanço hídrico da região (INPE, 2015). As imagens L8 possuem 11 bandas espectrais, porém foram usadas apenas as bandas 1 a 7 (1=*coastal aerosol* em 430-450 nm, 2=*blue* em 450-510 nm, 3=*green* em 530-590 nm, 4=*red* em 640-670 nm, 5=*near infrared* em 850-880 nm, 6=*short-wave infrared 1* em 1570-1650 nm), e 7= *short-wave infrared 2* em 2110-2290 nm). Possuem resolução radiométrica de 12 bits, resolução espacial de 30 m e são ortorretificadas. Para diminuir os efeitos da atmosfera, foi realizado procedimento de correção atmosférica nas imagens L8Dry e L8Wet através do pacote FLAASH[®] implementado no programa ENVI[®]. O posicionamento das imagens L8 foi ajustado através de registro polinomial de primeira ordem no programa ENVI[®], onde foram utilizadas as imagens RapidEye como referência. Além disso, foram calculados os mesmos índices de vegetação (NDVI, RVI e SAVI) para as imagens L8 (**Tabela 2**).

O conjunto de dados com baixa resolução espacial (90-m) foi composto pelas mesmas variáveis de alta (MDE-I e REWet e REDry) e média (MDEs1, L8Wet e L8Dry) resoluções espaciais, reamostradas de 10 para 30m e de 30 para 90 m por meio de interpolação *bicubic spline* através do programa Saga GIS[®]. Além destas, fez-se uso de imagens com baixa resolução espacial, sendo o MDEs3 e dados gamarradiométricos e magnetométricos com 125 m de resolução. O MDEs3 foi obtido através do sítio *SRTM 90m Digital Elevation Database v4.1* (CGIAR, 2014) e as 25 derivadas do terreno (**Tabela 2**) foram geradas no programa Saga GIS[®]. As imagens gamarradiométricas e magnetométricas foram obtidas a partir de levantamento aéreo (SERVIÇO GEOLÓGICO DO BRASIL, 2009) com 125-m de resolução espacial e reamostradas de 125 para 90 m por meio de interpolação *bicubic spline* através do programa Saga GIS[®], compreendendo 10 planos de informação (variáveis *parent* na **Tabela 2**).

4.3. MODELOS E MAPAS

As classes de solos em nível de grande grupo, de acordo com o SiBCS, foram preditas em função das covariáveis ambientais utilizando o algoritmo de AC C5 (QUINLAN, 1993, 2001). Para isso, foi utilizado o conjunto de dados de treinamento com 296 observações (item 4.1;

Tabela 1). Para isso, foi utilizado o programa C5.0®, sendo que o algoritmo funciona em duas etapas. Primeiro, uma grande árvore cresce; depois, ela é podada, sendo que este processo é aplicado em cada subárvore (TAGHIZADEH-MEHRJARDI et al., 2015). O ajuste da poda (*Pruning CF*) no C5.0® foi mantida como *default* (25%), sendo que valores menores que o *default* aumentam a poda da árvore inicial, enquanto valores maiores resultam em menos poda (QUINLAN, 2001). Além desta, a opção de número mínimo de casos em cada folha foi ajustada para dois casos, porque valores com mais de dois casos podem resultar em uma árvore inicial que se ajusta apenas aproximadamente ao conjunto de dados de treinamento (QUINLAN, 2001).

Modelos foram elaborados para estimar classes de solos no nível categórico de grande grupo. **Seis** modelos foram implementados utilizando somente as variáveis *topo*, sendo três a partir do MDE-I nas três resoluções espaciais (10-m, 30-m e 90-m), respectivamente, dois a partir do MDEs1 em duas resoluções (30-m e 90-m), e um usando o MDEs3 (90-m). **Dez** modelos foram formulados utilizando somente as variáveis *sat*, sendo seis com imagens RE nos períodos úmido (*Wet*) e seco (*Dry*) nas três resoluções (10-m, 30-m e 90-m), respectivamente, e quatro com imagens L8 nos dois períodos (*Wet* e *Dry*) em duas resoluções (30-m e 90-m). **Um** modelo foi preparado utilizando somente as variáveis *parent* (90-m). Dessa forma, foram gerados **17** modelos de predição (**Tabela 3**).

Tabela 3. Modelos, fontes e covariáveis ambientais utilizadas para prever classes de solos.

Tipos	Modelos	Fontes	Variáveis Preditoras
<i>topo</i> (relevo)	M_10m_D_I		
	M_30m_D_I	MDE Ikonos 10-m	
	M_90m_D_I		25 (ASP, CAR, CNBL, CVI, DIFIN, DRIN, LSF, MBI, MDE, MRRTF, MRVBF, MSP, PFCV, PLCV, RSP, SHT, SLG, SLO, TOIN, TPI, TRI, TST, TWI, VDCN, VDP)
	M_30m_D_s1	SRTM com 1 arco de segundo (~30-m)	
	M_90m_D_s1		
	M_90m_D_s3	SRTM com 3 arcos de segundo (~90-m)	
<i>sat</i> (organismos)	M_10m_D_REdry		
	M_30m_D_REdry	RapidEye (5-m) no período seco (<i>dry</i>)	
	M_90m_D_REdry		
	M_10m_D_REwet		RapidEye (5-m) no período úmido (<i>wet</i>)
	M_30m_D_REwet		
	M_90m_D_REwet		
	M_30m_D_L8dry	Landsat 8 (30-m) no período seco (<i>dry</i>)	
	M_90m_D_L8dry		10 (B1, B2, B3, B4, B5, B6, B7, NDVI, RVI, SAVI)
	M_30m_D_L8wet	Landsat 8 (30-m) no período úmido (<i>wet</i>)	
	M_90m_D_L8wet		
<i>parent</i> (material de origem)	M_90m_D_G	Aerolevantamento gamarradiométrico e magnetométrico (125-m)	10 (1DV, CT, K, Mag, Sinal, Th, ThK, U, UK, UTh)

MDE, elevation; SLO, slope gradient; PFCV, profile curvature; PLCV, plan curvature; ASP, aspect; LSF, LS-factor; VDP, valley depth; RSP, relative slope position; MRVBF, multiresolution index of valley bottom flatness; MRRTF, multiresolution index of ridge top flatness; MBI, mass balance index; TPI, topographic position index; TRI, terrain ruggedness index; TST, terrain surface texture; SLG, slope length; SHT, slope height; MSP, mid slope position; DRIN, direct insolation; DIFIN, diffuse insolation; TWI, topographic wetness index; CVI, convergence index; CAR, catchment area; CNBL, channel network base level; 1DV, 1ª derivada vertical do campo magnético total; CT, taxa de exposição da contagem total; K, concentração de potássio; Mag, campo magnético total; Sinal, sinal analítico do campo magnético total; Th, concentração de tório; U, concentração de urânio; ThK, razão Th/K; U/K, razão U/K; UTh, razão U/Th; B1 a B7, bandas espectrais; NDVI, normalized difference vegetation index; RVI, ratio vegetation index; SAVI, soil-adjusted vegetation index.

Para as etapas de treinamento, foram utilizados o conjunto de dados de treinamento com 296 observações e as covariáveis ambientais (**Tabela 2**) extraídas no programa Saga GIS®.

Os mapas de classes de solo foram produzidos a partir dos modelos de predição ajustados, aplicando-se, para cada modelo ajustado, o conjunto de regras de decisão obtido pelo algoritmo C5 em todos os *pixels* que preenchem a área de estudo, nas respectivas resoluções espaciais, de acordo com os modelos (90-m, 30-m e 10-m). As matrizes de 90-m, 30-m e 10-m possuem, respectivamente, 12.380, 111.600 e 1.004.643 *pixels*. Em todos os mapas foram extraídas as áreas correspondentes a corpos d'água e afloramentos rochosos delineadas no mapa de solos convencional (COELHO et al., 2013; **Figura 5B**).

4.4. VALIDAÇÃO

A fim de validar os modelos e mapas de classes de solos ao nível de grande grupo gerados por AC, foi usado um conjunto 128 observações independentes (**Figuras 5; Tabela 1**). Para avaliar a incerteza dos mapas de solos fez-se uso dos erros de classificação, representando a porcentagem de casos classificados erroneamente, e foram geradas matrizes de confusão, onde as colunas representam os dados de referência (observados) e as linhas as predições (classes nos mapas de solos). Segundo Congalton (1991), a matriz de confusão é eficiente para descrever erros de inclusão (comissão) e erros de exclusão (omissão). Com base na matriz de confusão, quatro índices foram calculados (CONGALTON, 1991) a fim de comparar os modelos e mapas, sendo, a acurácia global (AG), acurácia do usuário (AU), acurácia do produtor (AP) e índice KAPPA (\hat{K}), conforme segue:

1. Acurácia global (AG): representa o percentual de observações corretamente classificadas. É calculada dividindo-se o total de observações corretamente classificadas (diagonal principal da matriz de confusão) pelo número total de observações da matriz (**Equação 4**; CONGALTON; GREEN, 2009).
2. Acurácia do produtor (AP): representa quanto uma dada classe foi corretamente classificada. Ela indica a probabilidade da observação de referência ter sido corretamente classificada. É calculada dividindo-se o número total de observações corretamente classificados em uma classe pelo número total dos dados de referência (**Equação 5**; CONGALTON; GREEN, 2009). Essa medida indica a probabilidade do *pixel* de referência ter sido corretamente classificado, é também conhecida como erro de omissão.
3. Acurácia do usuário (AU): representa a probabilidade de uma observação em um mapa efetivamente representar determinada classe no terreno. É calculada dividindo-se o número total de observações corretamente classificadas em uma classe pelo número total de observações classificadas nesta classe (**Equação 6**; CONGALTON; GREEN, 2009), também conhecido como medida comissão.
4. Índice KAPPA (\hat{K}): outra medida para avaliar a acurácia de um classificador, sendo que esse índice faz distinção entre o que foi corretamente classificado por causa do classificador em si e o que foi corretamente classificado por mero acaso. Em seu

cálculo, leva em conta os elementos da diagonal principal e os elementos fora da diagonal principal (**Equação 7**; CONGALTON; GREEN, 2009), ou seja, considera os erros de omissão e comissão.

Equação 4. Acurácia global (AG)

$$AG = \frac{\sum_{i=1}^k n_{ii}}{n}$$

Onde i representa as linhas (predições) da matriz, $\sum_{i=1}^k n_{ii}$ representa o número total de observações corretamente classificadas na diagonal principal da matriz, e n o número total de observações.

Equação 5. Acurácia do produtor (AP)

$$AP = \frac{n_{jj}}{n_{+j}}$$

Onde j representa as colunas (observações, dados de referência) da matriz de confusão, n_{jj} o número de observações corretamente classificadas em uma classe, e n_{+j} o número total de observações de referência daquela classe.

Equação 6. Acurácia do usuário (AU)

$$AU = \frac{n_{ii}}{n_{i+}}$$

Onde i representa as linhas (predições) da matriz, n_{ii} o número de observações corretamente classificadas em uma classe, e n_{i+} o número total de observações classificadas nesta mesma classe.

Equação 7. Índice KAPPA (\hat{K})

$$\hat{K} = \frac{p_o - p_e}{1 - p_e}$$

Onde p_o representa a proporção de amostras corretamente classificadas na diagonal principal da matriz, ou seja, a AG, e p_e representa o acerto por mero acaso. O índice pode variar de -1 a 1.

Para avaliar a performance dos 17 modelos de predição de classes de solos, fez-se uso das faixas de concordância em função do índice KAPPA (\hat{K}), propostas por Landis e Koch (1977; **Tabela 4**).

Tabela 4. Faixas de concordância em função do índice KAPPA (\hat{K}), adaptadas de Landis e Koch (1977).

Índice KAPPA	Faixas de concordância
< 0	Insignificante
0 a 0,20	Fraco
0,21 a 0,40	Justo
0,41 a 0,60	Razoável
0,61 a 0,80	Bom
0,81 a 1	Quase perfeito

O melhor modelo selecionado foi aquele com a maior \hat{K} medido usando o conjunto de dados de validação (128 observações). O mapa de solos convencional (COELHO et al., 2013) foi generalizado ao nível de grande grupo, a fim de permitir uma comparação com os mapas gerados pelos 17 modelos. O mapa de solos convencional, como em Vasques et al. (2016), não foi utilizado como verdade de campo, ou seja, não faz parte da validação dos modelos, mas como uma fonte de informação alternativa com o mesmo objetivo para o mesmo local.

