UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE AGRONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DO SOLO

CLASSIFICAÇÃO ORIENTADA A OBJETOS E REDES NEURAIS ARTIFICIAIS PARA MAPEAMENTO DIGITAL DE CLASSES DE SOLOS

Fabrício Fernandes Coelho

(Tese de Doutorado)

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE AGRONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DO SOLO

CLASSIFICAÇÃO ORIENTADA A OBJETOS E REDES NEURAIS ARTIFICIAIS PARA MAPEAMENTO DIGITAL DE CLASSES DE SOLOS

FABRÍCIO FERNANDES COELHO
Geógrafo (UFRGS)

Tese apresentada como um dos requisitos à obtenção do Grau de Doutor em Ciência do Solo

Porto Alegre (RS) Brasil
Outubro de 2020

CIP - Catalogação na Publicação

Coelho, Fabrício Fernandes
CLASSIFICAÇÃO ORIENTADA A OBJETOS E REDES NEURAIS ARTIFICIAIS PARA MAPEAMENTO DIGITAL DE CLASSES DE SOLOS / Fabrício Fernandes Coelho. -- 2020.
70 f.
Orientador: Elvio Giasson.

Tese (Doutorado) -- Universidade Federal do Rio Grande do Sul, Faculdade de Agronomia, Programa de Pós-Graduação em Ciência do Solo, Porto Alegre, BR-RS, 2020.

1. GEOBIA. 2. Revisão Sistemática. 3. Fatores de formação do solo. 4. Variáveis preditoras. 5. Classificadores. I. Giasson, Elvio, orient. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

FABRÍCIO FERNANDES COELHO
Bacharel em Geografia - UFRGS
Mestre em Sensoriamento Remoto - UFRGS

TESE

Submetida como parte dos requisitos
para obtenção do Grau de

DOUTOR EM CIÊNCIA DO SOLO

Programa de Pós-Graduação em Ciência do Solo
Faculdade de Agronomia
Universidade Federal do Rio Grande do Sul
Porto Alegre (RS), Brasil

Aprovado em: 27/10/2020
Pela Banca Examinadora

Homologado em: 28/02/2021
Por

ELVIO GIASSON
Orientador-PPG Ciência do Solo

ALBERTO VASCONCELLOS INDA JUNIOR
Coordenador do
Programa de Pós-Graduação em
Ciência do Solo

PAULO CÉSAR DO NASCIMENTO
PPG Ciência do Solo - UFRGS
(através do Mconf da UFRGS)

CLÁUDIO WILSON MENDES JÚNIOR
Departamento de Geodésia/UFRGS
(através do Mconf da UFRGS)

ELPIDIO INACIO FERNANDES FILHO
Departamento de Solos/UFV
(através do Mconf da UFRGS)

CARLOS ALBERTO BISSANI
Diretor da Faculdade
de Agronomia

“O maior risco para um universitário,
mestre ou doutor é se adaptar ao mundo das respostas.
As respostas são o câncer das novas ideias.”

Augusto Cury

Dedico à minha família.

AGRADECIMENTOS

Agradeço a “aquele que é, que era e que há de vir’, que desde o princípio me conduziu e me deu forças para chegar ao fim dessa jornada.

Meus agradecimentos à minha esposa Elisani e aos meus pais Odila e Samuel pelo amor incondicional. Ao meu irmão Cássio, sua esposa Isabel e minha sobrinha Giovana, por todo apoio e carinho de todos os momentos.

Agradeço pela orientação, pelo conhecimento compartilhado, pelas palavras de confiança e de incentivo que me foram dadas pelo professor e amigo Elvio Giasson.

Aos colegas de pós-graduação, em especial ao Alcinei, Anthony, Elisângela, Janderson, Israel, João, Ryshardson e Saulo, pela amizade, parceria e construção de conhecimento.

Aos professores do Programa de Pós-Graduação em Ciência do Solo; à UFRGS por me possibilitar uma sólida formação desde a graduação e à CAPES por financiar parte de meu doutorado.

CLASSIFICAÇÃO ORIENTADA A OBJETOS E REDES NEURAIS ARTIFICIAIS PARA MAPEAMENTO DIGITAL DE CLASSES DE SOLOS [1]

Autor: Fabrício Fernandes Coelho
Orientador: Elvio Giasson

RESUMO

Passados mais de 70 anos do início dos levantamentos sistemáticos de solos brasileiros, mapas semidetalhados ou detalhados ainda são muito escassos no território nacional. Visando maior eficiência para cartografia de solos, a estratégia de mapeamento digital de classes de solos teve início em 2006 no Brasil. Desde então, diversos estudos já foram realizados porém não se percebe uma conexão entre a maioria desses e, até então, não se teve nenhuma tentativa de analisar os estudos agrupando-os por suas características em comum. Além disso, a maioria dos estudos de mapeamento digital de classes de solos brasileiros dependem de um mapa legado para calibrar modelos preditivos, porém os mapas existentes são majoritariamente pobres em detalhes e podem apresentar limitações. Assim, faz-se necessária a busca de alternativas para calibrar modelos preditivos sem a necessidade de mapas legados de solos. O problema é a dificuldade de delinear as unidades de mapeamento com base, somente, em pontos de campo. Sendo assim, o objetivo geral dessa tese foi analisar os aspectos metodológicos mais promissores encontrados em estudos de mapeamento digital de classes de solos no Brasil e, a partir de então, desenvolver proposta metodológica para produção de mapas digitais de solos detalhados e semidetalhados, a partir de pontos levantados em campo, baseada em classificação orientada a objetos (GEOBIA) e Redes Neurais Artificiais. Na revisão sistemática apresentada no CAPÍTULO III foram encontrados 334 estudos publicados em 42 artigos, em território nacional, que mostram evidências de que o tamanho de pixel adequado com escala de estudo e utilização de variáveis preditoras relacionadas ao maior número de fatores de formação do solos são importantes para se obter melhores resultados. Por outro lado, quanto maior a densidade de unidades de mapeamento por unidade de área, menores são as acurácias obtidas. Além disso, ficou evidenciado que os métodos de Redes Neurais Artificiais e de classificadores em árvore são os mais promissores dos métodos utilizados. No CAPÍTULO IV é apresentada uma nova metodologia baseada GEOBIA e Redes Neurais Artificiais para mapeamento de solos em escala detalhada utilizando somente de pontos de campo. Os mapas preditos por GEOBIA tiveram melhores resultados que os métodos tradicionais baseados em pixels.

Palavras-chave: GEOBIA, revisão sistemática, fatores de formação do solo, variáveis preditoras, classificadores

[1] Tese de Doutorado em Ciência do Solo. Programa de Pós-Graduação em Ciência do Solo, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul. Porto Alegre. (70p.) Outubro, 2020.

GEOGRAPHIC OBJECT-BASED IMAGE ANALYSIS AND NEURAL ARTIFICAL NETWORK FOR DIGITAL SOIL CLASS MAPPING [2]

Author: Fabrício Fernandes Coelho
Adviser: Elvio Giasson

ABSTRACT

More than 70 years after the beginning of systematic surveys of Brazilian soils, semi-detailed or detailed maps are still very scarce in the national territory. In order to increase efficiency in soil cartography, the digital mapping of soil classes approach began in 2006 in Brazil. Several studies have been carried out but there is no connection between the majority of these and, until then, there has been no attempt to analyze the studies grouping them by their common characteristics. In addition, most studies of digital mapping of soil classes depend on a legacy map to calibrate predictive models, however the existing maps are poor in detail and may have limitations. Thus, it is necessary to search for alternatives to calibrate predictive models without the legacy soil maps. The problem is the difficulty of to delimit the mapping units based only on field points. Therefore, the objective of this thesis was to analyze the most promising methodological aspects found in studies of digital mapping of soil classes in Brazil and to develop a methodological proposal for the production of detailed and semi-detailed digital maps, based on field point, Geographic Object-Based Image Analysis (GEOBIA) and Artificial Neural Networks. In the systematic review presented in CHAPTER III, 334 studies were found in 42 articles, in the national territory, which show evidence that the appropriate pixel size with study scale and use of predictor variables related to the greater number of soil-formation factors are found important to obtain better results. On the other hand, the higher the density of mapping units per unit area, the lower the accuracy obtained. In addition, it was shown that the methods of Artificial Neural Networks and Decision Tree Classification are the most promising of the methods used. CHAPTER IV presents a new methodology based on GEOBIA and Artificial Neural Networks for soil mapping on a detailed scale using only field points. The maps predicted by GEOBIA had better results than the traditional pixel-based methods.

Keywords: GEOBIA, Systematic Review, soil-forming factors, predictor variables, classifiers

---

[2] Doctoral thesis in Soil Science. Graduate Program in Soil Science, Faculty of Agronomy, Universidade Federal do Rio Grande do Sul. Porto Alegre. (70p.) October, 2020.

SUMÁRIO

RELAÇÃO DE TABELAS

RELAÇÃO DE FIGURAS

RELAÇÃO DE ABREVIATURAS

ADGV - Aquisição de Dados Geoespaciais Vetoriais

ALT - Altimetry

ANN - Artificial Neural Network

ANOVA – Analysis of Variance

CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

CART - Classification and Regression Trees

CI - Confidence Interval

CURV - Curvature

CURVH - Horizontal Curvature

CURVV - Vertical Curvature

DEM - Digital Elevation Model

DSM - Digital Soil Mapping

DT - Decision Tree

DTM - Digital Terrain Model

ED - Euclidean Distance

ET - Especificação Técnica

FA - Flow Accumulation

GEOBIA - Geographic Object-Based Image Analysis

GNSS - Global Navigation Satellite System

HAND - Height Above the Nearest Drainage

IA – Inteligência Artifical

LMT - Logistic Model Trees

LR - Logistic Regression

MDS – Mapeamento Digital de Solo

MLP – Multilayer Perceptron

MRS - Multiresolution Segmentation

MU - Mapping Unit

NA - Not Available

NDVI - Normalized Difference Vegetation Index

NDWI - Normalized Difference Water Index

OA - Overall Accuracy

PA - Producer’s Accuracy

PC - Principal Component

PCA - Principal Component Analysis

PCD - Produtos Cartográficos Digitais

PDI – Processamento Digital de Imagens

PRONASOLOS – Programa Nacional de Solos

RADAM - Radar da Amazônia

SciELO - Scientific Electronic Library Online

SIG – Sistema de Informação Geográfica

SP - Scale Parameter

SSPFe - Soil spatial prediction function with spatially autocorrelated erros

SVM - Support Vector Machine

TWI - Topographic Wetness Index

UA - User’s Accuracy

CAPÍTULO I - INTRODUÇÃO GERAL

O Brasil tem a totalidade de seu território coberto por mapas de solos exploratórios ou de reconhecimento de baixa intensidade que variam na escala de 1:250.000 até 1:1.000.000 (SANTOS *et al.*, 2013), sendo que parte desses mapas são produtos de levantamentos iniciados na década de 1950, quando o conhecimento a respeito dos solos nacionais era pequeno. Mapas de reconhecimento de média ou alta intensidade com escala de 1:100.000 cobrem aproximadamente 5% do território nacional ao passo que mapas de reconhecimento de alta intensidade ou semidetalhado, com escala de 1:50.000, cobrem aproximadamente 1,5% do território. Mapas semidetalhados ou detalhados, com escala igual ou superior a 1:25.000 são escassos, com área de cobertura inferior a 0,2% do território brasileiro (SANTOS *et al.*, 2013). A escassez de mapas de solos detalhados pode ser explicada pela grande área territorial brasileira, má distribuição de estradas que restringem o acesso a diversas partes do país e, principalmente, pela falta de recursos. A utilização de métodos convencionais de mapeamento de solos com maior detalhamento demanda alto custo monetário e de tempo para que os mapeamentos sejam confeccionados.

É nesse contexto brasileiro que, em 2018, foi instituído o PRONASOLOS (Programa Nacional de Solos), que tem por objetivo retomar as atividades de levantamentos pedológicos que foram iniciadas na década de 1950 e que foram descontinuadas a partir da década de 1980, com os levantamentos executados no âmbito do Projeto RADAMBRASIL. A partir dos levantamentos a serem realizados no âmbito do PRONASOLOS, estima-se que em até 30 anos se alcance mais de 6,9 milhões de km² em escala 1:100.000, 1 milhão de km² mapeados na escala 1:50.000 e 250 mil km² em escala 1:25.000

(POLIDORO *et al.*, 2016). Considerando-se o volume de trabalho de campo e a quantidade de dados que serão obtidas pelo PRONASOLOS, as técnicas de mapeamento digital de solos (MDS), que já estão sendo testadas a mais de uma década pela comunidade científica brasileira, com a finalidade de subsidiar a falta de mapas de solos mais detalhados, podem auxiliar no delineamento das unidades de mapeamento de solos e ajudar a identificar feições e reconhecer padrões que podem passar desapercebidos pela interpretação visual do pedólogo.