5. RESULTADOS E DISCUSSÃO

5.1. AVALIAÇÃO DOS MODELOS NA ETAPA DE TREINAMENTO

Em geral, os melhores modelos foram os que utilizaram variáveis topográficas (*topo*; **Tabela 5**), embora a área de estudo apresente pouca variação no relevo (predominantemente plano e suave ondulado). Isto pode estar relacionado ao que Behrens et al. (2014) observou: que os atributos do terreno podem revelar complexos processos de formação do solo, mesmo quando este não é o principal condutor do processo de formação do solo, servindo como indicador de condições climáticas ou parentais e fornecendo informações sobre a relação solo-paisagem em múltiplas escalas. Ademais, todos os modelos *topo* e *parent* conseguiram representar todas as 13 classes de solos da área de estudo. No entanto, apenas três modelos *sat* conseguiram estimar todas as classes de solos, sendo dois com imagens no período seco (M10m_D_REdry e M30m_D_L8dry) e um com imagens do período úmido (M90m_D_L8wet) (**Tabela 5**).

Os maiores índices de AG e \hat{K} , na etapa de treinamento, ocorreram nos modelos *topo* e *parent*, e os piores foram encontrados nos modelos *sat*. A AG variou de 88,5 (M10m_D_I) a 91,6% (M90m_D_I) nos modelos *topo*, e de 75,7 (M30m_D_REdry) a 81,4% (M90m_D_L8wet) nos modelos *sat*. Segundo a classificação do índice KAPPA de Landis e Koch (1977; **Tabela 4**), poderiam ser considerados quase perfeitos todos os modelos *topo* e *parent* (\hat{K} de 0,85 a 0,89; **Tabela 5**) e bons todos os modelos *sat* (0,68 a 0,76). Estes resultados confirmam a hipótese de McBratney et al. (2003): que os fatores de formação do solo representados por variáveis ambientais podem estimar classes de solos com tolerável margem de erro. Nossos resultados apresentaram valores semelhantes ou melhores de AG e \hat{K} nos modelos *topo* e *parent* do que os observados em trabalhos semelhantes usando o algoritmo C5, como em: Bui e Moran (2003), em que o valor de AG variou de 53,8 a 79,3% e o do \hat{K} de 0,33 a 0,74; Minasny e McBratney (2007), que encontraram AG de 85,5%; e Mendonça-Santos et al. (2008), com AG de 79 a 88,3%, exceto em relação ao melhor modelo (AG de 97,6%), no qual foi utilizado um mapa de solos como variável preditora.

Tabela 5. Resultados dos modelos ordenados do maior para o menor índice KAPPA (\hat{K}) calculado no conjunto de dados de validação.

Modelos	Tipo	Classes de solos preditas	Treinamento			Validação		
			Erros %	AG	\hat{K}	Erros %	AG	\hat{K}
M90m_D_s3	<i>topo</i>	13 Classes	10,5	89,5	0,87	46,9	53,1	0,40
M90m_D_s1	<i>topo</i>	13 Classes	11,1	88,9	0,86	50,8	49,2	0,34
M30m_D_s1	<i>topo</i>	13 Classes	10,8	89,2	0,86	53,9	46,1	0,30
M90m_D_I	<i>topo</i>	13 Classes	8,4	91,6	0,89	55,5	44,5	0,29
M90m_D_G	<i>parent</i>	13 Classes	11,5	88,5	0,86	56,3	43,8	0,28
M90m_D_L8dry	<i>sat</i>	12 (- RY)	21,6	78,4	0,72	54,7	45,3	0,27
M90m_D_REdry	<i>sat</i>	12 (- RY)	23,0	77,0	0,70	56,3	43,8	0,23
M10m_D_REdry	<i>sat</i>	13 Classes	22,6	77,4	0,70	59,4	40,6	0,22
M30m_D_L8dry	<i>sat</i>	13 Classes	19,6	80,4	0,75	60,9	39,1	0,22
M30m_D_REdry	<i>sat</i>	11 (- LVd e RY)	24,3	75,7	0,68	58,6	41,4	0,21
M30m_D_I	<i>topo</i>	13 Classes	10,5	89,5	0,87	62,5	37,5	0,20
M30m_D_L8wet	<i>sat</i>	12 (- RY)	23,3	76,7	0,70	60,2	39,8	0,19
M90m_D_L8wet	<i>sat</i>	13 Classes	18,6	81,4	0,76	64,1	35,9	0,17
M10m_D_I	<i>topo</i>	13 Classes	11,5	88,5	0,85	64,8	35,2	0,14
M10m_D_REwet	<i>sat</i>	12 (- RY)	22,6	77,4	0,71	68,8	31,3	0,11
M90m_D_REwet	<i>sat</i>	12 (- RY)	24,3	75,9	0,68	68,8	31,3	0,09
M30m_D_REwet	<i>sat</i>	12 (- RY)	24,0	76,0	0,69	71,9	28,1	0,07

Erros %, erros de classificação; \hat{K} , índice KAPPA; AG, acurácia global; RY, NEOSSOLOS FLÚVICOS; LVd, LATOSSOLOS VERMELHOS Distróficos.

5.2. IMPORTÂNCIA DAS VARIÁVEIS AMBIENTAIS, FONTES, RESOLUÇÃO ESPACIAL E PERÍODOS DE AQUISIÇÃO DAS IMAGENS

A melhor resolução espacial para estimar classes de solos com variáveis *topo* foi 90-m (**Tabela 5**). Isso fica evidente quando se observa que, para as fontes de MDE Ikonos (MDE-I) e SRTM 30-m (MDEs1), os melhores resultados foram obtidos usando os MDE's reamostrados para 90-m, respectivamente, o que indica que a variação solo-paisagem no PEMS ocorre a distâncias/escalas próximas de 90-m. Os resultados dos modelos *topo* são consistentes com os resultados de Cavazzi et al. (2013), que analisaram três áreas com relevos diferentes e observaram que, na área com relevo homogêneo, as melhores estimativas de classes de solos ocorreram nas resoluções espaciais mais baixas (acima de 140 m), sendo que o desempenho de validação foi comparável com os nossos melhores resultados, com AG de 45,3 a 53,1%.

Em comparação, não foi possível perceber um padrão claro referente à qual a melhor resolução espacial para os modelos com variáveis *sat*. Apesar de os modelos *sat* com resolução espacial de 90-m apresentaram resultado ligeiramente melhor, eles não conseguiram estimar todas as 13 classes de solos (**Tabela 5**). Por outro lado, ficou claro que os modelos *sat* que utilizaram imagens no período seco apresentaram melhores resultados para estimar classes de solos em FTS do que imagens no período úmido (**Tabela 5**), sendo que os melhores resultados nos modelos *sat* ocorreram com imagens no período seco e com baixa resolução espacial, os quais tendem a concordar com outras pesquisas, tais como Kalacska et al. (2007), Brungard et al. (2015), Kim et al. (2012) e Vasques et al. (2016). Kalacska et al. (2007) analisaram imagens hiperespectrais de três períodos, a fim de acessar a descrição espacial da estrutura e diversidade da FTS e concluíram que as imagens da estação seca são melhores para estimar os parâmetros biofísicos do dossel e da biodiversidade florística. Também em FTS, Vasques et al. (2016), estimando atributos do solo por regressão linear com seleção de variáveis *stepwise*, observaram que as variáveis selecionadas de imagens RE no período seco foram em geral preferidas às imagens RE no período úmido. Brungard et al. (2015) estimaram classes de solos em três áreas semiáridas e, em duas destas, razões de bandas espectrais Landsat no período seco (*dormant season*) foram selecionadas como parte do melhor conjunto de variáveis ambientais preditoras. No que diz respeito à resolução espacial em diferentes imagens de satélite, Kim et al. (2012) mapearam classes de solos a nível de série e

avaliaram três sensores orbitais (SPOT, Landsat ETM e MODIS), observando que em modelos sem o uso de profundidade do solo como covariável, os melhores resultados foram com imagens de maior (alta) resolução espacial. Por outro lado, quando a profundidade do solo foi utilizada nos modelos, os melhores resultados foram aqueles com imagens de satélite com menor (baixa) resolução espacial, semelhante aos nossos resultados.

Nos modelos *topo* foram selecionadas de 22 a 23 variáveis nas resoluções média e baixa (30-m e 90-m) e 21 variáveis na alta resolução, sendo que as variáveis com maior relevância nos modelos *topo* para estimar classes de solos a nível de grande grupo foram MRRTF, MRVBF, CAR e PFCV (**Tabela 6**). Os índices multirresolução de suavidade de áreas elevadas (MRRTF) e de deposição (MRVBF) foram úteis para distinguir fundos de vales e encostas (GALLANT; DOWLING, 2003). De maneira geral, a importância das variáveis *topo* selecionadas nos modelos de AC variou de acordo com a resolução espacial e a fonte do MDE.

Tabela 6. Tipo de covariável utilizada em cada modelo, número de covariáveis e cinco covariáveis mais utilizadas mostrando a frequência em que aparecem na árvore de classificação, e covariáveis que não foram utilizadas nos modelos.

Modelos	Tipo	Nº	Covariáveis mais utilizadas (%)	Covariáveis não utilizadas
M10m_D_I	<i>topo</i>	21	VDP (11,5), DRIN (9,8), CNBL (8,2), MDE (6,6), MRRTF (6,6)	PLCV, TPI, TRI, VDCN
M30m_D_I		23	MDE (7), DIFIN (7), MRVBF (7), MSP (7), PFCV (7)	CVI, VDCN
M90m_D_I		22	CAR (8,9), MRRTF (8,9), PFCV (7,1), TPI (7,1), TST (7,1)	CVI, TOIN, TRI
M30m_D_s1		22	MRRTF (10,4), MRVBF (8,3), TST (8,3), ASP (6,3), PFCV (6,3)	LSF, TOIN, TPI
M90m_D_s1		23	MRRTF (14), MRVBF (9,3), VDCN (9,3), ASP (4,7), CAR (4,7)	MBI, TPI
M90m_D_s3		22	MRRTF (14), CAR (8), CNBL (8), MBI (8), MRVBF (8)	ASP, PFCV, RSP
M90m_D_G	<i>parent</i>	10	1DV (16,4), Th (14,5), K (12,7), Mag(12,7), Sinal (10,9)	
M10m_D_REdry	<i>sat</i>	8	B1 (27), B3 (17,5), B2 (12,7), B5 (12,7), NDVI (11,1)	
M10m_D_REwet		7	B1 (27), B3 (23), B2 (18,9), B4(12,2), B5(9,5)	SAVI
M30m_D_REdry		7	B1 (21,8), B3 (21,8), NDVI (20), RVI (16,4), B5 (9,1)	SAVI
M30m_D_REwet		8	NDVI (20,7), B2 (19), B5 (19), B4 (13,8), B1 (8,6)	
M30m_D_L8dry		9	B1 (22,1), B3 (17,6), B4 (17,6), B7 (11,8), RVI (7,4)	SAVI
M30m_D_L8wet		10	B3 (15,9), B4 (15,9), B5 (15,9), B1 (14,3), B6 (11,1)	
M90m_D_REdry		7	B1 (22,4), B3 (18,4), NDVI (16,3), RVI (16,3), B5 (14,3)	SAVI
M90m_D_REwet		8	B5 (21,7), B1 (18,7), B3 (11,7), B4 (11,7), RVI (11,7)	
M90m_D_L8dry		10	B1 (27,9), NDVI (13,1), B4 (11,5), B5 (11,5), B6 (9,8)	
M90m_D_L8wet		9	B3 (19,7), B1 (18), B6 (16,4), B5 (11,5), B7 (11,5)	SAVI

Tipo, tipo de covariável ambiental utilizada; Nº, número de covariáveis utilizadas; Covariáveis mais utilizadas (%), cinco covariáveis ambientais mais utilizadas, com a frequência em %; Covariáveis não utilizadas, covariáveis ambientais não utilizadas nos modelos; MDE, elevation; SLO, slope gradient; PFCV, profile curvature; PLCV, plan curvature; ASP, aspect; LSF, LS-factor; VDP, valley depth; RSP, relative slope position; MRVBF, multiresolution index of valley bottom flatness; MRRTF, multiresolution index of ridge top flatness; MBI, mass balance index; TPI, topographic position index; TRI, terrain ruggedness index; TST, terrain surface texture; SLG, slope length; SHT, slope height; MSP, mid slope position; DRIN, direct insolation; DIFIN, diffuse insolation; TWI, topographic wetness index; CVI, convergence index; CAR, catchment area; CNBL, channel network base level; 1DV, 1ª derivada vertical do campo magnético total; Th, concentração de tório; K, concentração de potássio; Mag, campo magnético total; Sinal, sinal analítico do campo magnético total; B1 a B7, bandas espectrais; NDVI, normalized difference vegetation index; RVI, ratio vegetation index; SAVI, soil-adjusted vegetation index.