No mapeamento digital de solos, técnicas de inteligência artificial, como as redes neurais artificiais, podem ser utilizadas para descobrir o conhecimento embutido nos mapas legados de solos, confeccionados subjetivamente por métodos convencionais, e então reproduzir esses padrões para áreas não amostradas. Apesar de que diversos estudos de mapeamento digital de classes de solos já foram realizados em território brasileiro, não se percebe uma conexão entre a maioria dos trabalhos. Essa falta de conexão entre os trabalhos se percebe tanto pelas diferenças geográficas de áreas de estudos quanto de métodos e técnicas utilizadas. Até o momento desse trabalho, não se teve nenhuma tentativa de analisar os estudos agrupando-os por suas características em comum.

Em uma revisão realizada em 2012 (TEN CATEN *et al.*, 2012), os autores destacaram a necessidade de evolução quanto a uma padronização metodológica, tanto no no que diz respeito à execução quanto à avaliação dos resultados de estudos de MDS. De toda forma, os autores não realizaram uma análise estatística mais aprofundada dos trabalhos de MDS para apontar os fatores promissores e aqueles que não deram resultados positivos para o mapeamento digital de classes de solos no Brasil. Outro fator que não foi considerado pelos autores é que a grande maioria dos estudos brasileiros dependem de um mapa legado para que sejam calibrados os modelos preditivos. Em uma revisão bibliométrica realizada em 2018 (CANCIAN; DALMOLIN; CATEN, 2018), os autores buscaram caracterizar a produção brasileira em mapeamento digital de solos no período de 1996 até 2017, porém a pesquisa foi focada em quantitativos somente da evolução de trabalhos publicados e não nas características dos métodos empregados nos estudos e seus resultados.

No presente trabalho, foram avaliadas novas estratégias tanto para redução de erros nos resultados da predição espacial de solos, em especial para mapas mais detalhados, quanto para calibrar modelos sem a necessidade de um mapa de solos existente. Foi utilizada a abordagem de classificação orientada a objetos (GEOBIA - *Geographic Object-Based Image Analysis*), uma técnica que não foi utilizada em nenhum estudo brasileiro até o momento e é incipiente em estudos de DSM a nível global.

Com base no contexto apresentado, nessa tese buscou-se primeiramente levantar os dados dos trabalhos já publicados de mapeamento digital de classes de solos realizados no Brasil a partir de uma série de critérios de consulta. Os resultados desse levantamento são apresentados no CAPÍTULO III. A partir da análise dos resultados encontrados, foi proposta uma nova abordagem metodológica que integra técnicas de Geoprocessamento, Processamento Digital de Imagens (PDI) e de Inteligência Artificial (IA) para mapeamento digital de classes de solos, apresentado no CAPÍTULO IV. Sendo assim, o objetivo geral dessa tese foi analisar os aspectos metodológicos mais promissores encontrados em estudos de mapeamento digital de classes de solos no Brasil a partir de uma revisão sistemática e, a partir de então, desenvolver proposta metodológica para produção de mapas digitais de solos detalhados e semidetalhados a partir de pontos levantados em campo, baseada em classificação orientada a objetos e redes neurais artificiais.

CAPÍTULO II - REVISÃO BIBLIOGRÁFICA

1. Mapeamento digital de solos

O mapeamento de solos envolve a determinação da localização e distribuição dos diferentes solos que ocorrem numa região, coleta de informações sobre sua localização, natureza, propriedades, potencial e usos. Os métodos convencionais baseiam-se nas relações entre solos e fisiografia (IPPOLITI R. *et al.*, 2005). Tais métodos utilizam insumos cartográficos e produtos de sensoriamento remoto que vão desde pares de ortofotos para observação estereoscópica, modelos digitais de elevação, imagens de satélite multi e hiperespectrais para subsidiar análises e interpretação visual das feições do relevo e paisagem. Na etapa de levantamento de campo, com apoio de instrumentos de navegação como dispositivos GNSS (*Global Navigation Satellite System*), aplicativos móveis ou mesmo mapas analógicos, os tipos de solos são avaliados e uma relação solo-paisagem é definitivamente estabelecida e que serve de base para o mapeamento dos solos.

Conforme observado por Minasny e McBratney (2016) em uma revisão sobre mapeamento digital de solos, em meados da década de 1970 mapas confeccionados por métodos convencionais que passaram por processo de digitalização por meio da utilização de SIG's foram rotulados como mapas digitais de solos. Contudo, apesar do mapa ser digital, o processo de produção não era baseado em modelos matemáticos e estatísticos de inferência, então esses não devem ser considerados mapas digitais de solos, mas sim mapas digitalizados de solos.

O mapeamento digital de solos pode ser definido como a criação e a inserção de dados em sistemas de informações espaciais de solos por meio de métodos de observação de campo e laboratório combinados com sistemas de inferência espacial e não espacial de solos (LAGACHERIE; MCBRATNEY, 2007). Para tanto, os sistemas de informações espaciais de solos devem apresentar três principais componentes (Lagacherie e Mcbratney, 2007; Minasny e McBratney, 2016):

1) Entrada de dados: consiste na inserção de dados no sistema por métodos de observação de campo e de laboratório. Estão inclusas informações de levantamento de solo, mapas legados, novos dados por meio técnicas estatísticas de amostragem e novos dados de campo.
2) Processo: consiste nos modelos de inferência espacial e não espacial de solos, como métodos matemáticos e estatísticos, que relacionam as observações de solos com as variáveis ambientais.
3) Saída: consiste na informação espacial resultante, que inclui mapas e planos de informações matriciais que representam a predição espacial de solos com o erro conhecido.

De acordo com Lagacherie e Mcbratney (2007), o mapeamento digital de solos se baseia na geração de sistemas de informações que permitem estabelecer relações matemáticas entre variáveis ambientais e classes de solos e, assim, predizer a distribuição espacial das classes de solos. Dessa forma, a estruturação de uma base de dados espacial e não espacial de solos em conjunto com métodos computacionais de inferência em sistema de informação geográfica (SIG) assume grande importância no mapeamento digital de solos. A grande vantagem do mapeamento digital de solos é produzir novos mapas pedológicos, com erro conhecido, para áreas que não têm informações sobre a distribuição espacial das classes de solos. Dentre as abordagens que estão sendo testadas de MDS, destacam-se o mapeamento digital de atributos do solo, mapeamento digital de classes de solos e mapeamento digital de classes de solos por desagregação de polígonos, que tem por foco o aumento do detalhamento de mapas convencionais legados de escala pequena.

Para estudos de mapeamento digital de solos envolvendo predição de classes e atributos de solos, McBratney et al. (2003), estabeleceram o modelo genérico chamado de scorpan-SSPFe (*soil spatial prediction function with spatially autocorrelated erros* – função de precição espacial de solo com erros espacialmente correlacionados). Para estudos envolvendo a predição de classes de solos, o modelo scorpan-SSPFe é representado pela Equação 1.

$$S_c = f\ (s, c, o, r, p, a, n) + e \qquad \text{Equação 1}$$

Onde S_c são as classes de solos em função do solo (s), clima (c), organismos (o), relevo (r), material de origem (p), tempo (a), posição espacial (n) e erros espacialmente correlacionados (e). O modelo scorpan é utilizado para predição quantitativa de classes de solos e atributos do solo e não tem por objetivo tentar explicar os fatores de formação do solo, como é o caso do modelo fatorial de Jenny (1941) (MINASNY; MCBRATNEY, 2016).

Conforme Arrouays et al. (2017), a avaliação da incerteza da predição de solos é o aspecto que mais demanda progresso para o mapeamento digital de solos. Conforme os mesmos autores, em alguns estudos, os mapas digitais de solos contêm consideráveis graus de incertezas o que os leva a ter pequeno uso prático para os usuários finais desses produtos.

Como exemplo de aplicação do mapeamento digital de solos, Giasson et al., (2006) realizaram estudo na região sul do Brasil; os resultados chegaram a apresentar até 71% de coincidência (acurácia global) nos mapas produzidos com o mapa original, utilizando-se do método de regressões logísticas múltiplas multinomiais. Também utilizando-se do modelo logístico ten Caten et al. (2011c) realizaram mapeamento digital de solos com aplicação de componentes principais. Os resultados mostraram a aplicabilidade do uso de análise de componentes principais para redução das variáveis e análise da correlação entre as mesmas e ainda consideraram satisfatórios os resultados do mapeamento.

Coelho e Giasson (2010) compararam cinco métodos de classificação, sendo eles regressões logísticas múltiplas multinomiais, classificador de Bayes e três métodos de classificação em árvore (J48, CART – *Classification and Regression Trees* – e LMT – *Logistic Model Trees*). Os resultados mostraram que os classificadores em árvore apresentaram melhores acurácias e não tiveram grande variação entre os três algoritmos testados, além disso, o estudo mostrou que a utilização de legenda simplificada proporciona maior confiança dos mapas preditos pelo aumento das acurácias.

Utilizando-se de métodos hierárquicos de classificação, (BEHRENS, T.; SCHOLTEN, 2007), testaram os métodos de árvores de decisão e *support vector machine* (SVM) para o mapeamento digital de solos. Segundo os autores, os melhores resultados ocorreram nos métodos de árvores de decisão e concluíram que o método SVM não é adequado para mapeamento digital de solos pois retorna baixas acurácias e o processamento dos dados é lento.

Em relação aos dados utilizados, majoritariamente são utilizados dados provenientes de sensoriamento remoto. De toda forma, em estudos mais recentes, percebe-se a tentativa de integração entre dados de sensores remotos e de sensores próximos para modelagem pedológica em escalas cada vez mais detalhadas (Godinho Silva et al., 2016; Rizzo et al., 2016; Demattê et al., 2016; Demattê et al., 2017; Poppiel et al., 2019; Silvero et al., 2019).

1.1 Estratégias de mapeamento digital de classes de solos sem mapas legados

Em estudos de mapeamento digital que necessitam mapas legados para calibrar modelos preditivos devem ser consideradas as limitações dos mapas existentes. Muitas vezes os mapas existentes foram confeccionados com dados obsoletos e os polígnos das unidades de mapeamento não necessariamente apresentam a verdadeira distribuição dos solos pois são produtos de modelos mentais baseados em altos níveis categóricos do sistema de classificação utilizado (ARROUAYS *et al.*, 2020). Assim, surge a necessidade de buscar alternativas sem a dependência da utilização de mapas legados, isto é, utilizar observações de campo de perfis de solos. Um dos principais problemas encontrados é a baixa quantidade e, por consequência, a pequena representatividade espacial das amostras de campo (CAMPOS *et al.*, 2019a).

Na abordagem orientada a pixels, com a utilização de pontos georreferenciados de perfis de solos, cada ponto fica restrito a apenas um *pixel* por variável, assim sua representação espacial é de uma pequena área de mesma abrangência da resolução espacial (tamanho do pixel) das variáveis utilizadas (CHEN *et al.*, 2018). Quanto maior for o detalhamento do mapa, maior resolução espacial das variáveis é necessária, isto é, menor tamanho de *pixels* e consequentemente menor representatividade territorial da unidade amostral dos perfis de solos.

Considerando essas limitações, alguns autores utilizaram a estratégia de buffers no entorno de pontos georreferenciados de perfis de solos para proceder com a calibração de modelos preditivos. Pelegrino et al. (2016) realizou a predição de classes de solos a partir de buffers no entorno de pontos de campo com raios de 25 m, 50 m, 75 m e 100 m. Com isso, o ponto amostral foi extrapolado para uma área de aproximadamente 1.960 m² (com buffer de 25 m), 7.850 m² (com buffer de 50 m), 17.670 m² (com buffer de 75 m) e de 31.415 m² (com buffer de 100 m). Campos et al., (2019a) testaram cinco raios de buffers no entorno de pontos de campo, sendo de 50 m, 100 m, 150m, 200 m e 250 m. Com isso o ponto amostral foi extrapolado para uma área de aproximadamente 7.850 m² (com buffer de 50 m), 31.415 m² (com buffer de 100 m), 70.685 m² (com buffer de 150 m), 125.663 m² (com buffer de 200 m) e 196.345 m² (com buffer de 250 m). Machado et al. (2019) utilizaram buffers de 30 m no entorno dos pontos georreferenciados, extrapolando a informação do ponto para uma área de aproximadamente 2.827 m².

Em todos os estudos citados, os autores obtiveram melhores resultados a partir da extrapolação do ponto amostral para uma área de entorno a partir da estratégia dos buffers, todavia um dos problemas ao se utilizar um raio com tamanho específico é que as características ambientais do entorno dos pontos são ignoradas na delimitação do *buffer*. Assim, se um ponto amostral estiver, por exemplo, no limite entre uma área plana e uma declivosa, a poligonal do *buffer* irá contemplar os valores dos pixels dessas duas posições topográficas e relacioná-las com a classificação de um perfil de solo, o que pode acarretar erros na classificação. Dessa forma, as técnicas baseadas em classificação orientada a objetos surjem como uma alternativa a ser utilizada pois produzem polígonos que consideram as características ambientais do terreno.