Geralmente, a covariável do relevo mais relevante em todas as resoluções foi o MRRTF, com mais importância nos modelos de baixa resolução espacial (90-m) e também no com maior acurácia (M90m_D_s3; **Tabela 5**). Além dessa, outras variáveis importantes no melhor modelo (M90m_D_s3) foram, CAR, CNBL, MBI e MRVBF (**Tabela 6**). O MBI foi útil para indicar padrões de solos em superfícies convexas na parte superior da encosta, ressaltando áreas de deposição (depressões) com valores negativos e superfícies de erosão (encostas) com valores positivos (BÖHNER; SELIGE, 2006). Particularmente, no M90m_D_s3 o MBI foi útil para distinguir as classes CXbe e CXve. Por outro lado, Böhner e Selige (2006) indicam que os valores positivos de MBI denotam zonas de acumulação, sendo nessas áreas comum observar solos com influência de rochas carbonáticas e afloramentos rochosos próximos, o que condiz com a realidade de campo. As variáveis de relevo que representam a incidência de calor na superfície (DRIN e DIFIN) apresentaram maior relevância nos modelos a partir do MDE-I de alta e média resolução espacial (**Tabela 6**). Este resultado é similar ao encontrado em Vasques et al. (2016), que mapearam atributos do solo na mesma área de estudo através de técnicas de MDS usando o MDE-I de 10-m. A mesma relevância não foi observada nas outras fontes de MDE (MDEs1 e MDEs3) utilizadas e nem no MDE-I com baixa resolução. As variáveis do relevo menos utilizadas nos modelos foram TPI, ausente em três modelos, além de CVI, TOIN, TRI e VDCN. Foi possível observar que estas variáveis estavam geralmente associadas à mesma fonte de MDE, por exemplo, CVI não foi utilizada nos modelos de média (30-m) e baixa (90-m) resolução espacial a partir do MDE-I (**Tabela 6**).

Modelos L8wet (M30m_D_L8wet e M90m_D_L8wet) apresentaram concordância entre quatro das variáveis mais relevantes dos modelos. As quatro variáveis mais relevantes foram as bandas espectrais verde (B3), que enfatiza o pico de vegetação, *coastal aerosol* (B1), útil para estudos costeiros e de aerossóis, infravermelho próximo (B5), que acentua teor de biomassa, e infravermelho de ondas curtas (B6), que discrimina teores de umidade do solo e da vegetação (USGS, 2016; **Tabela 6**). Por outro lado, nos modelos L8dry (M30m_D_L8dry e M90m_D_L8dry) apenas duas variáveis apresentaram concordância entre as mais relevantes dos modelos, quais sejam: *coastal aerosol* (B1) e vermelho (B4), discriminando a vegetação verde devido à grande absorção desta banda (INPE, 2016; **Tabela 6**). Os índices de vegetação avaliados não apresentaram grande relevância para estimar classes de solos nos modelos

L8wet. No entanto, nos modelos L8dry, os índices de vegetação NDVI e RVI apresentaram maior importância (**Tabela 6**).

As variáveis mais relevantes nos modelos REdry foram as bandas espectrais azul (B1), vermelho (B3) e NIR (B5) e o NDVI, enquanto nos modelos REwet as variáveis mais importantes foram as bandas espectrais azul (B1), *red edge* (B4) e NIR (**Tabela 6**). Segundo Filella e Peñuelas (1994), a banda *red edge* é um bom indicador do conteúdo de clorofila, do índice de área foliar e da umidade da planta. No entanto, segundo os autores, quando em condição de estresse hídrico, os parâmetros do comprimento de onda *red edge* podem ser afetados. Isso explica a maior relevância da banda *red edge* nos modelos no período úmido do que no período seco, o que também foi observado por Vasques et al. (2016). No geral, os índices de vegetação foram mais relevantes nos modelos com imagens do período seco, sendo que o SAVI foi selecionado em somente metade dos modelos.

No modelo *parent* (M90m_D_G), as variáveis gamarradiométricas mais relevantes a fim de estimar classes de solo através de AC foram: 1DV, Th, K, Mag e Sinal (**Tabela 6**). Na área de estudo, Vasques et al. (2016) observaram que a variável Mag, utilizada nos modelos para estimar atributos do solo no PEMS, estava relacionada ao material de origem dos solos. Duas das covariáveis mais relevantes do modelo *parent* podem ser observadas na **Figura 3**, sendo que as menores concentrações de K (**Figura 3A**) e as maiores concentrações de Th (**Figura 3B**) estão nas áreas onde ocorrem os LATOSSOLOS (**Figura 5B**), que são solos muito intemperizados, concordando com Wilford e Minty (2007).

5.3. MODELOS DE PREDIÇÃO E RELAÇÕES SOLO-PAISAGEM

Os modelos *topo* e *parent* representaram todas as 13 classes de solo (**Tabela 5**) em um padrão semelhante ao que ocorre na paisagem (**Figuras 10 e 11**), geralmente mais homogêneo quando observado em mapas de solos convencionais. No entanto, os mapas gerados a partir dos modelos *sat* (**Figuras 8 e 9**) apresentaram um padrão do tipo “*salt and pepper*”, comumente observado em classificação de uso e cobertura da terra baseada em procedimentos por *pixel*. Comparando-se o mapa de solos convencional (COELHO et al., 2013) para a área de estudo e considerando os padrões de distribuição espacial dos solos observados durante as campanhas de campo realizadas, é possível observar que os modelos *sat* com imagens no período seco dos sensores RE (**Figura 8**) e L8 (**Figura 9**), em geral, conseguiram distinguir os solos que ocorrem em depósitos quaternários (CY, GXve e RY), o que não foi possível nos modelos com imagens no período úmido. Outro ponto interessante nos modelos com imagens no período seco foi a melhor distinção dos solos dominantes no domínio do Grupo Bambuí, os CXbe e LVe. Como pode ser observado na **Tabela 7**, em geral, os modelos *sat* no período seco obtiveram melhor acurácia nas duas classes mencionadas acima. Outra característica importante foi o delineamento dos CAMBISSOLOS com alta atividade de argila (CXve) no modelo M30m_D_L8dry (**Figura 9**) na parte sul do PEMS, semelhante ao mapa convencional de solos (**Figura 5B**). No entanto, os modelos *sat* com imagens L8 (**Figura 9**) no período úmido foram melhores para distinguir os solos que ocorrem nos domínios do Grupo Urucua, parte oeste do PEMS, onde os LATOSSOLOS (LVAd, LVd e LAd) de textura média ocorrem, o que pode ser verificado através dos índices de acurácia desses modelos na predição dessas classes (**Tabela 7**).

Os modelos *topo*, em geral, foram os melhores para distinguir todas as 13 classes de solos, tal como apresentado na **Tabela 5**, sendo que os com menor resolução espacial (30-m e 90-m) apresentaram melhores resultados. Nas áreas dos depósitos quaternários, em análise visual, os modelos oriundos do MDE-I, de alta (10-m) e média (30-m) resoluções espaciais, foram melhores para separar os solos que ocorrem na planície de inundação do Rio São Francisco – CY, GXve e RY (**Figura 10**). O mapa derivado do modelo M30m_D_I (**Figura 10**) apresentou algumas semelhanças com o mapa convencional de solos (**Figura 5B**), discriminando, principalmente, CY e GXve nos depósitos aluviais. Porém, ao observarmos os índices de acurácia do produtor (AP) e do usuário (AU) na etapa de validação, os melhores

resultados foram obtidos nos modelos com baixa resolução espacial, especialmente naquele que usou MDEs3 (**Tabela 8**).

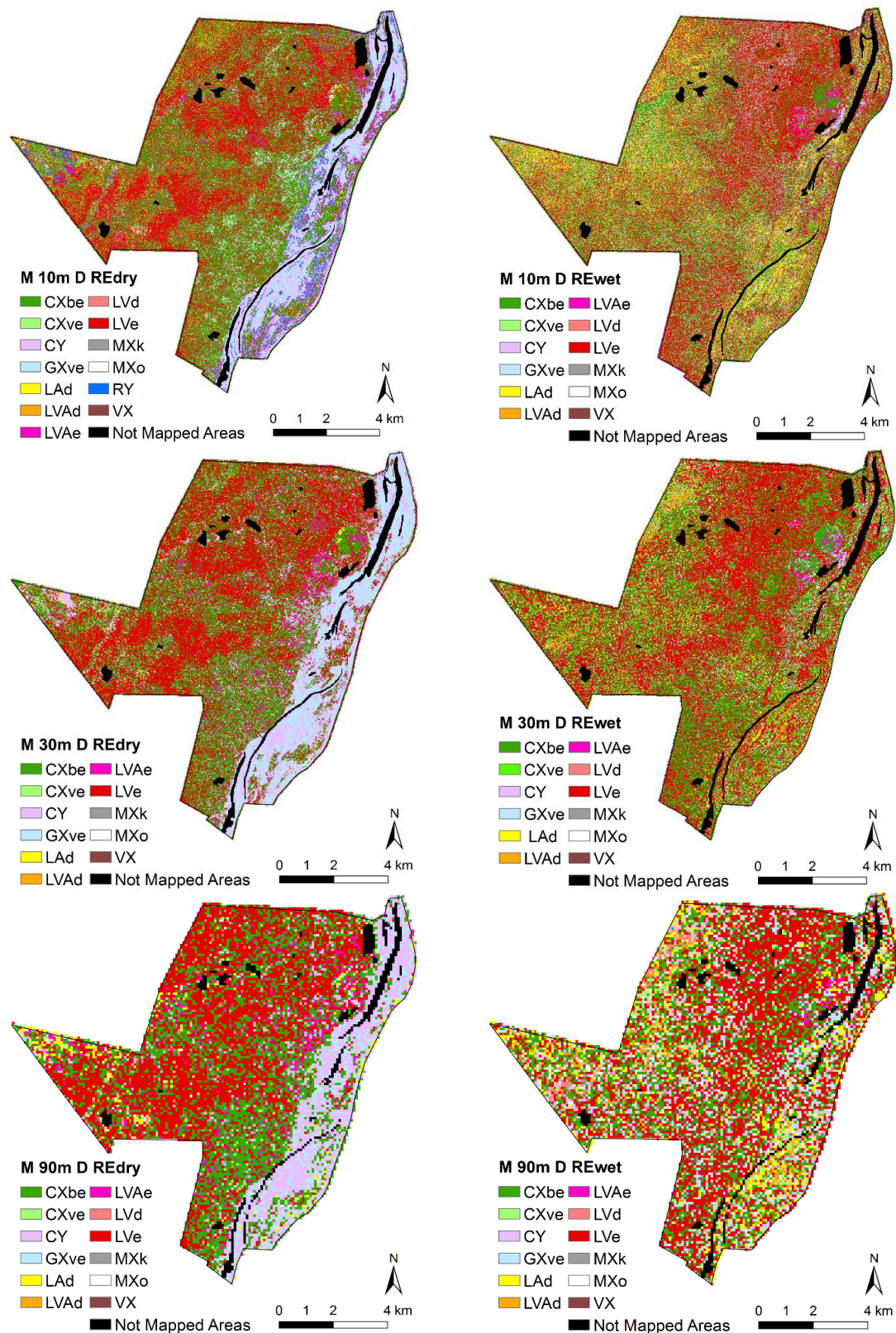


Figura 8. Mapas de solos obtidos pelos modelos usando imagens RapidEye (RE) nos períodos seco (*dry*) e úmido (*wet*).

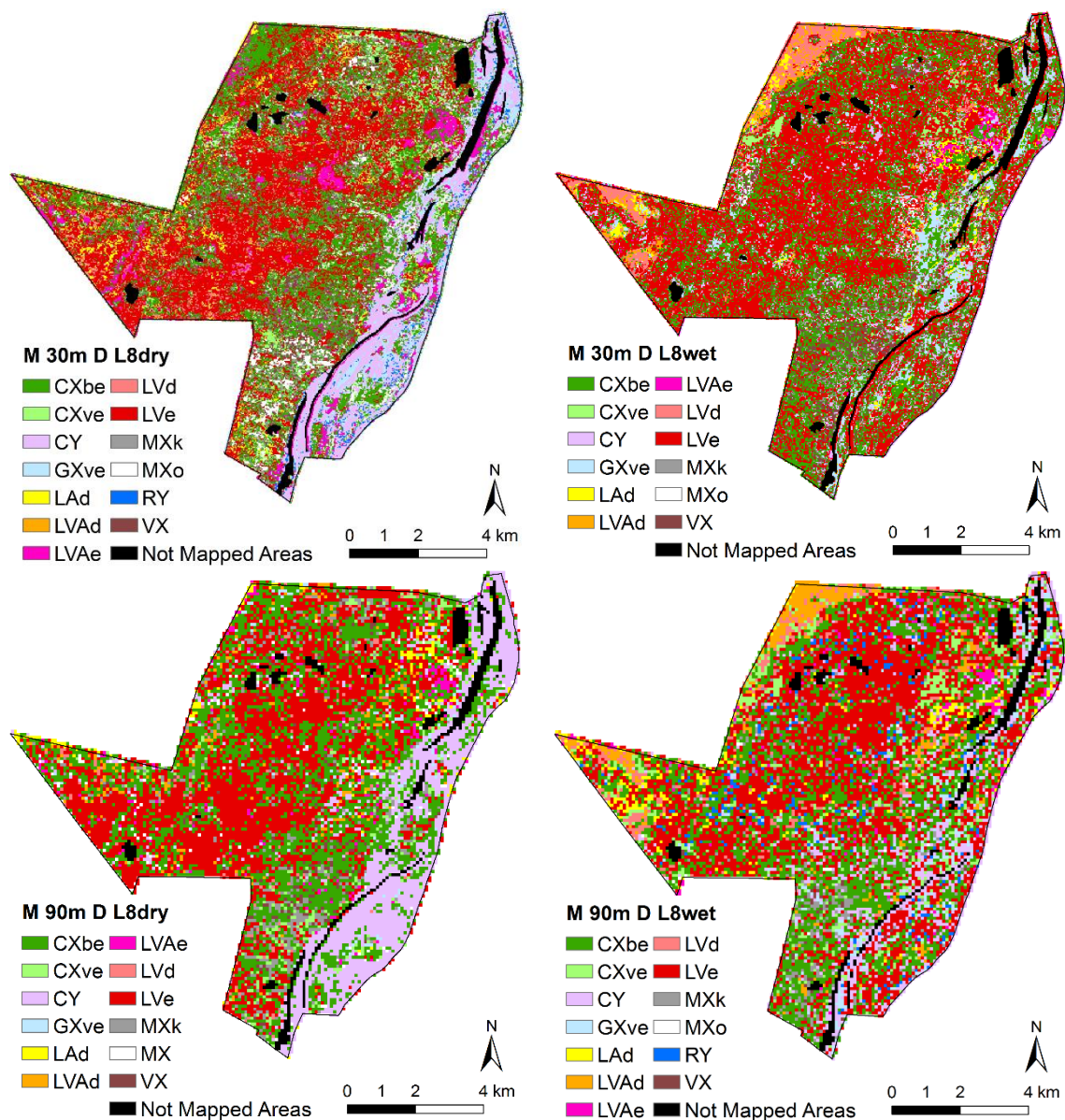


Figura 9. Mapas de solos obtidos pelos modelos usando imagens Landsat 8 (L8) nos períodos seco (dry) e úmido (wet) de diferentes fontes e resoluções espaciais.

Os modelos que utilizaram o MDEs1 (M30m_D_s1 e M90m_D_s1) e o MDEs3 (M90m_D_s3) caracterizaram bem os LATOSSOLOS (LVAd, LVd e LAd) de textura média na parte oeste do PEMS, associados à vegetação “Carrasco” e ao arenito Urucuia (**Figura 10; Tabela 8**). Na área do Grupo Bambuí, os solos dominantes, CXbe e LVe, foram, em geral, melhor discriminados pelos modelos MDEs1 (M30m_D_s1 e M90m_D_s1), MDEs3 (M90m_D_s3) e M30m_D_I (**Figura 10; Tabela 8**). Os CAMBISSOLOS com alta atividade de argila (CXve) geralmente foram representados próximos de talwegues secos ou próximo aos depósitos quaternários, provavelmente devido à posição inferior no relevo. Ponto

interessante do M30m_D_I foi a distinção da classe MXk na parte oeste do PEMS, já que é uma região de grande complexidade para distinguir os limites das classes de solos (COELHO et al., 2013). De maneira geral, três modelos (M30m_D_s1, M90m_D_s1 e M90m_D_s3) conseguiram identificar as principais classes (CXbe, MXk e VX) que ocorrem nessa região (**Figura 10**), o que também pode ser verificado através dos índices de AU e AP (**Tabela 8**).

Visualmente, o modelo *parent* (M90m_D_G) resultou no mapa de solos mais homogêneo (**Figura 11**), provavelmente devido à menor resolução espacial dos dados de gamarradiometria, sendo originalmente 125-m e reamostrados para 90-m. Comparado visualmente com o mapa de solos convencional (**Figura 5B**), o mesmo apresentou delimitação dos LATOSSOLOS (LVAd, LVd e LAd) da porção oeste do PEMS (**Figura 11**) mais próxima ao delineamento do mapa convencional. Porém, ao analisar os resultados de AU e AP (**Tabela 8**), os melhores índices foram obtidos nos modelos *topo* com baixa resolução espacial. Na região sob domínio do Grupo Bambuí, em geral, o mapa da **Figura 8** representou bem as principais classes de solos, LVe e CXbe. Quanto à área do depósito do Rio São Francisco, as três classes de solos mais frequentes (CY, RY e GXve) foram identificadas no mapa obtido de variáveis *parent*. Porém, foi observada a presença da classe LVe neste ambiente, sendo que a mesma não foi observada em campo e nem é esperada que ocorra neste ambiente (**Figura 11**). Conforme apontando por Wilford e Minty (2007), os raios gama conseguem penetrar na vegetação moderada, permitindo medidas geoquímicas do solo, o que os sensores multiespectrais (L8 e RE) não conseguem. Porém, sob vegetação densa, os raios gama podem sofrer atenuação, necessitando de atenção ao interpretar as imagens. Portanto, os maiores erros nas classes RY e GXve, sob floresta perenifólia (**Figura 5C**), bem como das classes CXve, MXo e VX (**Tabela 8**), sob Floresta Estacional Decidual Densa de porte alto, podem estar relacionados à atenuação dos raios gama devido à vegetação densa, como relatado por Wilford e Minty (2007). Problema similar foi observado por Moonjum et al. (2017), que compararam variáveis gamarradiométricas com um mapa de solos convencional a nível de série (escala nominal 1:50.000) na Tailândia, e observaram baixa correspondência nas planícies aluviais. Interessante observar que as áreas representadas na **Figura 11** como LATOSSOLOS correspondem às áreas com menor concentração de K (LAd e LVAd; **Figura 3A**) e maior concentração de tório (LVe; **Figura 3B**), respectivamente.

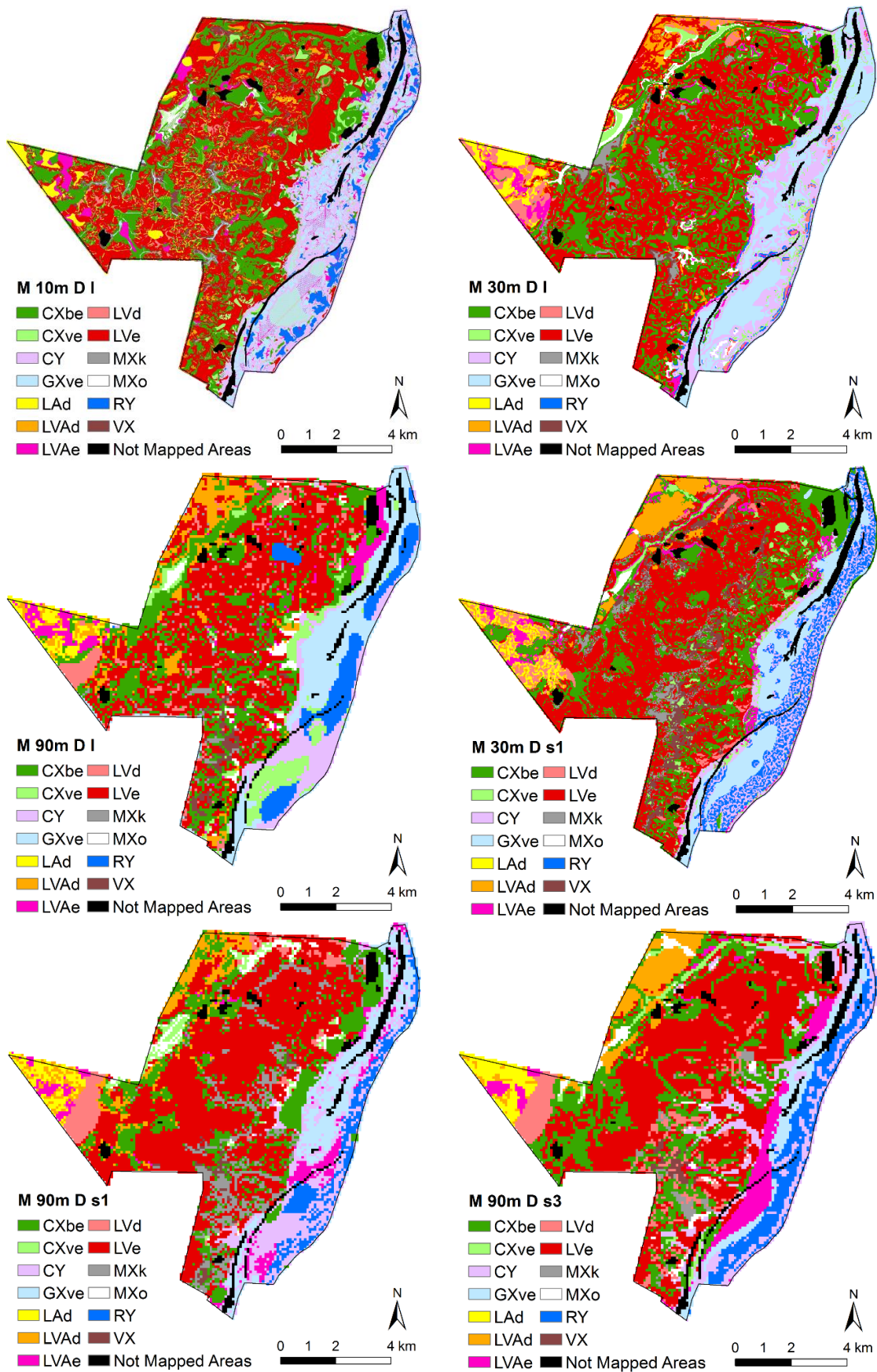


Figura 10. Mapas de solos obtidos pelos modelos usando covariáveis topográficas (*topo*) de diferentes fontes e resoluções espaciais.

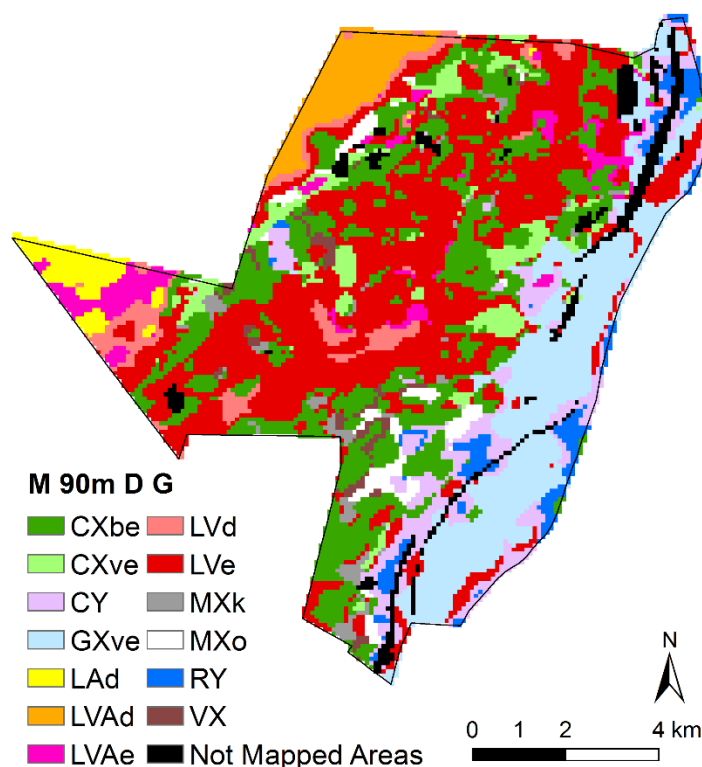


Figura 11. Mapa de solos obtido pelo modelo usando variáveis gamarradiométricas e magnetométricas (*parent*).

5.4. VALIDAÇÃO DOS MODELOS E MAPAS

Os melhores resultados de AG na validação independente dos modelos e mapas ocorreram nos modelos com 90-m de resolução espacial (**Tabela 5**), sendo que nos modelos *topo* ela variou de 44,5 (M90m_D_I) a 53,1% (M90m_D_s3), seguido pelo modelo *parent* (M90m_D_G), com AG de 43,8%, e cinco modelos *sat* com imagens no período seco (M90m_D_L8dry, M90m_D_REdry, M10mDREdry, M30m_D_L8dry e M30m_D_REdry), com AG variando de 39,1 a 45,3%. Os piores resultados de AG ocorreram nos modelos *sat* com imagens no período úmido, com valores variando de 28,1 (M30m_D_REwet) a 39,8% (M30m_D_L8wet), e no modelo *topo* com MDE de alta resolução espacial (M10m_D_I), com AG de 35,2%.