2. Classificação orientada a objetos

Em SIG, os modelos de representação espacial se dividem em modelo de dado matricial e modelo de dado vetorial. No modelo de dado vetorial, a representação das feições espaciais é definida por entidades com geometria específica de pontos, linhas ou polígonos e suas coordenadas são consideradas matematicamente exatas (BURROUGH, 1986). No modelo de dado matricial, as feições são representadas por células ou conjunto de células adjacentes uma a outra e é o modelo mais indicado para representar variáveis contínuas no espaço (elevação, declividade, curvatura, por exemplo). No modelo matricial, cada plano de informação representa uma variável onde cada célula ou *pixel* (*picture element*) recebe um valor numérico, o qual é utilizado para realizar as operações matemáticas e análises espaciais.

Nos métodos de classificação digital, os algoritmos de classificação extraem automaticamente feições por meio do reconhecimento de padrões representados em classe única (VIEIRA *et al.*, 2012). A classificação digital pode ser feita baseada em *pixels* ou em objetos. Na classificação baseada em *pixels*, os valores digitais dos *pixels* são utilizados como unidade mínima de classificação, já na classificação orientada a objetos, os *pixels* são agrupados em regiões mais homogêneas entre si, tornando os segmentos como unidades mínimas de classificação (Blaschke, 2010; Myint et al., 2011).

A segmentação é o particionamento de uma imagem em regiões até a obtenção do isolamento dos objetos de interesse ou regiões de interesse (GONZALES; WOODS, 2006). Dessa forma, a segmentação de imagens aproxima-se mais da forma como o cérebro humano faz o reconhecimento de padrões e a classificação dos alvos. Segmentos são regiões as quais são geradas por um ou mais critérios de homogeneidade em um ou mais dimensões. Esses segmentos têm informações adicionais se comparados aos pixels unitários, ou seja, média de valores, valores mínimos e máximos, variância, mas a maior vantagem é a informação espacial adicional dos objetos (BLASCHKE, 2010). A Figure 1 apresenta um exemplo dos objetos sobrepostos a um modelo 3D do terreno com diferentes tamanhos (níveis de homogeneidade).

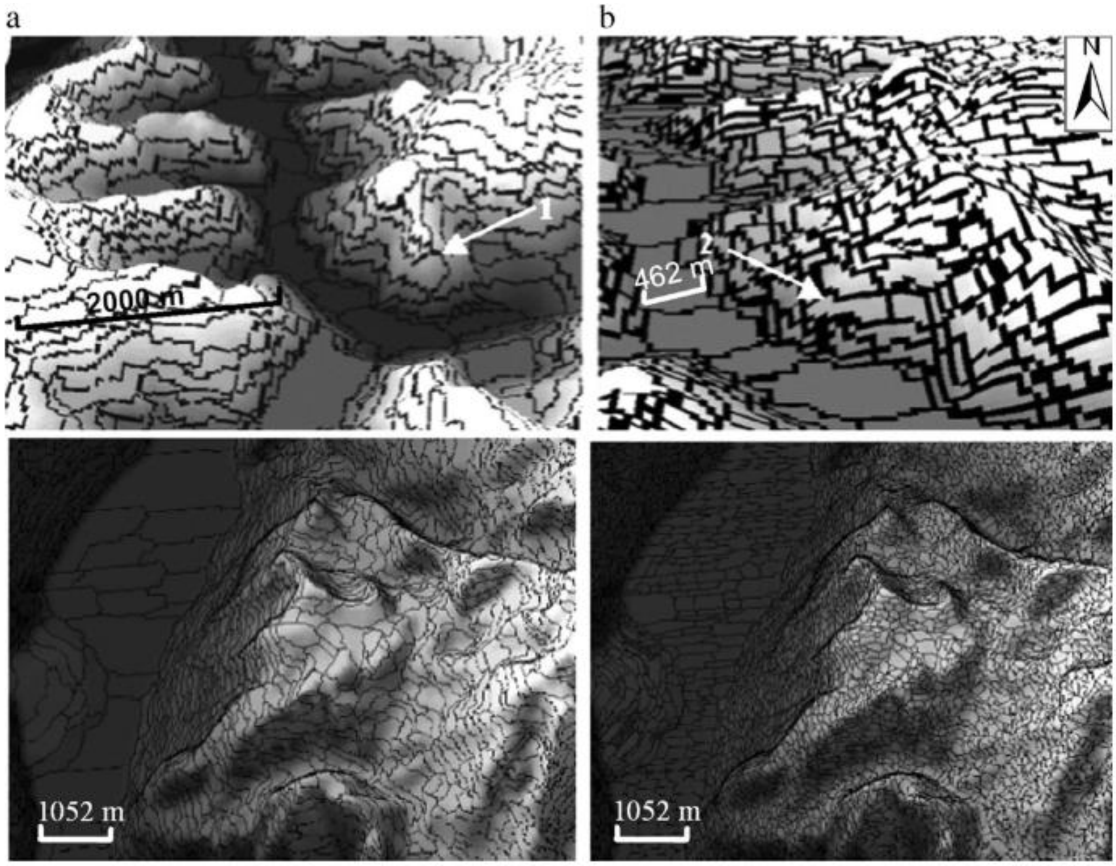

Figure 1. Exemplo de objetos com diferentes níveis de homogeneidades apresentados sobre uma superfície 3D do terreno. (Reproduzido de DRĂGUŢ; BLASCHKE, 2006).

Segundo Schultz, (2016), a classificação orientada a objetos possui as seguintes vantagens em relação aos métodos baseados em pixels: a) reduz os ruídos inerentes ao processo de classificação pixel a pixel, b) melhora a relação entre os valores dos atributos de cada segmento com as variáveis e c) reduz o tempo de processamento da classificação devido à redução do número de objetos a serem classificados.

A classificação orientada a objetos é uma técnica bastante difundida em classificação de imagens de sensoriamento remoto para delimitação das classes de uso e cobertura das terras. porém ainda é novidade para mapeamento digital de solos. Até o momento desse estudo, somente foi identificado um artigo com essa abordagem para mapeamento digital de solos (DORNIK; DRÂGUŢ; URDEA, 2017). Nesse estudo, os autores utiliaram 171 pontos georreferenciados de perfis de solos associados com 23 variáveis preditoras e compararam os resultados de classificação a partir da abordagem orientada a pixels e objetos, utilizando o classificador Random Forest. Os autores obtiveram valor de acurácia de 58% com a abordagem orientada a objetos, 10% a mais do que obtiveram com a abordagem orientada a pixels.

Estudos de sensoriamento remoto que compararam métodos de classificação orientada a objetos com métodos baseados em *pixels* mostraram que a classificação orientada a objetos apresentou maior capacidade na separação de diferentes classes (Desclée et al. 2006; Fernandes et al. 2011). Conforme Desclée et al. (2006) foi obtida maior acurácia na detecção de mudanças de cobertura florestal com o método baseado em objetos, em relação aos métodos baseados em *pixels*. Ao comparar a classificação orientada a objetos com a baseada em *pixels*, Fernandes et al. (2011) verificaram melhores resultados para detectar áreas de cobertura vegetal alterada com a classificação orientada a objetos.

Além de ser utilizada para classificação de uso e cobertura das terras, a classificação orientada a objetos também está sendo utilizada para classificação de formas do terreno a partir de Modelos Digitais de Elevação (MDEs) (Drăguţ and Blaschke, 2006; Gerçek et al., 2011; Drăguţ and Eisank, 2012; Martha et al., 2018).

3. Redes Neurais Artificiais

As redes neurais artificais são consideradas como uma das técnicas de inteligência artificial e são uma abstração das redes neurais biológicas, ou seja, o neurônio biológico. O neurônio biológico é composto por um núcleo, axônio, dentritos e sinapses que são as junções entre os neurônios. Um neurônio artificial modela os axônios e dentritos por meio de conexões e as sinapses utilizando de ponderações ou pesos de ajuste (COSTA, H. R. do N., 2006) (**Figure 2**).

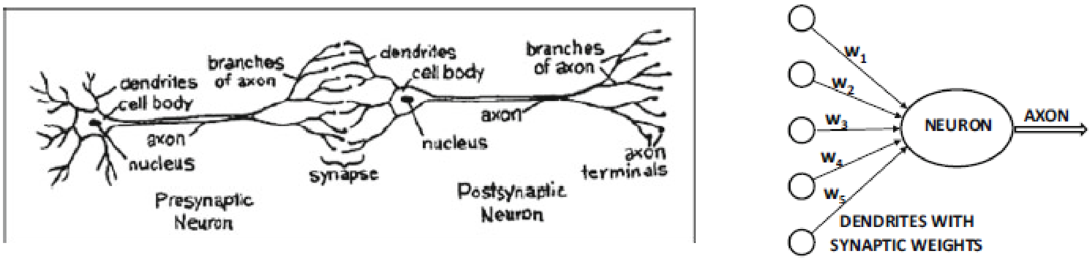

Figure 2. Comparação da rede neural biológica e artificial. (Reproduzido de AGGARWAL, 2018)

Uma rede neural artificial calcula uma função das entradas, propagando os valores calculados dos neurônios de entrada para os neurônios de saída, utilizando os pesos como parâmetros intermediários. A aprendizagem ocorre mudando os pesos conectados aos neurônios. Existem diversas tipologias de redes neurais e, aquela utilizada no presente trabalho foi a rede Perceptron Multicamada (multilayer perceptron) (MLP) que consiste e um sistema de neurônios interconectados como ilustra a **Figure 3**.

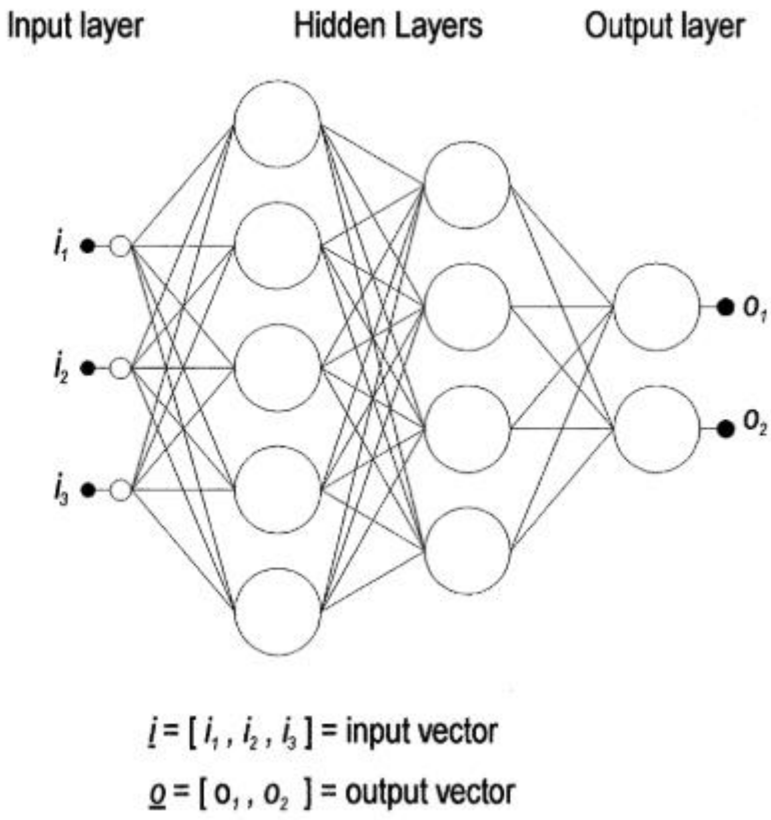

Figure 3. Estrutura de uma rede Perceptron Multicamada. (Reproduzido de GARDNER; DORLING, 1998).

No caso de estudos de mapeamento digital de classes de solos, a camada de entrada (input layer) da MLP consiste nos valores das variáveis preditoras das amostras e a camada de saída (output layer) consiste nas classes de solos. As camadas escondidas (hidden layers) consistem em neurônios intermediários para o cálculo das relações não lineares.

Em estudos de mapeamento digital de classes de solos, Calderano Filho et al. (2014) utilizaram redes neurais artificiais para predição de tipos de solo utilizando variáveis espectrais de produtos de sensoriamento remoto, variáveis geomorfométricas e geologia em uma área de alto grau de diversidade litológica da Serra do Mar. Os autores confirmaram o potencial de se utilizar redes neurais artificiais para predição de classes de solos pois os resultados apresentaram dados consistentes e similares aos mapas produzidos por métodos convencionais, porém com maior detalhamento espacial, com reultados variando de 93 a 95% de acurácia global e de 0.92 a 0.95 de índice Kappa.