O melhor \widehat{K} de validação alcançado (0,40) ocorreu no modelo *topo* com baixa resolução espacial (M90m_D_s3). Segundo Landis e Koch (1977; **Tabela 4**), dez modelos obtiveram concordância justa e outros sete concordância fraca (**Tabela 5**). Os dez modelos considerados com concordância justa são os modelos *topo* com baixa resolução espacial (90-m), os modelos *sat* no período seco e o modelo *parent*, enquanto os sete modelos com

concordância fraca foram os *sat* com imagens no período úmido e os *topo* com média (30-m) e alta (10-m) resoluções espaciais.

O melhor resultado de validação deste trabalho foi semelhante aos de pesquisas que utilizaram procedimento de AC com conjunto de amostras independentes (HEUNG et al., 2016; LEMERCIER et al., 2012; KIM et al., 2012; LACOSTE et al., 2011; GRINAND et al., 2008) e que utilizaram o algoritmo C5 especificamente (MINANSNY; MCBRATNEY, 2007; MENDONÇA-SANTOS et al., 2008). No entanto, Taghizadeh-Mehrjardi et al. (2015) obtiveram resultados superiores (0,71 de AG e 0,69 de \widehat{K}) utilizando o algoritmo C5 para estimar classes de solo ao nível de família (*Soil Taxonomy*) em ambiente semiárido. Eles estimaram cinco classes de solos, enquanto no presente trabalho foram estimadas 13 classes, onde provavelmente existem maiores confusões entre estas. Além disso, a generalização das classes de solos possibilita aumento da acurácia da predição, como observado em alguns trabalhos em MDS (MINANSNY; MCBRATNEY, 2007; DART et al., 2010; KEMPEN et al., 2011; VASQUES et al., 2015).

Tabela 7. Acurácias do usuário (AU) e do produtor (AP) nos modelos *sat*, para o conjunto de dados de validação.

Modelos / Classes	M10m_D_REdry		M10m_D_REwet		M30m_D_REdry		M30m_D_REwet		M90m_D_REdry		M90m_D_REwet		M30m_D_L8dry		M30m_D_L8wet		M90m_D_L8dry		M90m_D_L8wet	
	AU	AP	AU	AP	AU	AP	AU	AP	AU	AP	AU	AP	AU	AP	AU	AP	AU	AP	AU	AP
CXve	0,0	0,0	0,0	0,0	0,0	0,0	33,3	50,0	33,3	33,3	33,3	25,0	33,3	14,3	33,3	50,0	0,0	0,0	33,3	20,0
CXbe	55,3	43,8	39,5	36,6	50,0	39,6	47,4	35,3	55,3	41,2	39,5	38,5	52,6	44,4	36,8	38,9	47,4	46,2	44,7	37,8
CY	0,0	0,0	16,7	14,3	16,7	25,0	0,0	0,0	50,0	50,0	0,0	0,0	66,7	80,0	0,0	0,0	83,3	50,0	16,7	20,0
GXve	20,0	14,3	20,0	50,0	20,0	16,7	20,0	12,5	20,0	33,3	20,0	33,3	20,0	20,0	40,0	66,7	0,0	0,0	40,0	66,7
LAd	0,0	0,0	25,0	20,0	0,0	0,0	25,0	16,7	0,0	0,0	50,0	25,0	25,0	25,0	0,0	0,0	25,0	33,3	0,0	0,0
LVe	60,5	60,5	41,9	40,0	69,8	54,5	34,9	40,5	67,4	52,7	44,2	33,3	51,2	59,5	72,1	50,0	76,7	57,9	46,5	44,4
LVd	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	25,0	25,0
LVAAd	25,0	25,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	25,0	33,3	50,0	33,3
LVAe	16,7	14,3	50,0	60,0	16,7	33,3	0,0	0,0	16,7	25,0	16,7	50,0	0,0	0,0	16,7	50,0	0,0	0,0	33,3	28,6
MXo	25,0	25,0	0,0	0,0	25,0	50,0	0,0	0,0	0,0	0,0	25,0	25,0	0,0	0,0	25,0	100,0	0,0	0,0	0,0	0,0
MXk	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
RY	33,3	100,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	33,3	100,0	0,0	0,0	0,0	0,0	0,0	0,0
VX	0,0	0,0	25,0	11,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0

CXbe, CAMBISSOLOS HÁPLICOS Tb Eutróficos; CXve, CAMBISSOLOS HÁPLICOS Ta Eutróficos; CY, CAMBISSOLOS FLÚVICOS; GXve, GLEISSOLOS HÁPLICOS Ta Eutróficos; LAd, LATOSSOLOS AMARELOS Distróficos; LVAAd, LATOSSOLOS VERMELHO-AMARELOS Distróficos; LVAe, LATOSSOLOS VERMELHO-AMARELOS Eutróficos; LVd, LATOSSOLOS VERMELHOS Distróficos; LVe, LATOSSOLOS VERMELHOS Eutróficos; MXk, CHERNOSSOLOS HÁPLICOS Carbonáticos; MXo, CHERNOSSOLOS HÁPLICOS Órticos; RY, NEOSSOLOS FLÚVICOS; VX, VERTISSOLOS HÁPLICOS.

Tabela 8. Acurácias do usuário (AU) e do produtor (AP) nos modelos *topo* e *parent*, para o conjunto de dados de validação.

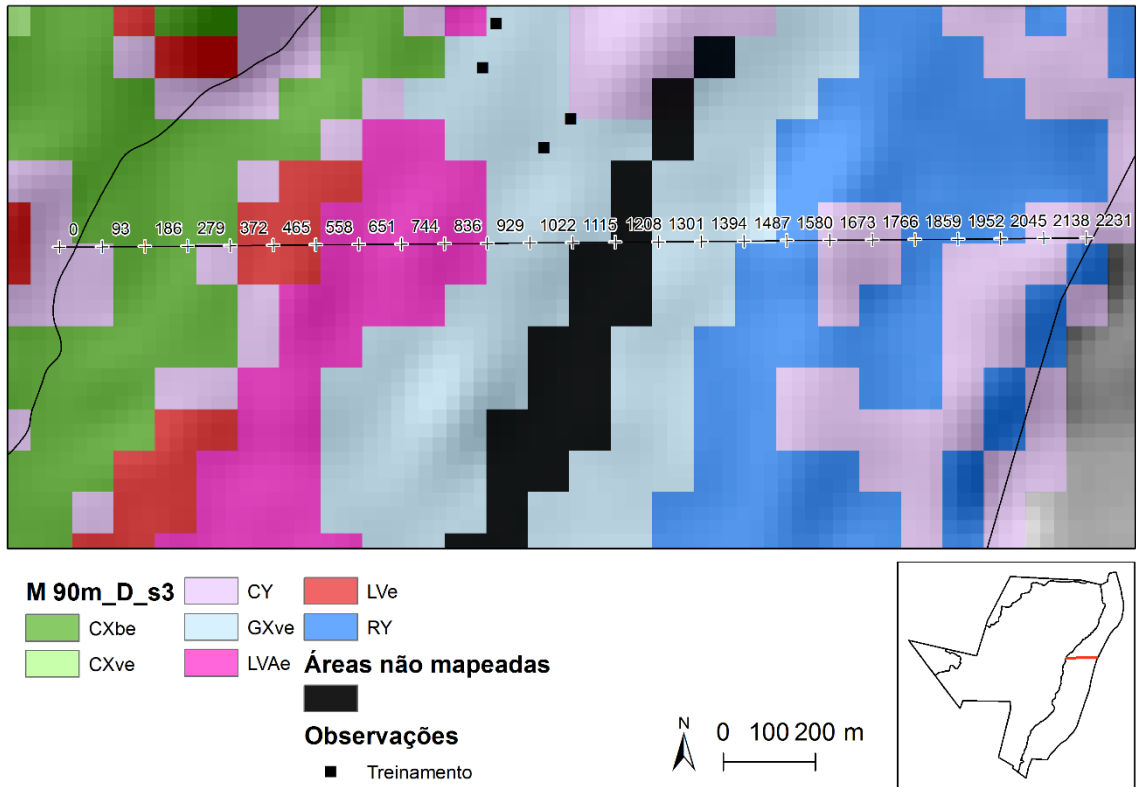
Modelos / Classes	M90m_D_s3		M90m_D_s1		M90m_D_I		M30m_D_s1		M30m_D_I		M10m_D_I		M90m_D_G	
	AU	AP	AU	AP	AU	AP	AU	AP	AU	AP	AU	AP	AU	AP
CXve	33,3	50,0	33,3	25,0	33,3	33,3	33,3	20,0	33,3	33,3	33,3	50,0	0,0	0,0
CXbe	31,6	42,9	36,8	50,0	39,5	40,5	47,4	45,0	26,3	35,7	23,7	25,0	39,5	42,9
CY	66,7	80,0	33,3	33,3	33,3	50,0	0,0	0,0	33,3	28,6	16,7	50,0	66,7	44,4
GXve	40,0	100,0	20,0	50,0	60,0	37,5	20,0	100,0	20,0	50,0	40,0	33,3	20,0	25,0
LAd	25,0	50,0	25,0	33,3	25,0	33,3	25,0	50,0	0,0	0,0	25,0	25,0	25,0	50,0
LVe	74,4	57,1	79,1	58,6	65,1	59,6	60,5	54,2	67,4	52,7	62,8	45,0	65,1	58,3
LVd	25,0	33,3	25,0	33,3	25,0	33,3	25,0	25,0	0,0	0,0	0,0	0,0	25,0	33,3
LVAAd	75,0	50,0	25,0	50,0	50,0	28,6	50,0	66,7	25,0	14,3	0,0	0,0	50,0	40,0
LVAe	50,0	50,0	33,3	28,6	33,3	25,0	33,3	33,3	33,3	50,0	0,0	0,0	16,7	12,5
MXo	50,0	25,0	25,0	50,0	25,0	33,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
MXk	50,0	50,0	75,0	30,0	0,0	0,0	50,0	28,6	25,0	20,0	25,0	33,3	50,0	66,7
RY	100,0	100,0	0,0	0,0	33,3	50,0	66,7	33,3	33,3	33,3	100,0	75,0	0,0	0,0
VX	50,0	66,7	50,0	100,0	0,0	0,0	75,0	60,0	0,0	0,0	0,0	0,0	25,0	100,0

CXbe, CAMBISSOLOS HÁPLICOS Tb Eutróficos; CXve, CAMBISSOLOS HÁPLICOS Ta Eutróficos; CY, CAMBISSOLOS FLÚVICOS; GXve, GLEISSOLOS HÁPLICOS Ta Eutróficos; LAd, LATOSSOLOS AMARELOS Distróficos; LVAAd, LATOSSOLOS VERMELHO-AMARELOS Distróficos; LVAe, LATOSSOLOS VERMELHO-AMARELOS Eutróficos; LVd, LATOSSOLOS VERMELHOS Distróficos; LVe, LATOSSOLOS VERMELHOS Eutróficos; MXk, CHERNOSSOLOS HÁPLICOS Carbonáticos; MXo, CHERNOSSOLOS HÁPLICOS Órticos; RY, NEOSSOLOS FLÚVICOS; VX, VERTISSOLOS HÁPLICOS.

5.5. RELAÇÕES SOLO-PAISAGEM DO MELHOR MODELO

No melhor modelo, M90m_D_s3, todas as 13 classes de solos foram estimadas, em geral, com boa representação espacial e alocadas onde são esperadas de acordo com as relações solo-paisagem do PEMS observadas em campo e descritas em Coelho et al. (2013) e Vasques et al. (2016). O modelo M90m_D_s3, em algumas áreas, conseguiu separar classes de solos que o pedólogo não conseguiu separar nitidamente no campo. As diferenças mais marcantes, em comparação com o mapa de solos convencional (COELHO et al., 2013), foram o encolhimento da área de quatro classes, CXbe, CXve, GXve e CY, com o aumento da área de outras cinco classes, RY, LVd, LVe, LVAe e MXo. Por exemplo, as classes de solos com maior área no mapa convencional são LVe (30% da área) e CXbe (28%), enquanto no mapa predito pelo modelo M90m_D_s3, elas ocuparam, respectivamente, 34 e 18% da área. A classe CXbe, que apresentava grande expressão no domínio do Grupo Bambuí no mapa convencional, diminuiu a sua área, enquanto as classes LVe e LVd aumentaram a sua área de abrangência nesse domínio. Essas variações de classes são esperadas, uma vez que, segundo Coelho et al. (2013), os CAMBISSOLOS deste ambiente apresentam morfologia típica de LATOSSOLOS, com diferenças sutis na capacidade de troca catiônica (CTC) da fração de argila e no índice Ki; LATOSSOLOS devem ter índice Ki inferior a 2,2 (EMBRAPA, 2006), sendo que um perfil que representa a classe CXbe nesse ambiente possui índice Ki de 2,44, próximo ao limite entre classes. A classe CY ocupou área da classe CXbe do mapa convencional no domínio do Grupo Bambuí nas áreas de sopé. Além disso, a classe CY ocupou área da classe CXbe do mapa convencional no domínio do Grupo Bambuí nas áreas de sopé. Nos depósitos quaternários do Rio São Francisco, duas classes, CY e GXve, ficaram menores comparadas com o mapa de solos convencional, sendo substituídas pela classe RY. Como mencionado por Coelho et al. (2013), nas áreas de depósitos de sedimentos aluviais, as classes de solos são difíceis de serem individualizadas em unidades simples. Não obstante, no mapa resultante do modelo M90m_D_s3, a classe GXve pode ser observada com clareza nas áreas deprimidas, permanente- ou temporariamente inundadas, como expresso na **Figura 12**. A classe LVAe expandiu na área dos depósitos quaternários, provavelmente delimitada erroneamente devido à semelhança das variáveis de relevo utilizadas na definição de classes pela AC.

A



C

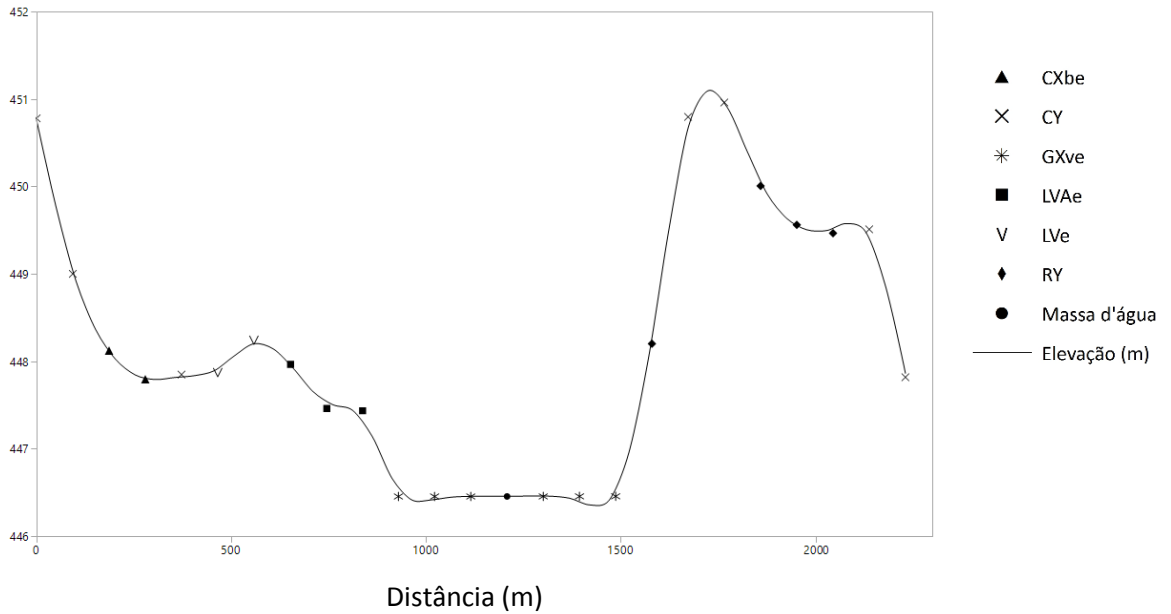


Figura 12. Detalhe do mapa de classes de solo produzido a partir do modelo M90m_D_s3, mostrando topossequência de solos na área de depósitos sedimentares do Rio São Francisco. A) Extrato do mapa de solos obtido pelo modelo M90m_D_s3; B) Localização da topossequência de solos; e C) Perfil topográfico mostrando a topossequência de solos previstos pelo modelo M90m_D_s3, com destaque para a identificação da classe GXve na posição mais baixa da paisagem (~446 m).

O modelo M90m_D_s3 conseguiu separar bem os solos sob influência do arenito do Grupo Urucua, sob vegetação do tipo "Carrasco", dos solos no domínio do Grupo Bambuí, sob vegetação do tipo Floresta Tropical Seca. A evidência dessa área como uma região de transição geológica é a presença de LATOSSOLOS com baixa saturação por bases e textura média (LAd, LVAd e LVd), na parte oeste do PEMS, sob influência da vegetação "Carrasco", adjacentes a LATOSSOLOS com alta saturação por bases e textura média (LV Ae) mais ao centro do PEMS. Essa transição pode ser observada na **Figura 13**, onde os LAd, LVAd e LVd estão dispostos no topo e ombro da topossequência, solos com alta saturação por bases e texturas média para argilosa (CXbe e LVe) ocorrem na encosta, e sobre os talwegues secos ocorrem os MXo, estes formados a partir de rochas carbonáticas, concordando com Coelho et al. (2013).

Olhando em detalhe a sua matriz de confusão de treinamento (**Tabela 9**), o modelo M90m_D_s3 obteve ótimos resultados baseando-se na AU (84,7%) e AP (86,5%). Ao avaliar as classes de solos em nível de grande grupo individualmente, podemos observar que as melhores taxas de concordância ocorreram nas classes LAd e LVe, com AU e AP acima de 90%. Além destas, outras sete classes apresentaram valor de AU e AP acima 80%. O pior resultado ocorreu nos CAMBISSOLOS com alta atividade de argila (CXve), provavelmente devido à dificuldade de separar esta classe apenas pelas variáveis topográficas.

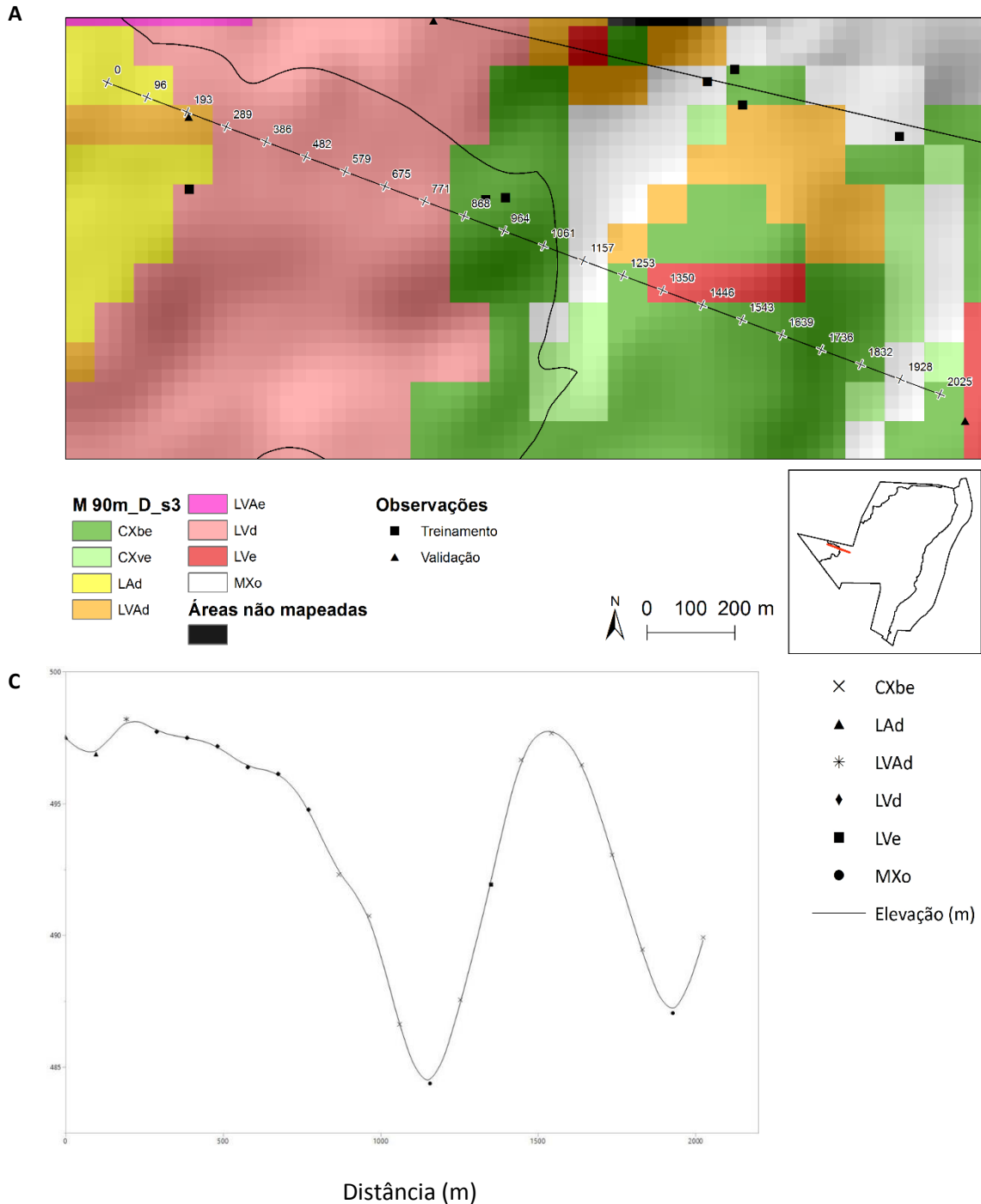


Figura 13. Detalhe do mapa de classes de solo produzido pelo modelo M90m_D_s3, mostrando topossequência de solos na área de transição do Grupo Urucua para o Grupo Bambuí. A) Mapa de solos obtido pelo modelo M90m_D_s3; B) Localização da topossequência de solos; C) Perfil topográfico mostrando a topossequência de solos preditos pelo modelo M90m_D_s3.

Avaliando a matriz de confusão de validação do modelo M90m_D_s3 (**Tabela 10**), observa-se que o mesmo estimou os solos dos depósitos quaternários (CY, GXve e RY) com as melhores taxas de concordância, sendo que a classe RY, que possui pequeno número de observações, alcançou resultado perfeito. Por outro lado, a classe GXve apresentou 100% de

AP e apenas 40% de AU, estimando corretamente apenas duas amostras de um conjunto de cinco. A classe CXbe apresentou AP de 42,9% e AU de 31,6%, sendo que, dentre as 38 observações dessa classe no conjunto de validação, apenas 12 foram corretamente classificadas. Essa taxa de erro elevada é esperada, uma vez que a classe CXbe ocorre em diferentes posições da paisagem e se confunde com outras seis classes de solos (CXve, CY, LVe, LVAe, MXo e MXk). A confusão mais marcante ocorreu com o LVe, com 17 erros. Isso pode ser explicado porque, segundo Coelho et al. (2013), os CAMBISSOLOS (CXbe) apresentam morfologia similar a de LATOSSOLOS (LVe) e ocorrem nas mesmas superfícies geográficas.

Os LATOSSOLOS com baixa saturação por bases e textura média, LAd, LVAd e LVd, na parte oeste do PEMS (**Figura 11**), geralmente apresentaram baixa taxa de acerto na validação, com exceção do LVAd (75% de AU), enquanto os LATOSSOLOS com alta saturação por bases apresentaram AU de 50% (LVAe) e 74% (LVe) (**Tabela 10**). Os VERTISSOLOS (VX) e CHERNOSSOLOS (MXo e MXk), que ocorrem sob influência de rochas carbonáticas e em posições mais baixas da paisagem, apresentaram AU de 50%. Geralmente, o mapa obtido pelo modelo M90m_D_s3 (**Figura 11**) delimitou bem as classes VX, MXo e MXk, com base no mapa de solos convencional (**Figura 5B**).

Tabela 9. Matriz de confusão do modelo M90m_D_s3, na etapa de treinamento.

Predito	Observado / Referência													Total	AU
	CXve	CXbe	CY	GXve	LAd	LVe	LVd	LVAd	LVAe	MXo	MXk	RY	VX		
CXve	5	2				1				1				9	55,6
CXbe	2	79				2			2		1		1	87	90,8
CY			14	1								1		16	87,5
GXve		2		10										12	83,3
LAd					9									9	100,0
LVe		3				95								98	96,9
LVd		2					7							9	77,8
LVAd					1			8						9	88,9
LVAe		1	1				1	1	10					14	71,4
MXo		1								8				9	88,9
MXk						1	1				9			11	81,8
RY												4		4	100,0
VX		1				1							7	9	77,8
Total	7	91	15	11	10	100	9	9	12	9	10	5	8	296	
AP	71,4	86,8	93,3	90,9	90,0	95,0	77,8	88,9	83,3	88,9	90,0	80,0	87,5		

CXbe, CAMBISSOLOS HÁPLICOS Tb Eutróficos; CXve, CAMBISSOLOS HÁPLICOS Ta Eutróficos; CY, CAMBISSOLOS FLÚVICOS; GXve, GLEISSOLOS HÁPLICOS Ta Eutróficos; LAd, LATOSSOLOS AMARELOS Distróficos; LVAd, LATOSSOLOS VERMELHO-AMARELOS Distróficos; LVAe, LATOSSOLOS VERMELHO-AMARELOS Eutróficos; LVd, LATOSSOLOS VERMELHOS Distróficos; LVe, LATOSSOLOS VERMELHOS Eutróficos; MXk, CHERNOSSOLOS HÁPLICOS Carbonáticos; MXo, CHERNOSSOLOS HÁPLICOS Órticos; RY, NEOSSOLOS FLÚVICOS; VX, VERTISSOLOS HÁPLICOS; AU, acurácia do usuário; AP, acurácia do produtor.