Estudo desenvolvido por Arruda et al., (2016) teve como objetivo a avaliação da eficiência das redes neurais artificiais para a extrapolação de classes de solos para áreas adjacentes não mapeadas, a partir de um mapa legado detalhado de solos. Os autores verificaram que o mesmo padrão de distribuição de classes de solos do mapa legado foi estimado para as áreas não mapeadas e concluíram que a técnica apresenta potencial para ser utilizada no mapeamento digital de solos, chegando a concordância de mais de 80% entre os pontos de referência e o mapa digital de solos.

Combinando imagens de satélite de alta resolução espacial (imagens Quickbird) com variáveis do terreno, Chagas et al. (2011) obtiveram resultados de acurácias próximo a 70% no mapeamento de tipos de solos com o uso de redes neurais artificiais. Khaledian and Miller, (2020) compararam diversos algoritmos e descobriram que as Redes Neurais Artificiais (RNA’s) apresentam os melhores resultados na predição de propriedades de solos quando utilizada uma grande amostra.

Uma das vantagens em utilizar uma rede Perceptron Multicamada é que, ao contrário de outras estatísticas, a MLP não depende pressupostos sobre a distribuição de dados e pode modelar funções altamente não lineares (GARDNER; DORLING, 1998). Além disso, a MLP tende a apresentar bons resultados de acurácia, porém pode ser considerada como uma “caixa preta” e a interpretação da sua modelagem é difícil (KHALEDIAN; MILLER, 2020). Dessa forma, ao escolher um método de classificação a ser utilizado, os usuários devem estar cientes de que, se a interpretação da modelagem é importante, então a rede neural nao é o método mais indicado. Por outro lado, se o usuário está mais interessado no resultado do que na interpretação da modelagem, a MLP pode ser considerada para utilização.

CAPÍTULO III – DIGITAL SOIL CLASS MAPPING IN BRAZIL: A SYSTEMATIC REVIEW [3]

1. Introduction

Soil maps are crucial for environmental and agricultural management but conventional soil mapping is costly and time-consuming and the existing soil maps are lack details. In the past decades, digital soil mapping (DSM) methods are being tested and analyzed by the scientific community to maximize the use of existing maps and information and to provide estimates for wider areas.

There is geographic variation in the uptake of digital soil mapping technologies and some countries have made considerable progress. However, Minasny e McBratney (2016) and Arrouays et al. (2017) found that methodologies and assessment of DSM results need to be standardized, errors should be minimized and better evaluated, and strategies devised to overcome the lack of detailed cartographic bases and dearth of soil maps and data.

Recent reviews indicate others challenges such as soil mapping at the flat terrains, simulation of the soil heterogeneity at the regional scale, linking DSM and soil spectroscopy (ZHANG; LIU; SONG, 2017) and the use process-based soil-landscape evolution modelling with interactions between pedology and DSM (MA *et al.*, 2019).

Studies of digital soil class mapping in Brazil began in 2006 (GIASSON, Elvio *et al.*, 2006) were re-analyzed six years later by

---

[3] Artigo publicado na Scientia Agricola. COELHO, Fabrício Fernandes et al . Digital soil class mapping in Brazil: a systematic review. Sci. agric. (Piracicaba, Braz.), Piracicaba , v. 78, n. 5, e20190227, 2021.

ten Caten et al. (2012) that found an agreement average of 48 % (Kappa coefficient) in 11 articles. The overall accuracy (OA) was not analyzed and the classifier most used were logistic regressions, but there is not consensus on which methods have better results in the prediction soil maps Cancian et al., (2018) conducted a bibliometric analysis of the scientific production of DSM in the period from 1996 to 2017 and perceived that publications are increasing and that Brazilian research is gaining prominent position in the world scenario. The authors found approximately 200 researchers working with DSM in Brazil.

Those who intend to produce digital maps of soil classes find publication that test different classifiers, sets of predictor variables and sample size. However, there is no study that presents the data from these publications in an integrated and systematic way.

In this study we sought to produce information from data from several publications to assist in decision making for those who want to produce digital soil maps. The aim was to analyze the factors used in digital soil class mapping and to assess the accuracy of the studies based on a systematic review of articles published between 2006 and 2019 in Brazil.

2. Material and Methods

2.1 Bibliographic Survey

The following criteria were used for a systematic survey of articles (inclusion criteria): a) study area in Brazil, to analyze results of the Brazilian approach in studies of digital mapping of soil classes; b) period: from 2006 to 2019; c) object of study: digital soil class mapping; d) articles with quantitative validation; and e) articles that used supervised learning methods as: Support Vector Machines, Artificial Neural Networks, naïve Bayes, Logistic Regression and Decision Trees.

The following articles were excluded (exclusion criteria): a) articles whose object of study was the digital mapping of soil attributes or soil polygon disaggregation mapping; b) articles with no spatial soil class prediction using classification algorithms (e.g., map algebra) and/or mapping units (MU) delimited manually; c) articles with only qualitative validation of spatial soil class prediction;

and d) articles that used unsupervised classification methods (clustering) and Fuzzy because the process of modelling is very different from supervised learning methods that are the focus of this study.

Two electronic libraries were used for this study a) *Portal de Periódicos CAPES* (Coordination for the Improvement of Higher Education Personnel) and b) SciELO (Scientific Electronic Library Online). These are electronic libraries financed by public funds to promote free access of scientific journals. The *Portal de Periódicos CAPES* offers free access for professors, researchers, students and employees of participating institutions such as all federal institutions of Brazilian higher education among others. The SciELO promotes free public access to scientific journals from developing countries. Currently, *Portal de Periódicos CAPES* offers access to complete texts available in more than 38,000 national and international journals, while SciELO has 1,285 active journals.

For the moment and objectives of this study, the research in these two electronic libraries proved to be adequate. However, researchers are being encouraged to internationalize Brazilian research and for future studies it is worth considering the consultation in other databases as (e.g., Web of Science and Scopus).

The following strings and filters were used: a) topic: “Mapeamento Digital de Solos”; type of resource/literature: “article”; period: “2006 – 2016” (Surveys 1 and 3) and b) topic: “Digital Soil Mapping AND Brazil”; type of resource/literature: “article”; period: “2006 – 2016”; (Surveys 2 and 4).

All the articles from de survey were tabulated and analyzed. If an article met the inclusion criteria then it was included for participation in this study; if an article met the exclusion criteria then it was not included. The Mendeley Reference Manager was used for bibliographic management.

2.2 Database

For the construction of the database, data of all studies contained in the articles were extracted. We consider all the soil class prediction tests carried in an article as studies; e.g., if in an article three learning algorithms were compared for prediction soil classes, then these tests were considered as three studies. The same occurs if two or more sets of predictor variables were compared or any tests that were performed that present the respective validation values. Therefore, there is a relationship “one-to-many”; i.e., one article to many studies.

The following quantitative and qualitative data were extracted from the studies of the selected articles: a) year of publication; b) reference city of the study area; c) size of study area (km^2); d) cartographic scale; e) number of mapping units; f) number of samples used in the predictive models (pixels from raster data of legacy map and/or points of fields observations); g) digital elevation model (DEM) used; h) pixel size (m); i) number of predictor variables used; j) learning algorithms; k) overall accuracy (%); and l) Kappa coefficient (%).

2.3 Data grouping

2.3.1 Reproducibility and exactness assessment groups

The studies were assigned to the reproducibility assessment group when they were validated using conventional maps; when they were validated using points of fields observations then were assigned to the exactness assessment group.

2.3.2 Soil-forming factors

All the predictor variables used in the selected articles were extracted and assigned to a soil-forming factor attribute like climate, parent material, organisms, relief and time (**Table 1**). Thus, it was possible to calculate the number of soil-forming factors used in each study.

Table 1. Relation between predictor variables and soil-forming factors.

Soil-Forming Factors	**Predictor Variables**
Climate	Aspect, Diffuse Insolation, Direct Insolation, Direct to Diffuse Ratio, Diurnal Anisotropic Heating, Flow Accumulation, Hydrology, Hillshade, Relative Radiation Available, Solar Radiation, Stream Density, Thermal

	Bands (satellite sensors), Topographic Wetness Index, Total Insolation
Parent material	Clay Mineral Index, Fe content, Geological Units, Iron Oxide Index, Magnetic Susceptibility, SiO2 content
Organisms	Normalized Difference Vegetation Index, Soil Adjusted Vegetation Index
Relief	Altitude Above the Channel Network, Catchment Area, Catchment Slope, Channel Network Base Level, Compound Topographic Index, Convergence Index, Curvature, Cross-Sectional Curvature, Diagonal Distance from Streams, Downslope Curvature, Elevation, Flow Direction, Flow Length, Flow Line Curvature, Generalized Surface, Landforms, Local Curvature, Local Downslope Curvature, Local Upslope Curvature, Longitudinal Curvature, Maximum Curvature, Maximum Flow Current Index, Mid-slope Position, Minimum Curvature, Multiresolution Index of Ridge Top Flatness, Multiresolution Index of Valley Bottom Flatness, Negative Openness, Normalized Height, Planar Curvature, Planar Distance from Streams, Positive Openness, Profile Curvature, Relative Altimetry of Sub-basins, Relative Slope Position, Sediment Transport Capacity, Slope, Slope Height, Standardized Height, Stream Network Base Level, Stream Power Index, Terrain Roughness Index, Tangential Curvature, Terrain View Factor, Topographic Factor (LS Factor), Topographic Position Index, Upslope Curvature, Valley Depth, Vector Terrain Roughness
Time	Geomorphic Surface

Some of the predictor variables have indirect or multiple-factor relations with soil-forming factor (MA *et al.*, 2019); in this study we associated variables that were directly related to a particular factor.

For the climate soil forming factor, predictor variables that influence temperature and soil moisture were assigned. For the parent material the characteristics of the mineral solid soil phase as well as the lithology of the environment used whereas for the organism the biological characteristics of the environment, and for relief the terrain model derivatives were used. For the time factor, only the geomorphic surface variable was used and this was used in two studies (Arruda et al., 2013 and Arruda et al., 2016).

The principal components were used as predictor variables for elevation, hydrology and curvature (TEN CATEN *et al.*, 2011c), only component characteristics explained by the authors were assigned to a soil-forming factors.

2.3.3 Compatibility with the Brazilian map accuracy standard

The Cartographic Accuracy Standard for Digital Cartographic Products (Padrão de Exatidão Cartográfica dos Produtos Cartográficos Digitais – PEC-PCD) is the Brazilian standard for the evaluation of the map accuracy published in the version 2.1.3 of the “Especificação Técnica para Aquisição de Dados Geoespaciais Vetoriais – ET-ADGV” (BRASIL, 2011). According to this standard, the digital products are classified in four classes (“A” – more accurate, “B”, “C” and “D” – less accurate) that indicate acceptable both altimetric and planimetric errors at different cartographic scales; e.g., for the scale 1:10,000 the PEC-PCD planimetric values are: 2.8 m (“A”), 5.0 m (“B”), 8.0 m (“C”) and 10.0 m (“D”). Thus, we infer the compatibility of the studies to the PEC-PCD planimetric by the pixel size and scale used; e.g., if a study used pixel size of 15 m and scale 1:10,000, it was considered incompatible with the PEC-PCD.

The studies were grouped in compatible and incompatible with PEC-PCD. In this study. the compatibility with planimetric PEC-PCD only indicates the studies that exhibited compatible scale and pixel size and not the positional precision of variables in relation to field coordinates.

2.3.4 Classifiers groups

The studies were grouped according to the type of learning algorithms used. The list of all algorithms used in the studies and the types of classifiers to which they were associated are presented in the results and discussion item.

2.4 Statistical methods for comparing the groups

The method of Zuur et al. (2010) was used for exploratory data analysis which includes graphical observations and statistical tests in an R environment. Overall accuracy data normality was tested by the Shapiro-Wilk and Kolmogorov-Smirnov statistical test, whereas Brown-Forsythe test (modified Levene test) was applied to analyze the homogeneity of variance. When both normality and homogeneity assumptions were met, parametric tests were applied. The Student's t-test was used to compare the means between two groups.

When both normality and homogeneity assumptions were not met, non-parametric tests were applied to the non-transformed data; i.e., the original overall accuracy data. The Wilcoxon-Mann-Whitney test was used to determine whether the distributions between two groups were equally located. The Kruskal-Wallis test was applied to verify if there were differences between three or more groups. When an inter-group difference was observed, the Dunn post-hoc test was used in each pair of groups.

When the overall accuracy data met one of the assumptions (i.e., normality or homogeneity), they were transformed using the Box-Cox method (BOX; COX, 1964). The transformed overall accuracy data were tested by the Shapiro-Wilk and Kolmogorov-Smirnov for test data normality; and the Brown-Forsythe for test of homogeneity of variance. If both normality and homogeneity assumptions were met, parametric tests were applied to the transformed data; otherwise non-parametric tests were applied to the non-transformed data.

3. Results and Discussion

3.1 Bibliographic Analysis

We included 42 articles that met the requirements (i.e., inclusion and exclusion criteria) for participation in this study (**Table 2**). These articles contained 334 digital soil class mapping studies. The first digital soil class mapping article in Brazil were carried out in Rio Grande do Sul state (GIASSON, Elvio *et al.*, 2006); the authors evaluated logistic regressions to reproduce soil maps from a reference area. More than half of all articles were conducted in Rio Grande do Sul state (24), followed by São Paulo (7), Minas Gerais (6) and Rio de Janeiro state (5). In most of the country, no digital soil class mapping articles, that met the requirements for participation in this study, has conducted. An average of 3 articles by year were published during the study period (2006 to 2019).

Table 2. Chronological list of selected articles that focus on digital soil class mapping in Brazil (2006-2016).

Year	Reference
2006	(GIASSON, Elvio *et al.*, 2006)
2008	(FIGUEIREDO *et al.*, 2008)
2009	(CRIVELENTI *et al.*, 2009)
2010	(CHAGAS, C. da S. *et al.*, 2010); (COELHO; GIASSON, 2010)
2011	(CARVALHO JÚNIOR *et al.*, 2011); (CHAGAS, C. S.; CARVALHO JÚNIOR; BHERING, 2011); (GIASSON, Elvio *et al.*, 2011); (TEN CATEN *et al.*, 2011c); (ten Caten et al., 2011b); (TEN CATEN *et al.*, 2011d); (TEN CATEN *et al.*, 2011a)
2012	(SARMENTO *et al.*, 2012); (TEN CATEN; DALMOLIN; RUIZ, 2012)
2013	(ARRUDA; DEMATTÊ; CHAGAS, 2013); (CHAGAS, C. da S.; OLIVEIRA; FERNANDES, 2013); (GIASSON, E *et al.*, 2013); (SILVA, C. C. *et al.*, 2013); (TEN CATEN *et al.*, 2013)
2014	(CALDERANO FILHO *et al.*, 2014); (HÖFIG; GIASSON; VENDRAME, 2014); (TESKE; GIASSON; BAGATINI, 2014)
2015	(BAGATINI; GIASSON; TESKE, 2015); (GIASSON, Elvio *et al.*, 2015); (TESKE; GIASSON; BAGATINI, 2015a); (TESKE; GIASSON; BAGATINI, 2015b); (VASQUES *et al.*, 2015)
2016	(ARRUDA *et al.*, 2016); (BAGATINI; GIASSON; TESKE, 2016); (DEMATTÊ *et al.*, 2016); (DIAS *et al.*, 2016); (SILVA, S. H. G. *et al.*, 2016); (PELEGRINO *et al.*, 2016)
2017	(CHAGAS, C. da S. *et al.*, 2017); (WOLSKI *et al.*, 2017)
2018	(COSTA, E. M.; SAMUEL-ROSA; ANJOS, 2018); (MEIER *et al.*, 2018)

2019	(CAMPOS *et al.*, 2019a); (CAMPOS *et al.*, 2019b); (MOURA-BUENO *et al.*, 2019); (SILVA, B. P. C. *et al.*, 2019); (SILVERO *et al.*, 2019)

3.2 Descriptive statistics of the data extracted from the studies

The study areas varied from 1.75 km^2 (175 ha) (PELEGRINO *et al.*, 2016) to approximately 120,000 km^2, which represents 48 % of the São Paulo State (SILVERO *et al.*, 2019). Approximately 50 % of the studies were conducted in areas up to 120 km^2. There is no information about the cartographic scale size in 70 studies. In the 264 studies containing this information, 169 were at a scale 1:50,000. The most detailed cartographic scale was 1:10,000 used in 49 studies (Giasson et al. (2011); Sarmento et al. (2012); Arruda et al. (2016); Pelegrino et al. (2016); Wolski et al. (2017)). The number of MU varied between 3 (FIGUEIREDO *et al.*, 2008), using a simplified legend, and 34 (Vasques et al. (2015); Silvero et al. (2019)) with an average of 9.5 MU per study. About 75 % of the studies had a ratio of up to approximately 0.4 MU km^{-2} but one study carried out by (Pelegrino et al. (2016) stands out for its high ratio of 2.8 MU km^{-2}, with a study area of 1.75 km^2 and five soil classes.

The median number of samples was 2,463. The lowest number of samples (74) was found in studies that used field observations to spatial soil class prediction models based on Decision Trees and Logistic Regressions algorithms (SILVA, B. P. C. *et al.*, 2019); still in this article the authors carried out others studies with additional points that improved the prediction performance of each model. The largest number of samples (794,273) was found in a study carried out by Crivelenti et al. (2009) that used pixels from a raster data to spatial soil class prediction models based on Decision Trees models.

The highest number of samples per area was 10,024 per km^2, in a study by Pelegrino et al. (2016), where the authors used 17,542 samples in an area of 1.75 km^2. Nevertheless Dias et al. (2016) used 1,710 samples in an area of 1,100 km^2. In both studies, pixels from a raster data was used as samples. The size of the study area and the number of samples are not correlated (Pearson (*r*

= 0.01)) which may reveal the lack of standardization in digital soil mapping (ten Caten et al. (2012a); Minasny e McBratney (2016); Arrouays et al. (2017).

A mean of nine predictor variables were used per study; i.e., the variables selected and used in the predictive models per study. The maximum number of variables used in the same study was 43 (SILVA, B. P. C. *et al.*, 2019); on the other hand, in one of the studies (PELEGRINO *et al.*, 2016) only two variables were used (aspect and wetness index) obtaining overall accuracy of 50 %.

Of the 334 studies, 263 presented cartographic scale and spatial resolution (pixel size) information used (**Table 3**); 38 studies were found incompatible with the planimetric PEC-PCD, since their pixel size is higher than that indicate by cartographic scale.

Table 3. Number of studies and the relation between pixel size and the cartographic scale.

Pixel size	**Cartographic Scale**									
	S1	S2	S3	S4	S5	S6	S7	S8	NA	Total
5 m	30	-	16	-	-	-	-	-	8	54
10 m	8	8	-	-	-	-	-	-	-	16
12.5 m	-	-	-	-	-	16	-	-	6	22
20 m	9	-	-	-	9	4	-	-	6	28
30 m	2	2	3	-	-	125	-	2	18	152
50 m	-	-	-	-	-	7	-	-	-	7
90 m	-	-	3	-	-	17	2	-	32	54
NA	-	-	-	1	-	-	-	-	-	1
Total	49	10	22	1	9	169	2	2	70	334

Not Available (NA). Studies incompatibles with the PEC-PCD are within the highlighted area. S1 = 1:10,000; S2 = 1:12,500; S3 = 1:20,000; S4 = 1:25,000; S5 = 1:30,000; S6 = 1:50,000; S7 = 1:80,000; S8 = 1:100,000.

3.3 Learning algorithms and types of classifiers

The following learning algorithms were used in the selected articles: Bagged AdaBag, BF Tree, C5 Decision Tree, CART, ExtraTree, J48, Logistic Model Trees, Maximum Likelihood, Multinomial Logistic Regression, Multilayer Perceptron, PART, Random Forest, Ranger Random Forest, Rep Tree, Support Vector Machine with Linear Kernel, Support Vector Machine with Polynomial Kernel, Weighted Subspace Random Forest -WSRF and xgBoost.

All learning algorithms were assigned to a type of classifier like Artificial Neural Network (ANN), Bayes classifiers, Decision Tree (DT), Logistic Regression (LR) and Support Vector Machine (SVM). Approximately 95 % of the studies used DT, ANN and LR classifiers (**Table 4**).

Table 4. Classifiers used in the studies.

	Proportion (%)	**Accumulated Proportion (%)**
Decision Trees	54	54
Artificial Neural Network	25	79
Logistic Regression	16	95
Bayes	3	98
Support Vector Machine	2	100

3.4 Digital soil map validation

3.4.1 Overall accuracy and Kappa coefficient

Of the 334 studies, the OA was used in 320 (96 %) while the Kappa in 190 (57 %). Although we do know that Kappa is often seen as problematic if not flawed because attempt to compare accuracy to a baseline of randomness (PONTIUS; MILLONES, 2011), we analyzed it considering its use is frequent in studies of digital mapping of soil classes.

The OA and Kappa medians were 62 % and 48 % of agreement respectively. The estimated population confidence interval (CI) for the OA median was 59 % to 63 % (CI (95 %) = 59 % - 63 %). The Kappa value remained the same as that found by ten Caten et al. (2012a) when carrying out an evaluation of 11 Brazilian studies of digital mapping of soil classes.

The agreement variation in the Kappa is higher than the OA; the OA agreement probability density is higher than the Kappa showed by the shape of the plot (**Figure 4**). The OA outliers are those below 23 %, identified in articles by Chagas et al. (2010), Vasques et al. (2015), Pelegrino et al. (2016) and Silva et al. (2019).

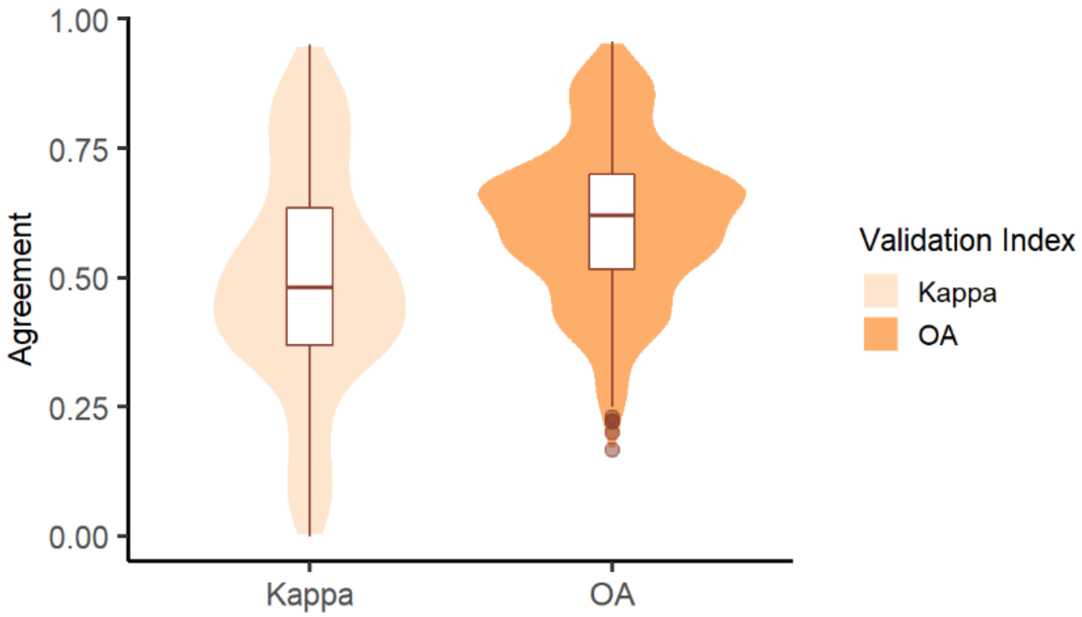

Figure 4. Results of OA and Kappa validation indices.

3.4.2 Reproducibility and exactness assessment for Overall Accuracy

When conventional maps were used for validation (reproducibility assessment group), the studies showed average values of 63 % for OA; when field samples were used for validation (exactness assessment group), the studies showed average values of 56 %.

Through the Box - Cox transformation of the OA data, it was possible to use parametric test for the reproduction and accuracy groups. Student's t-test results (t (150.53) = 3.73, $p < 0.05$)) indicates that the two validation groups are different. The 95 % confidence interval of the difference between estimated population means was 4 % and 12 %.

These findings indicate that the digital soil maps generated tend to have higher agreement in reproducibility than in exactness assessment; i.e., they were more accurate for reproduce legacy maps than to represent the actual soil distribution. That is understandable because when the legacy maps are used for training prediction models, the soil mapping units already correspond to well-identified landscape units, which makes adding more precise and up-to-date predictor variables useful for producing a better map (MA *et al.*, 2019).

Furthermore, the conventional maps used for validation are composed of polygons of MU and there is a relationship "one-to-many"; i.e., one MU to many classified pixels. The validation by field data is usually performed by points and there is a relationship "one-to-one"; i.e., one point to one classified pixel. In this way, validation by conventional maps increases the chances of random classification hits occurring within a MU. In addition, in the validation by field data any positional inaccuracy both of the points and the of predictor variables can compromise the validation. Therefore, the evaluation of the reproducibility is not enough for evaluate the exactness of predicted maps, for that, field data is necessary.