Tabela 10. Matriz de confusão do modelo M90m_D_s3, na etapa de validação.

Predito	Observado / Referência													Total	AU
	CXve	CXbe	CY	GXve	LAd	LVe	LVd	LVAd	LVAe	MXo	MXk	RY	VX		
CXve	1	1								1				3	33,3
CXbe	1	12				17			2	3	2			38	31,6
CY			4			2								6	66,7
GXve		2		2					1					5	40,0
LAd		2			1			1						4	25,0
LVe		9				32	1	1						43	74,4
LVd						3	1							4	25,0
LVAd								3		1				4	75,0
LVAe					1	1		1	3					6	50,0
MXo						1	1			2				4	50,0
MXk										1	2		1	4	50,0
RY												3		3	100,0
VX		2											2	4	50,0
Total	2	28	5	2	2	56	3	6	6	8	4	3	3	128	
AP	50,0	42,9	80,0	100,0	50,0	57,1	33,3	50,0	50,0	25,0	50,0	100,0	66,7		

CXbe, CAMBISSOLOS HÁPLICOS Tb Eutróficos; CXve, CAMBISSOLOS HÁPLICOS Ta Eutróficos; CY, CAMBISSOLOS FLÚVICOS; GXve, GLEISSOLOS HÁPLICOS Ta Eutróficos; LAd, LATOSSOLOS AMARELOS Distróficos; LVAd, LATOSSOLOS VERMELHO-AMARELOS Distróficos; LVAe, LATOSSOLOS VERMELHO-AMARELOS Eutróficos; LVd, LATOSSOLOS VERMELHOS Distróficos; LVe, LATOSSOLOS VERMELHOS Eutróficos; MXk, CHERNOSSOLOS HÁPLICOS Carbonáticos; MXo, CHERNOSSOLOS HÁPLICOS Órticos; RY, NEOSSOLOS FLÚVICOS; VX, VERTISSOLOS HÁPLICOS; AU, acurácia do usuário; AP, acurácia do produtor.

6. CONCLUSÕES

De maneira geral, os objetivos do trabalho foram alcançados com êxito, de forma que, entre as covariáveis avaliadas, foram identificadas quais fontes, resoluções espaciais e períodos de aquisição são mais adequados para predizer classes de solos em Floresta Tropical Seca (FTS) no Brasil.

Os resultados do trabalho permitem concluir que:

- O procedimento de árvore de classificação (AC) permite predizer solos em ambiente de FTS em nível de grande grupo com AG de 53,1% de validação;
- A representação espacial de todas as classes observadas no campo pelo modelo obtido por AC depende das covariáveis utilizadas no mesmo;
- Covariáveis de relevo com baixa resolução espacial (90-m) e obtidas por imagens de satélite no período seco permitem predizer com maior acerto classes de solos em FTS;
- Covariáveis obtidas a partir de imagens gamarradiométricas/magnetométricas fornecem melhores estimativas de classes de solos do que os sensores multiespectrais e fornecem um bom retrato das relações solo-geologia;
- A fonte, a resolução espacial e o período de aquisição das covariáveis influenciam na qualidade dos modelos de predição de classes de solos em FTS e nos seus padrões de distribuição espacial nos mapas derivados. A seleção das covariáveis deve levar em consideração, além da qualidade dos mapas, os custos envolvidos, pois covariáveis obtidas gratuitamente derivam mapas de igual ou melhor qualidade do que covariáveis derivadas de fontes pagas;
- Os procedimentos adotados são aplicáveis a áreas similares, ou mesmo outras áreas, sendo encorajado o uso de covariáveis obtidas de fontes de dados gratuitas, tais como, dados da *Shuttle Radar Topography Mission* e do satélite Landsat 8.

7. REFERÊNCIAS BIBLIOGRÁFICAS

BEHRENS, T.; SCHMIDT, K.; RAMIREZ-LOPEZ, L.; GALLANT, J.; ZHU, A. X.; SCHOLTEN, T. Hyper-scale digital soil mapping and soil formation analysis. **Geoderma**, v. 213, n. 0, p. 578-588, 2014.

BEHRENS, T.; ZHU, A.X.; SCHMIDT, K.; SCHOLTEN, T. Multi-scale digital terrain analysis and feature selection for digital soil mapping. **Geoderma**, v. 155, p. 175–185, 2010.

BOCK, M.; BÖHNER, J.; CONRAD, O.; KÖTHE, R.; RINGELER, A. Methods for creating Functional Soil Databases and applying Digital Soil Mapping with SAGA GIS. In: HENGL, T.; PANAGOS, P.; JONES, A.; TOTH, G. **Status and prospect of soil information in south-eastern Europe: soil databases, projects and applications**. Luxemburg: Office for Official Publications of the European Communities, 2007, p. 149-162.

BÖHNER, J.; SELIGE, T. Spatial prediction of soil attributes using terrain analysis and climate regionalization. In: BÖHNER, J.; MCCLOY, K. R.; STROBL, J. **SAGA – Analysis and Modelling Applications**. Hamburg: Göttinger Geographische Abhandlungen, 2006, p. 13-28.

BRASIL. **Geo Catálogo MMA**: Catálogo de imagens de satélite RapidEye do Ministério do Meio Ambiente. Ministério do Meio Ambiente, Brasília, Disponível em: <<http://www.geocatalogomma.com.br>>. Acesso em: 28 mar.2014

BREIMAN, L. Random forests. **Machine Learning**, v.45, p.5-32, 2001.

BREIMAN, L.; FRIEDMAN, J.; OLSHEN, R.; STONE, C. **Classification and Regression Trees**, New York: Wadsworth, 1984.

BRUNGARD, C. W.; BOETTINGER, J. L.; DUNIWAY, M. C.; WILLS, S. A.; EDWARDS JR, T. C. Machine learning for predicting soil classes in three semi-arid landscapes. **Geoderma**, v. 239–240, n. 0, p. 68-83, 2015.

BUI, E. N.; MORAN, C. J. A strategy to fill gaps in soil survey over large spatial extents: an example from the Murray–Darling basin of Australia. **Geoderma**, v. 111, n. 1–2, p. 21-44, 2003.

BURROUGH, P. A.; MCDONNELL, R. A. **Principles of Geographical Information Systems**. Oxford: Oxford University Press, 2004.

CAVAZZI, S.; CORSTANJE, R.; MAYR, T.; HANNAM, J.; FEALY, R. Are fine resolution digital elevation models always the best choice in digital soil mapping? **Geoderma**, v. 195–196, n. 0, p. 111-121, 2013.

CGIAR CONSORTIUM FOR SPATIAL INFORMATION. **SRTM 90m Digital Elevation Database v4.1**. Disponível em: <<http://www.cgiar-csi.org/data/srtm-90m-digital-elevation-database-v4-1>>. Acesso em: 17 dez.2014

COELHO, M. R.; DART, R. O.; VASQUES, G. M.; TEIXEIRA, W. G.; OLIVEIRA, R. P.; MENDONÇA-SANTOS, M. L.; BERBARA, R. L. L. Levantamento Pedológico Semi-detalhado (1:30.000) do Parque Estadual da Mata Seca, Município de Manga - MG. **Boletim de Pesquisa e Desenvolvimento**, v. 217, p. 264, 2013.

CONGALTON, R. G. A review of assessing the accuracy of classifications of remotely sensed data. **Remote Sensing of Environment**, v. 37, n. 1, p. 35-46, 1991.

CONGALTON, R. G.; GREEN, K. **Assessing the Accuracy of Remotely Sensed Data: Principles and Practices**. 2.ed. Boca Raton, CRC Press, 2009.

DART, R. O., MENDONÇA-SANTOS, M. L.; COELHO, M. R.; BERBARA, R. L. L.; PARES, J. G. Digital Soil Mapping at Parque Estadual da Mata Seca, MG, Brazil: Applying Regression Tree to predict soil classes. In: 4TH INTERNATIONAL WORKSHOP ON DIGITAL SOIL MAPPING: FROM DIGITAL SOIL MAPPING TO DIGITAL SOIL ASSESSEMENT: IDENTIFYING KEY GAPS FROM FIELDS TO CONTINENTS, 2010, Rome, **Anais**. Não paginado.

DART, R. O.; VASQUES, G. M.; COELHO, M. R.; FERNANDES, N. F. Digital soil mapping for soil class prediction in a dry forest of Minas Gerais, Brazil. In: XXXV CONGRESSO BRASILEIRO DE CIÊNCIA DO SOLO, 2015, Natal. **Anais**. Não paginado.

DIETRICH, H.; BÖHNER, J. Cold Air Production and Flow in a Low Mountain Range Landscape in Hessa. In: BÖHNER, J.; BLASCHKE, T.; MONTANARELLA, L. **SAGA - Seconds Out**. Hamburg: Göttinger Geographische Abhandlungen, 2008, p. 37-48.

EMBRAPA. **Sistema Brasileiro de Classificação de Solos**. 2. ed. Rio de Janeiro: EMBRAPA-SPI, 2006. 306 p.

ESPÍRITO-SANTO, M. M.; FAGUNDES, M.; SEVILHA, A. C.; SCARIOT, A. O.; SANCHEZ-AZOFEIFA, G. A.; NORONHA, S. E.; FERNANDES, G. W. Florestas estacionais decíduas brasileiras: distribuição e estado de conservação. **MG BIOTA**, v. 1, n. 2, p. 5-13, 2008.

FILELLA, I.; PENUELAS, J. The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status. **International Journal of Remote Sensing**, v. 15, n. 7, p. 1459-1470, 1994.

FLORINSKY, I. V. The Dokuchaev hypothesis as a basis for predictive digital soil mapping (on the 125th anniversary of its publication). **Eurasian Soil Science**, v. 45, n. 4, p. 445-451, 2012.

GALLANT, J. C.; DOWLING, T. I. A multiresolution index of valley bottom flatness for mapping depositional areas. **Water Resources Research**, v. 39, n. 12, p. 1347-1359, 2003.

GRINAND, C.; ARROUAYS, D.; LAROCHE, B.; MARTIN, M. P. Extrapolating regional soil landscapes from an existing soil map: Sampling intensity, validation procedures, and integration of spatial context. **Geoderma**, v. 143, n. 1-2, p. 180-190, 2008.

GUISAN, A.; WEISS, S. B.; WEISS, A. D. GLM versus CCA spatial modeling of plant species distribution. **Plant Ecology**, v. 143, n. 1, p. 107-122, 1999.

HANSEN, M. K.; BROWN, D. J.; DENNISON, P. E.; GRAVES, S. A.; BRICKLEMYER, R. S. Inductively mapping expert-derived soil-landscape units within dambo wetland catenae using multispectral and topographic data. **Geoderma**, v. 150, n. 1-2, p. 72-84, 2009.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: data mining, inference and prediction**. 2.ed. California, Springer, 2009, 745p.

HEUNG, B.; HO, H. C.; ZHANG, J.; KNUDBY, A.; BULMER, C. E.; SCHMIDT, M. G. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. **Geoderma**, v. 265, p. 62-77, 2016.

HEUNG, B.; HODÚL, M.; SCHMIDT, M. G. Comparing the use of training data derived from legacy soil pits and soil survey polygons for mapping soil classes. **Geoderma**, v. 290, p. 51-68, 2017.

HUETE, A.R. A soil-adjusted vegetation index (SAVI). **Remote Sensing of Environment**, v. 25, n. 3, p. 295-309, 1988.

HUTCHINSON, M. F. A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. **Journal of Hydrology**, v. 106, n. 3-4, p. 211-232, 1989.

IBGE. Mapa de Vegetação do Brasil. Rio de Janeiro: IBGE. 2006. Escala 1:5.000.000. Disponível em: <http://downloads.ibge.gov.br/downloads_geociencias.htm> . Acesso em: 24 ago. 2010.