3.5 Factors affecting the overall accuracy

3.5.1 Environmental factors

There is a small correlation between OA values and the size of the study area (Pearson (r = 0.18)); Spearman (ρ = 0.23). There is no correlation between OA values and number of MUs (Pearson (r = –0.08); Spearman (ρ = 0.03)). These results suggest a random or practically non-existent association between OA and the size of the study area and the number of mapping units. These results are different from those founding by (BRUNGARD *et al.*, 2015) that machine learning models are most accurate when there a few soil classes.

To verify the effect of the two variables (the size of the study area and number of MUs) representing the density per km^2 (MU km^{-2}), the k-means technique was applied to perform iterative data segmentation. After the tests, the data were partitioned into 15 groups (clusters) (**Table 5**).

Table 5. Cluster statistics.

Cluster	**MU km^{-2}**		
	Minimum	Centroid	Maximum
1	0.0000	0.0081	0.0100
2	0.0111	0.0122	0.0158
3	0.0206	0.0223	0.0239
4	0.0246	0.0262	0.0263
5	0.0283	0.0283	0.0284
6	0.0300	0.0301	0.0316
7	0.0632	0.0632	0.0632
8	0.0750	0.0764	0.0812
9	0.0846	0.0890	0.1200
10	0.1507	0.1647	0.1700
11	0.2000	0.2302	0.2615
12	0.4100	0.4245	0.5000
13	0.6275	0.7218	1.0571
14	1.9028	1.9028	1.9028
15	2.8571	2.8571	2.8571

As densities (MU km^{-2}) increase beyond the cluster 9 (lower limit = 0.08; centroid = 0.09; upper limit = 0.12), OA is lower in most studies (**Figure 5**). Except for cluster 13, those above 9 showed median OA below the lower limit of the estimated population confidence interval for the OA median (CI (95 %) = 59 % - 63 %). On the other hand, for clusters below 9, median OA of the studies are predominantly higher or within the estimated population confidence interval, except cluster 4, whose median was below the CI (95 %). Approximately 50 % of the studies are in clusters 1 to 9 and 50 % between clusters 9 and 15. The density values (MU km^{-2}) depend on the scale work and relate the environmental complexity. The results indicate that the greater environmental complexity (with 0.08 MU km^{-2} or more) has a negative effect on the accuracy of the predicted maps.

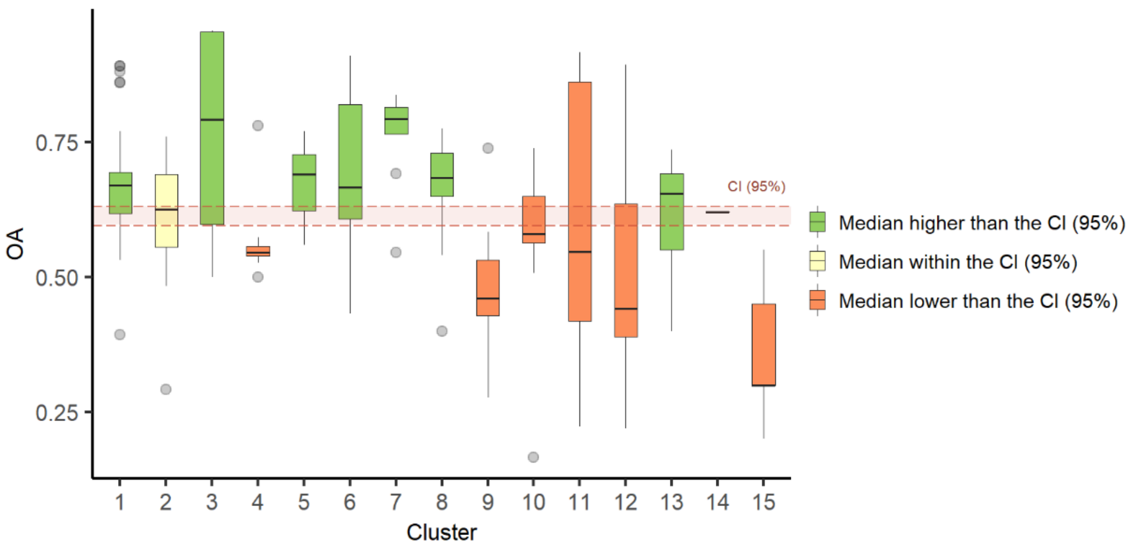

Figure 5. Relation between the OA and cluster of MU density.

3.6 Modelling factors

3.6.1 Scale and pixel size

Of the studies with information available of cartographic scale and pixel size, validated by the OA index, 85 % used spatial resolution compatible and 15 % incompatible with the planimetric PEC-PCD.

The result of the Wilcoxon Mann Whitney test ($p < 0.05$) suggests that the OA of the population median for the group of studies compatible with the PEC-PCD is higher than that of the incompatible group. The shape of the plot suggests that the OA probability density for the group of studies compatible with the PEC-PCD is higher than the CI (95 %) (**Figure 6**). The inter-group difference in the estimated population median is between 2 % and 12 % with a 95 % CI.

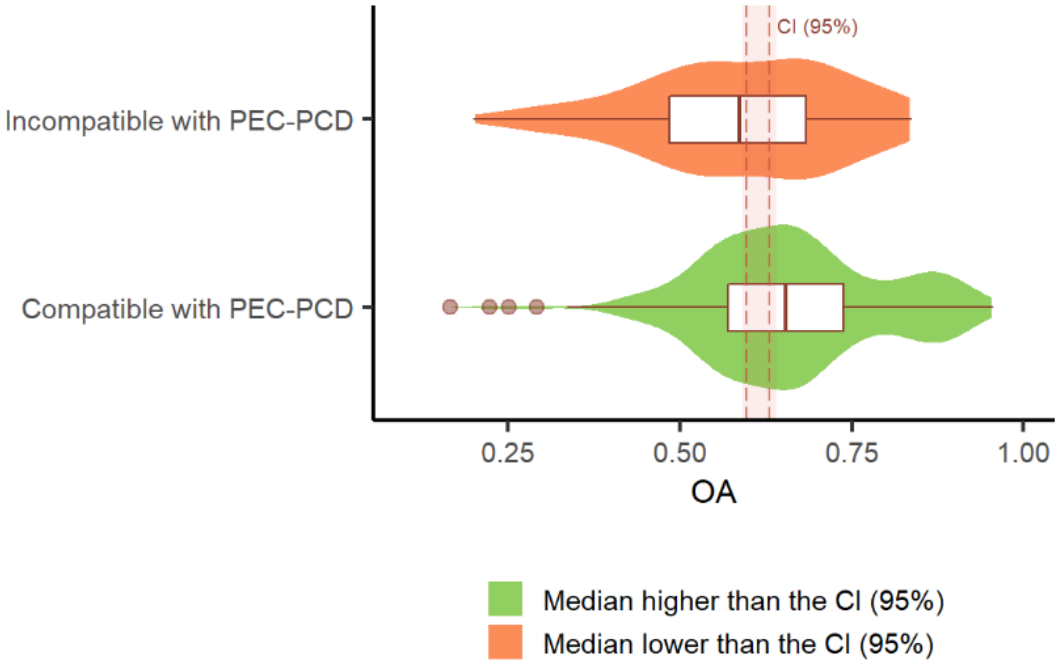

Figure 6. Relation between the OA results for groups of studies that used spatial resolution compatible and incompatible with PEC-PCD.

3.6.2 Sample size and density

There is not a correlation between OA values and the number of samples (r = 0.04) and also between OA values and sample density per km^2 (r = –0.08). In a study that aimed to extrapolate the soil map (GRINAND *et al.*, 2008) the authors found that the increase in the number of samples improved the prediction accuracy, whereas the increase in the sample density did not improve the accuracy.

It is important to underscore that these results are restricted only to sample number and density and not to sampling method. As such, prediction map agreement is associated with environmental and modelling factors, which may include sampling methods.

3.6.3 Number of predictor variables and of soil-forming factor

There is not a correlation between OA results and the number of predictor variables used (Spearman (ρ = –0.02); Pearson (r = 0.20)). The association between predictor variables and the respective soil-forming factor (**Table 1**) results in the number of formation factors used per study: a) only one formation factor (7 % of the studies); b) two factors (49 %); c) three factors (33 %); d) four factors (5 %); and e) no information available (6 %). None of the studies was associated with five soil-forming factors. There is a trend to higher OA results as the number of formation factors and variables increased (**Figure 7**).

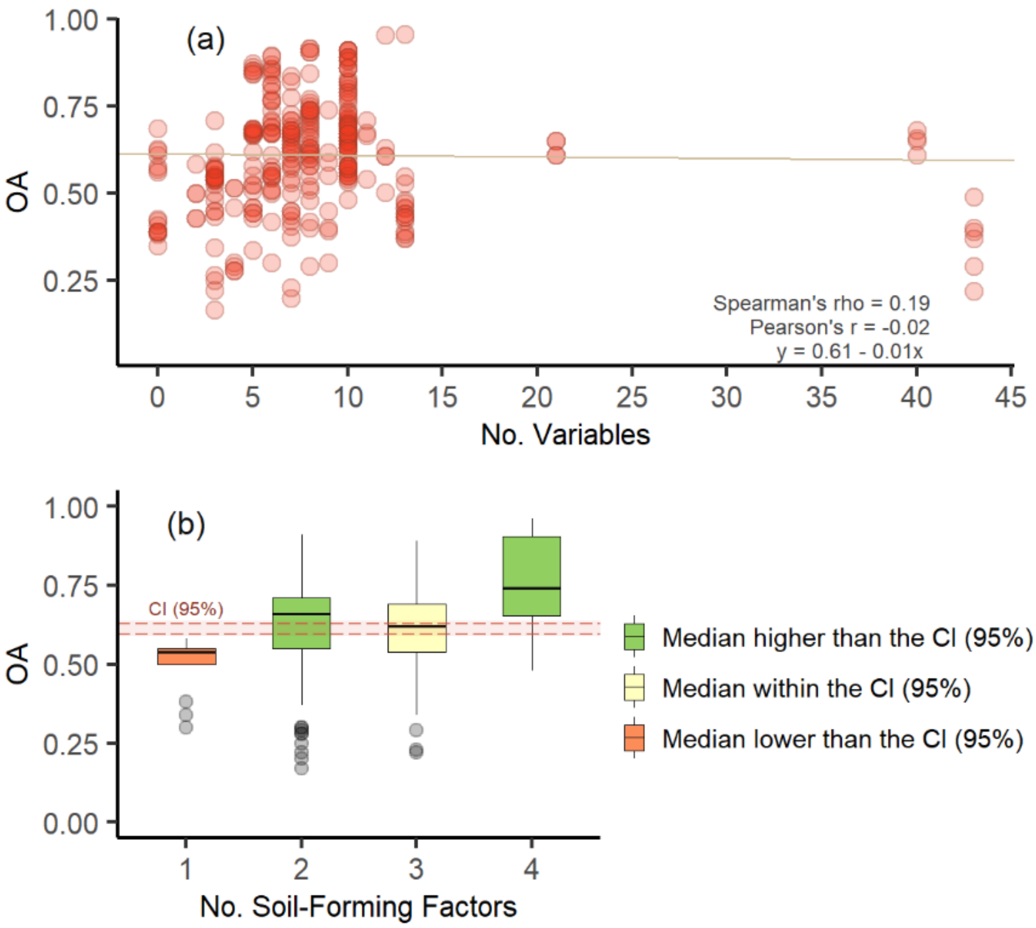

Figure 7. Relation between the OA and number of predictor variables used (A); relation between the OA and number of soil-forming factors used (B).

The result of the Kruskal-Wallis test (H (3) = 28.91, $p < 0.05$) indicated a difference between the groups of studies in which different numbers of soil-forming factors were used. The results of Dunn's post-hoc test indicated differences between groups with one and two, one and three, one and four, two and four, three and four formation factors ($p < 0.05$) and equality between those with two and three formation factors ($p > 0.05$).

The results indicate that the techniques applied in the set of studies analyzed here are sensitive to the conceptual structure given by the paradigm of soil-forming factors adapted to digital soil mapping using the scorpan model (MCBRATNEY; MENDONÇA SANTOS; MINASNY, 2003). The more completely the scorpan model is applied, the better the results obtained.

3.6.4 Classifiers

Due to the greater representativeness (95 % of studies), we compared the OA results of the following groups of classifiers: Decision Trees (DT; mean = 62 %; median = 63 %), Artificial Neural Network (ANN; mean = 67 %; median = 68 %) and Logistic Regressions (LR; mean = 45 %; median = 43 %). The results of the Kruskal-Wallis test (H (2) = 62.34, $p < 0.05$) indicated a difference between the groups analyzed. The results of Dunn's post-hoc test indicated that OA is different between LR and DT, LR and ANN ($p < 0.05$) but between DT and ANN ($p > 0.05$) there is not difference.

Figure 8 shows that the group of studies that used ANN had a median OA higher than the estimated populational CI (95 %), however there is no evidence that ANN models are better than DT models. On the other hand, there is evidence that ANN and DT models are better than LR for predicting soil classes. Evaluating prediction models of soil properties (KHALEDIAN; MILLER, 2020) concluded that there is no a correct learning algorithm. However, some algorithms are more appropriate than others considering the purpose of the mapping. According to the authors, if sample size is large, ANN would likely produce the best results, when interpretability of the resulting model is important LR and DT are more appropriate than others. Brungard et al. (2015) compared 11 learning algorithms and concluded that ANN and SVM were consistently more accurate than LR and DT algorithms.

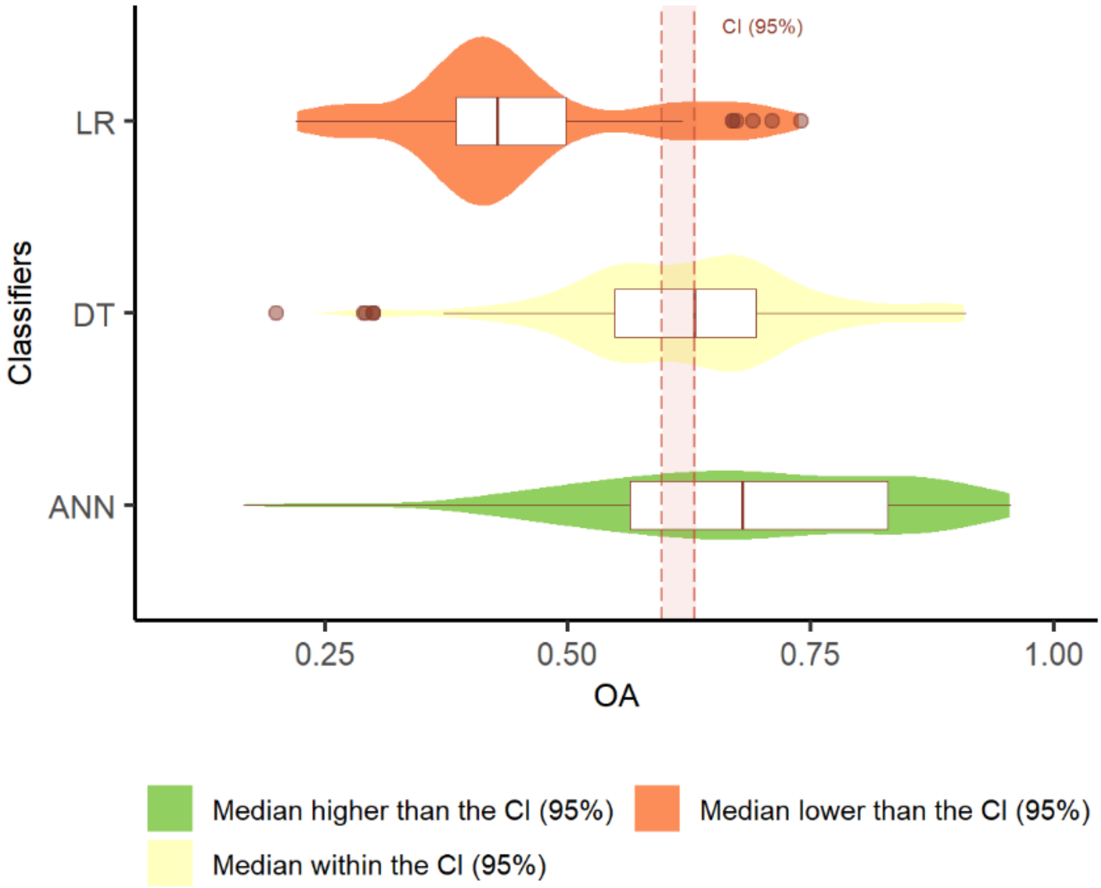

Figure 8. Relation between the main classifiers used and OA values. LR = Logistic Regression; DT = Decision Tree; ANN = Artificial Neural Network.

3.7 Relation between environmental and modelling factors

The results obtained in the present study demonstrate that the following factors determine higher OA values: a) density up to 0.08 MU km^{-2}; b) spatial resolution and scale compatible with planimetric PEC-PCD; c) use of four or more soil-forming factors associated with predictor variables; and d) use of ANN and DT classifiers.

Among these factors, only the density of MU km^{-2} is an environmental factor, which cannot be controlled by the user. The other factors are related to modelling of digital soil class mapping, which are controlled by the user.

Graphical analyses (**Figure 9**) and statistical tests were used to determine the possibility of bias in establishing the main factors, that is, confirm if any modelling-related factor obtained better OA because they were primarily distributed into lower MU km^{-2} values.

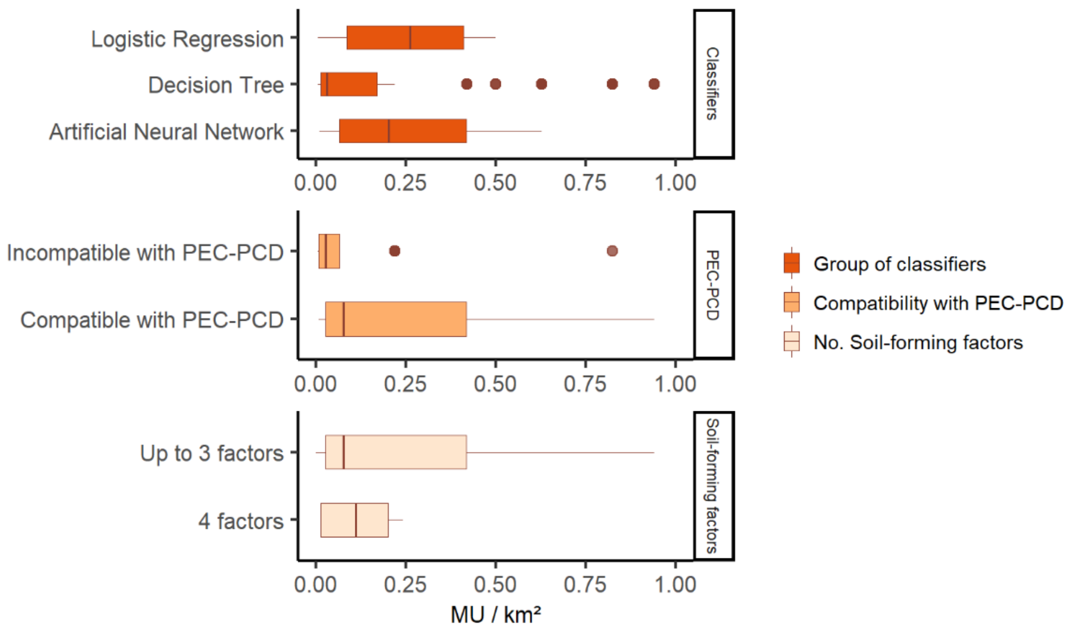

Figure 9. Relation between the density of mapping units per area and compatibility with PEC-PCD, number of soil-forming factors and group of classifiers.

For the group of compatibility with PEC-PCD, there is a greater concentration of studies compatibles in higher densities of MU km^{-2}. This indicates that there is no bias and reinforces the importance of using pixels sizes appropriate to the working scale to obtain better OA values.

For the groups of soil-forming factors and of classifiers, visual analysis was not conclusive. As such, a statistical test was necessary to determine whether the groups exhibited similar values variation of MU km^{-2}.

According to the Wilcoxon-Mann-Whitney ($p > 0.05$), data from the group with four soil-forming factors does not differ from the group with up to three factors in relation to MU km^{-2}, so does not exhibit low density concentration bias. Thus, it can be considered an important factor for improving soil prediction map agreement.

According to the Kruskal-Wallis test (H (2) = 18.68, $p < 0.05$), the classifiers groups differ in relation to MU km^{-2}. The results of Dunn's post-hoc test indicated that there is not difference between LR and ANN ($p > 0.05$); between DT and LR, DT and ANN there are difference. These finding indicates that the ANN performed better even though they were tested at higher densities of MU per area. In addition, the good results of the DT models may have had bias because most studies have occurred in areas with lower environmental complexity.

4. Conclusions

Based on digital soil class mapping studies in Brazil conducted between 2006 and 2019 and considering that this is a small database, the results of this study may not be definitive, the following can be concluded:

i. The mean overall accuracy of the group of studies that used pixel size and cartographic scale compatible with planimetric PEC-PCD is greater than that of the group that did not use spatial resolution compatible with PEC-PCD.

ii. There is no evidence that an increase in the number of samples and predictor variables results in more accurate soil map prediction. On the other hand, there is evidence that the use of more heterogeneous predictor variables in terms of soil-forming factors could result in improved accuracy.

iii. The density of MU per area affects the agreement of prediction maps. From a density of 0.08 MU km^{-2} onwards, it was more difficult for studies to obtain better overall accuracy values than their estimated population median counterparts.

iv. There is evidence that ANN classifiers are more efficient than the LR to predict soil classes. There is no evidence that ANN are more efficient than the DT. However good DT accuracy may have occurred because most tests were performed in areas of lower MU km^{-2}; i.e., less environmental complexity.

CAPÍTULO IV - INTEGRATION OF THE GEOGRAPHIC OBJECT-BASED IMAGE ANALYSIS AND ARTIFICIAL NEURAL NETWORKS FOR DIGITAL SOIL MAPPING

1. Introduction

When good quality legacy soil maps are available, these can be used to calibrate soil prediction models (COLLARD *et al.*, 2014). However, legacy maps can have limitations. They are often based on obsolete cartographic data, generally coarse in scale and lack details (ZHANG; LIU; SONG, 2017) and the polygons of the mapping units are usually delimited by soil surveyors because they are considered homogeneous, while they are actually products of mental models based on higher categorical levels of various soil classification systems (ARROUAYS *et al.*, 2020). Additionally, to these limitations, the use of techniques to calibrate soil prediction models based on legacy soil maps is restricted to areas where conventional soil survey and mapping were used.

Thus, it is necessary to seek alternatives to produce digital soil maps for areas without legacy maps, i.e., from field observations and soil profiles. One of the major problems is the difficulty to delineate polygons soil classes based on soil sampling points; in addition, the small area covered by field samples which is used to calibrate predictive models (CAMPOS *et al.*, 2019a), once each point match only one pixel and not a region or a group of pixels in the pixel-based approach.

To extrapolate a pixel information to a region, some studies used buffers around the sampling points. Buffer zones with different radii were built, resulting in areas of 1,960 m² (PELEGRINO *et al.*, 2016), 7,800 m² (CAMPOS *et al.*, 2019a) and 2,800 m² (MACHADO *et al.*, 2019). The authors obtained better results with the extrapolation of areas of the soil profile with the use of buffers. However, one of the problems associated with the use of a buffer of a specific size is that the environmental characteristics of the surrounding area are ignored in a buffer delimitation. Thus, for example, if a sampling point is located at the boundary between a flat area and a sloping area, the buffer polygon will assign the values of the pixels of these two topographic positions and relate them to the classification of the source soil profile, which may result in classification errors.

One way to extrapolate a soil sample point to a region, considering the environmental conditions near each point, is the Geographic Object-Based Image Analysis (GEOBIA) approach. In GEOBIA approach each point match an image-object (not an individual pixel) that represent “meaningful” geographic entities at multiple scales (CHEN *et al.*, 2018). Image-objects are derived from the segmentation of variables that describe natural characteristics related to soil formation factors. The segmentation groups adjacent and similar pixels into a region which is generated based on one or more criteria of homogeneity in one or more dimensions (Blaschke, 2010; Myint et al., 2011). GEOBIA is an approach widely used in the classification of remote sensing images. It is also used for the classification of terrain features based on Digital Elevation Models (DEMs) (Drăguţ and Blaschke, 2006; Gerçek et al., 2011; Drăguţ and Eisank, 2012; Martha et al., 2018). However, this approach has not yet been extensively used in digital mapping of soil classes (DORNIK; DRÂGUŢ; URDEA, 2017).

The most commonly method used for GEOBIA classification is the nearest neighbor; however it may be less effective in the presence of high-dimensional data and there are other machine-learning algorithm potentially more powerful that can be used for GEOBIA classification (MAXWELL *et al.*, 2015). In a research to predict soil classes by machine learning methods, Brungard et al. (2015) compared 11 algorithms and found that complex models as Artificial Neural Networks (ANN’s) were consistently more accurate than others simple or moderately complex models as Random Forest or others decision trees algorithms. A study conducted by Arruda et al., (2016) aimed to evaluate the efficiency of ANN’s for the extrapolation of soil classes to adjacent unmapped areas, from a detailed legacy soil map. The authors found that the same pattern of distribution of soil classes of the legacy map was estimated for the unmapped areas and concluded that the technique has the potential to be used in digital soil mapping.