IEF (INSTITUTO ESTADUAL DE FLORESTAS). **Parque Estadual da Mata Seca**, Belo Horizonte. Disponível em: <http://www.ief.mg.gov.br/index.php?option=com_content&task=view&id=204&Itemid=37>. Acesso em: 23 mai. 2007

INPE (INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS). **Divisão de Geração de Imagens: Os satélites LANDSAT 5 e 7**, São José dos Campos. Disponível em: <http://www.dgi.inpe.br/Suporte/files/Cameras-LANDSAT57_PT.php> Acesso em: 21 out. 2016

INPE (INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS). **PROCLIMA: Programa de Monitoramento Climático em Tempo Real da Região Nordeste**. Disponível em: <<http://proclima.cptec.inpe.br>>. Acesso em: 15 jan. 2015

IWAHASHI, J.; PIKE, R. J. Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature. **Geomorphology**, v. 86, n. 3-4, p. 409-440, 2007.

JAFARI, A.; KHADEMI, H.; FINKE, P. A.; VAN DE WAUW, J.; AYOUBI, S. Spatial prediction of soil great groups by boosted regression trees using a limited point dataset in an arid region, southeastern Iran. **Geoderma**, v. 232-234, p. 148-163, 2014.

KALACSKA, M.; SANCHEZ-AZOFEIFA, G. A.; RIVARD, B.; CAELLI, T.; WHITE, H. P.; CALVO-ALVARADO, J. C. Ecological fingerprinting of ecosystem succession: Estimating secondary tropical dry forest structure and diversity using imaging spectroscopy. **Remote Sensing of Environment**, v. 108, n. 1, p. 82-96, 2007.

KEMPEN, B.; BRUS, D. J.; STOORVOGEL, J. J. Three-dimensional mapping of soil organic matter content using soil type-specific depth functions. **Geoderma**, v. 162, n. 1-2, p. 107-123, 2011.

KIM, J.; GRUNWALD, S.; RIVERO, R. G. Soil phosphorus and nitrogen predictions across spatial escalating scales in an aquatic ecosystem using remote sensing images. **IEEE Trans. on Geoscience and Remote Sensing J**, v. 52, n. 10, p. 6724-6737, 2014.

- KIM, J.; GRUNWALD, S.; RIVERO, R. G.; ROBBINS, R. Multi-scale modeling of soil series using remote sensing in a wetland ecosystem. **Soil Sci. Soc. Am. J.**, v. 76, p. 2327-2341, 2012.
- LACOSTE, M.; LEMERCIER, B.; WALTER, C. Regional mapping of soil parent material by machine learning based on point data. **Geomorphology**, v. 133, n. 1–2, p. 90-99, 2011.
- LAGACHERIE, P.; MCBRATNEY, A.B. Spatial soil information systems and spatial soil inference systems: perspectives for Digital Soil Mapping. In: LAGACHERIE, P.; MCBRATNEY, A.B.; VOLTZ, M. **Digital Soil Mapping: An Introductory Perspective**. Amsterdam: Elsevier, 2007, p. 3–24.
- LANDIS, J. R.; KOCH, G. G. The Measurement of Observer Agreement for Categorical Data. **Biometrics**, v. 33, n. 1, p. 159-174, 1977.
- LEMERCIER, B.; LACOSTE, M.; LOUM, M.; WALTER, C. Extrapolation at regional scale of local soil knowledge using boosted classification trees: A two-step approach. **Geoderma**, v. 171–172, p. 75-84, 2012.
- MAYNARD, J. J.; JOHNSON, M. G. Scale-dependency of LiDAR derived terrain attributes in quantitative soil-landscape modeling: Effects of grid resolution vs. neighborhood extent. **Geoderma**, v. 230–231, p. 29-40, 2014.
- MAYNARD, J. J.; LEVI, M. R. Hyper-temporal remote sensing for digital soil mapping: Characterizing soil-vegetation response to climatic variability. **Geoderma**, v. 285, p. 94-109, 2017.
- MCBRATNEY, A. B.; MENDONÇA SANTOS, M. L.; MINASNY, B. On digital soil mapping. **Geoderma**, v. 117, v. 1–2, p. 3-52, 2003.
- MENDONÇA-SANTOS, M. L.; SANTOS, H. G.; DART, R. O.; PARES, J. G. Digital Mapping of Soil Classes in Rio de Janeiro State, Brazil: Data, Modelling and Prediction. In: HATERMINK, A. E.; MCBRATNEY, A. B.; MENDONÇA SANTOS, M. L. **Digital Soil Mapping with Limited Data**. Springer, 2008. cap. 34, p. 381-396.
- MILES, L; NEWTON, A.C.; DEFRIES, R.S.; RAVILIOUS, C; MAY, I.; BLYTH, S.; KAPOS, V.; GORDON, J.E. A global overview of the conservation status of tropical dry forests. **Journal of Biogeography**, v.33, p.491-505, 2006.
- MINASNY, B.; MCBRATNEY, A. B. A conditioned Latin hypercube method for sampling in the presence of ancillary information. **Computers & Geosciences**, v. 32, n. 9, p. 1378-1388, 2006.
- MINASNY, B.; MCBRATNEY, A. B. Incorporating taxonomic distance into spatial prediction and digital mapping of soil classes. **Geoderma**, 142, n. 3–4, p. 285-293, 2007.
- MÖLLER, M.; VOLK, M.; FRIEDRICH, K.; LYMBURNER, L. Placing soil-genesis and transport processes into a landscape context: A multiscale terrain-analysis approach. **Journal of Plant Nutrition and Soil Science**, v. 171, n. 3, p. 419-430, 2008.
- MOONJUN, R.; SHRESTHA, D. P.; JETTEN, V. G.; VAN RUITENBEEK, F. J. A. Application of airborne gamma-ray imagery to assist soil survey: A case study from Thailand. **Geoderma**, v. 289, p. 196-212, 2017.

MOORE, I. D.; GRAYSON, R. B.; LADSON, A. R. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. **Hydrological Processes**, v. 5, n. 1, p. 3-30, 1991.

MORAN, C. J.; BUI, E. N. Spatial data mining for enhanced soil map modelling. **International Journal of Geographical Information Science**, v. 16, n. 6, p. 533-549, 2002.

MURPHY, P.G.; LUGO, A.E. Ecology of Tropical Dry Forest. **Annual Review of Ecology and Systematics**, v.17, p.67-88, 1986.

OLAYA, V.; CONRAD, O. Geomorphometry in SAGA. In: TOMISLAV, H.; HANNES, R. **Geomorphometry: Concepts, Software, Applications**. Elsevier, 2009, p. 293-308.

PEZZINI, F. F.; RANIERI, B. D., BRANDÃO, D. O.; FERNANDES, G. W.; QUESADA, M.; ESPÍRITO-SANTO, M. M.; JACOBI, C. M. Changes in tree phenology along natural regeneration in a seasonally dry tropical forest. **Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology**, v. 148, n. 5, p. 965-974, 2014.

PLANCHON, O.; DARBOUX, F. A fast, simple and versatile algorithm to fill the depressions of digital elevation models. **CATENA**, v. 46, n. 2-3, p. 159-176, 2002.

QGIS. **Documentation QGIS 2.6.** Disponível em:<https://docs.qgis.org/2.6/en/docs/user_manual/processing_algs/saga/terrain_analysis_morphometry/convergenceindex.html>. Acesso em: 21 set.2016

QUINLAN, J. R. **See5: An Informal Tutorial**. Sydney, 2001. Disponível em:<<https://www.rulequest.com/see5-win.html>>. Acesso em: 14.mai.2015

QUINLAN, R. **C4.5: Programs for Machine Learning**, San Mateo: Morgan Kaufmann, 1993.

RAWLINS, B. G.; MARCHANT, B. P.; SMYTH, D.; SCHEIB, C.; LARK, R. M.; JORDAN, C. Airborne radiometric survey data and a DTM as covariates for regional scale mapping of soil organic carbon across Northern Ireland. **European Journal of Soil Science**, v. 60, n. 1, p. 44-54, 2009.

RILEY, S. J.; DEGLORIA, S. D.; ELLIOT, R. A terrain ruggedness index that quantifies topographic heterogeneity. **Intermountain Journal of Sciences**, v. 5, n. 1, p. 23-27, 1999.

ROCSCIENCE INC. **Slope Height.** Disponível em:<https://www.rocscience.com/help/swedge/webhelp/swedge/Slope_Height.htm>. Acesso em: 29 set. 2016

SAMUEL-ROSA, A.; HEUVELINK, G. B. M.; VASQUES, G. M.; ANJOS, L. H. C. Do more detailed environmental covariates deliver more accurate soil maps? **Geoderma**, v. 243-244, p. 214-227, 2015.

SÁNCHEZ-AZOFEIFA, G.A.; CASTRO, K.; RIVARD, B.; KALACSKA, M.; HARRISS, R.C. Remote Sensing Research Priorities in Tropical Dry Forest Environments. **Biotropica**, v.35, p.134-142, 2003.

SÁNCHEZ-AZOFEIFA, G.A.; QUESADA, M.; RODRIGUEZ, J.P. et al. Research priorities for neotropical dry forests. **Biotropica**, v. 37, p. 477 – 485, 2005.

SCHMIDT, K.; BEHRENS, T.; SCHOLTEN, T. Instance selection and classification tree analysis for large spatial datasets in digital soil mapping. **Geoderma**, v. 146, n. 1–2, p. 138-146,. 2008.

SERVIÇO GEOLÓGICO DO BRASIL. **Levantamento aerogeofísico do Estado de Minas Gerais: área 11A (Jaíba – Montes Claros – Bocaiúva)**. Rio de Janeiro: Serviço Geológico do Brasil. 2009. Escala 1: 125.000. CD-ROM.

SHAMY, S.; DHEEBA, J. Review of QUEST, GUIDE, CRUISE, C4.5, and RPART Classification Algorithms. **International Journal of Advanced Tecnology in Engineering and Science**, v. 4, n.6, p. 116-123, 2016.

SILLEOS, N. G.; ALEXANDRIDIS, T. K.; GITAS, I. Z.; PERAKIS, K. Vegetation Indices: Advances Made in Biomass Estimation and Vegetation Monitoring in the Last 30 Years. **Geocarto International**, v. 21, n. 4, p. 21-28, 2006.

TAGHIZADEH-MEHRJARDI, R.; NABIOLLAHI, K.; MINASNY, B.; TRIANTAFILIS, J. Comparing data mining classifiers to predict spatial distribution of USDA-family soil groups in Baneh region, Iran. **Geoderma**, v. 253–254, n. 0, p. 67-77, 2015.

USGS (UNITED STATES GEOLOGICAL SURVEY). **Earth Explorer**. Disponível em:< <https://earthexplorer.usgs.gov>> Acesso em: 15 jan. 2015

USGS (UNITED STATES GEOLOGICAL SURVEY). **NDVI, the Foundation for Remote Sensing Phenology**. Disponível em: <https://phenology.cr.usgs.gov/ndvi_foundation.php>. Acesso em: 23 set. 2016

VALERIANO, M. M. Dados Topográficos. In: FLORENZANO, T. G. **Geomorfologia: conceitos e tecnologias atuais**, São Paulo, Oficina de Textos, 2008, p. 73-104.

VASQUES, G. M.; COELHO, M. R.; DART, R. O.; OLIVEIRA, R. P.; TEIXEIRA, W. G. Mapping soil carbon, particle-size fractions, and water retention in tropical dry forest in Brazil. **Pesquisa Agropecuária Brasileira**, v. 51, n. 9, p. 1371-1385, 2016.

VASQUES, G. M.; DEMATTÊ, J. A. M.; VISCARRA ROSSEL, R. A.; RAMÍREZ LÓPEZ, L.; TERRA, F. S.; RIZZO, R.; DE SOUZA FILHO, C. R. Integrating geospatial and multi-depth laboratory spectral data for mapping soil classes in a geologically complex area in southeastern Brazil. **European Journal of Soil Science**, v. 66, n. 4, p. 767-779, 2015.

VASQUES, G. M.; GRUNWALD, S.; MYERS, D. B. Influence of the spatial extent and resolution of input data on soil carbon models in Florida, USA. **Journal of Geophysical Research: Biogeosciences**, v. 117, n. G4, Não paginado. 2012.

WILFORD, J. A weathering intensity index for the Australian continent using airborne gamma-ray spectrometry and digital terrain analysis. **Geoderma**, v.183–184, n. 0, p. 124-142, 2012.

WILFORD, J.; MINTY, B. The use airborne Gamma-ray imagery for mapping soils and understanding landscape processes. In: LAGACHERIE, P.; MCBRATNEY, A.B.; VOLTZ, M. **Digital Soil Mapping: An Introductory Perspective**. Amsterdam: Elsevier, 2007, p. 206–218.

ZEVENBERGEN, L. W., THORNE, C. R. Quantitative Analysis of Land Surface Topography. **Earth Surface Processes and Landforms**, v. 12, n. 1, p. 47-56, 1987.