However, when the capacity of the resulting model explain soil distribution on is more important that the capacity of the model to predict, logistic or decision tree models are more appropriate than ANN, that are considered to be a “black box” (KHALEDIAN; MILLER, 2020). In the present study we prioritize using models that have presented better accuracies such ANN’s. The use of ANN’s in the prediction of soil classes has shown promising results (Behrens et al.,2005; Chagas et al., 2013) and, nevertheless, ANN’s have not yet been tested together with the GEOBIA approach in digital mapping of soil classes.

This paper presents a strategy to delineate polygons of soil classes based on georeferenced points of soil profiles. The main objective is to evaluate the integration of the GEOBIA approach at different segmentation levels with ANN’s models and sampling points to produce a digital soil map of the Vale dos Vinhedos region, in Rio Grande do Sul, Brazil.

2. Material and methods

2.1 Study area

The study area comprises the Vale dos Vinhedos region, located in the municipalities of Bento Gonçalves, Garibaldi and Monte Belo do Sul, in Rio Grande do Sul, extending over an area of approximately 8,100 hectares and at altitudes ranging from 200 to 700 meters (**Figure 10**). The region is located in the geomorphological unit of Serra Geral. The relief is complex, with great variability in slope and altitude (FLORES *et al.*, 2012). The soil genesis consists of rocks from the Fácies Caxias and Fácies Gramado unit, consisting of basalts, andesites and dacites. Soils of the following orders are found there: Cambisols (43%), Alisols (29%), Regosols (14%), Chernozems (11%), Nitisols (2%) and Planosols (1%) (FLORES *et al.*, 2012).

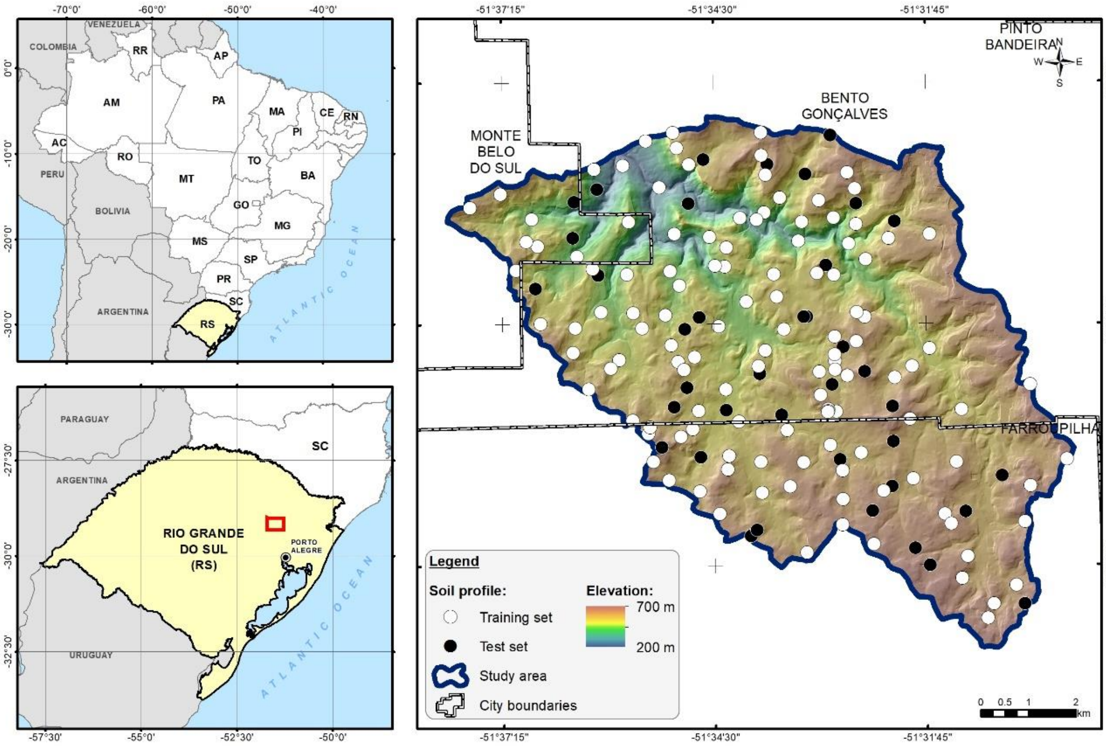

Figure 10. Location of the study area and distribution of soil profiles.

2.2 Digital soil mapping strategy

In ArcGIS 10.4 for Desktop environment, was created a Geodatase of the detailed soil survey of Vale dos Vinhedos in the scale 1:10,000 (FLORES *et al.*, 2012). The database includes 163 georeferenced points (legacy data) of the soil profiles and the polygons of the mapping units (legacy soil map). We extracted predictive variables from a Digital Terrain Model (DTM) and from satellite images from RapidEye’s sensors. A principal component analysis (PCA) of the variables was performed and a subset of principal components retained was by Scree Plot analysis and the Kaiser’s rule. The subset of principal components retained was an input to image segmentation processes.

In eCognition Developer 9 environment, were executed image segmentations processes performed by Multiresolution Segmentation algorithm at different levels by trial-and-error process to input parameters; and by Cheesboard Segmentation algorithm to divides the image according to the original pixels. The objects resulting from the segmentation processes were exported to polygon shapefiles.

In ArcGIS 10.4 for Desktop environment, the polygons resulting from the segmentation processes overlapping the georeferenced points of soil profiles were selected to create a sample dataset. Thus, a one-to-one ratio (1:1) was used, i.e., one soil profile for one object. Both, the polygon shapefiles resulting from the segmentation processes and the subset of samples were loaded in R environment (R Core Team, 2019) using the R-ArcGIS Bridge resources that enables to access ArcGIS data and bring it into R language for statistical analysis as well to access R data and bring it into ArcGIS for visualization and analysis.

In R environment, the sample dataset was subdivided into training (75%) and test (25%) subsets. ANN's models were trained and tested at each segmentation level. Based on the classification correctness of each soil profile, the accuracy of the models was calculated with overall accuracy (OA). ANOVA was performed to assess differences between OA's results at each segmentation level. The best model at each level was used to predict the soil classes at entire dataset of polygons resulting from segmentation processes at different levels. Each entire polygon dataset with soil prediction was loaded in ArcGIS environment as a polygon shapefile using the R-ArcGIS Bridge resources for visualization and analysis of each predicted map in the entire study area.

In ArcGIS 10.4 for Desktop environment, the predict soil maps were compared with a legacy map to evaluate the degree of agreement at each segmentation level. The soil map was predicted with the best model obtained for each scale parameter. The degree of reproducibility of the legacy soil map and the prediction model was calculated by comparing the two maps using OA, user's accuracy (UA) and producer's accuracy (PA) for each soil class. A flowchart of the methodology is presented in **Figure 11**.

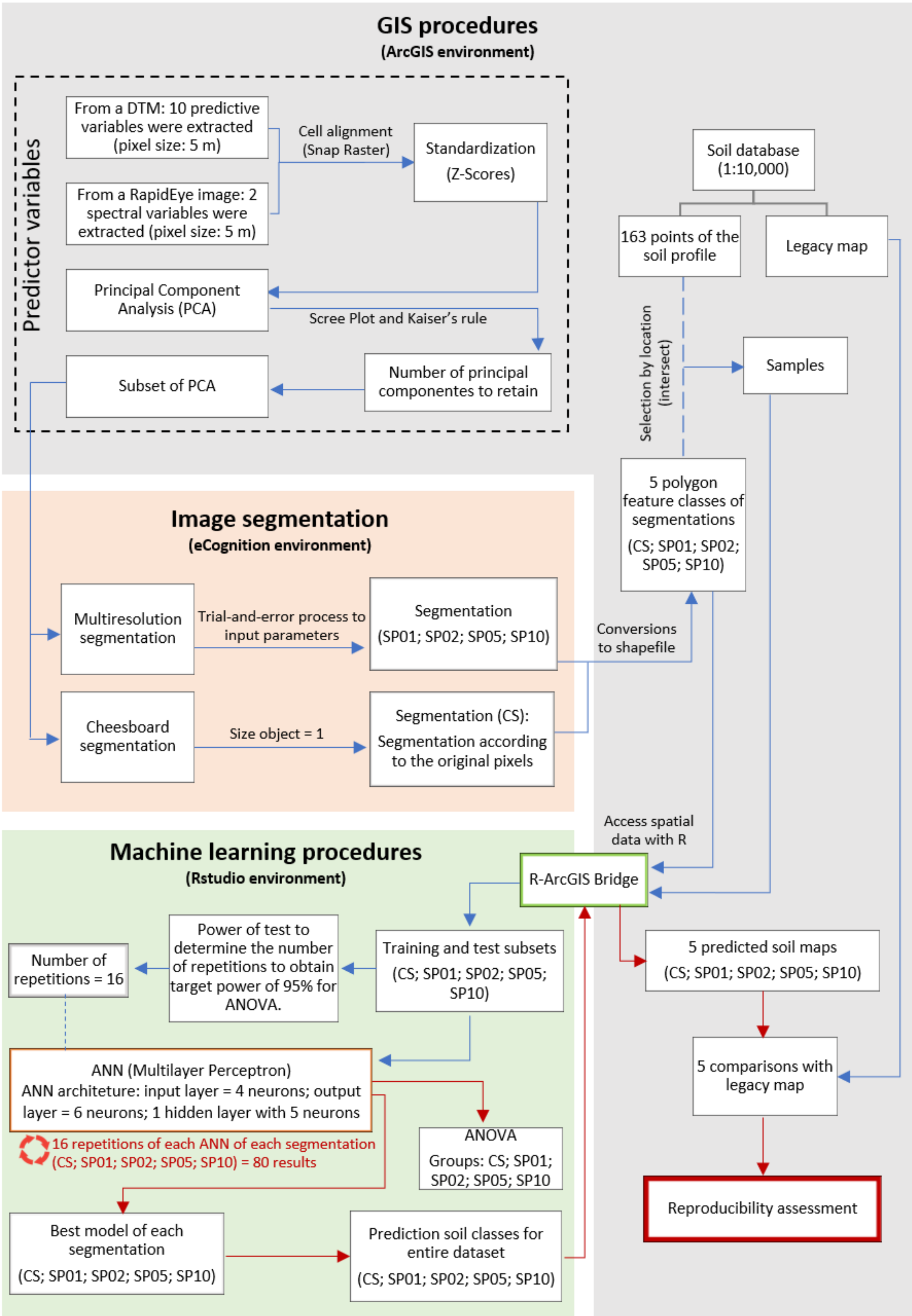

Figure 11. Illustration of the methodology process. DTM (Digital Terrain Model); PCA (Principal Component Analysis); SP (Scale Parameter); CS (Cheesboard Segmentation); ANN (Artificial Neural Network).

2.3 Predictor variables

From a DTM obtained with aerial digital photogrammetry (FLORES *et al.*, 2012), with 5m pixel size, 10 predictive variables most frequently used in digital soil mapping (DSM) studies were extracted in ArcGIS environment: Flow Accumulation (FA), Altimetry (ALT), Height Above the Nearest Drainage (HAND) (RENNÓ *et al.*, 2008), Curvature (CURV), Horizontal Curvature (CURVH), Vertical Curvature (CURVV), Slope (SLP), Euclidean Distance to stream (ED), Topographic Wetness Index (TWI) and Solar Radiation (SR). The drainage network was obtained from the corrected DTM (Fill) to remove minor imperfections in the model and ensure a coherent hydrological structure. The HAND variable was produced using the methodology of Rennó et al. (2008), implemented in a Model Builder (ArcGIS 10.4 for Desktop). All of these variables were produced with a pixel size of 5m.

From the RapidEye remote sensing image, with 5m spatial resolution and with 5 spectral bands (1 - Blue: 440-510 nm; 2 - Green: 520-590 nm, Red: 630 - 685 nm, Red Edge: 690 - 730 nm and Near Infrared: 760 - 850 nm), the following spectral variables were extracted: Normalized Difference Vegetation Index (NDVI) and Normalized Difference Water Index (NDWI), as follows:

$$NDVI = \frac{NIR-RED}{NIR+RED} \qquad \text{Equation 2}$$

$$NDWI = \frac{GREEN-NIR}{GREEN+NIR} \qquad \text{Equation 3}$$

Where NIR corresponds to the Near Infrared band, RED corresponds to the Red Band and GREEN corresponds to the Green Band. The spectral variables were recorded with the DTM (Snap Raster), to obtain the best overlap between the pixels of all the 12 variables. As these variables have different units and magnitudes, it was necessary to standardize them (z-score).

Python programming was used to standardize the variables. Then, a new set of variables not correlated with each other was produced using PCA in ArcGIS environment. Scree Plot analysis was used to determine the number of components to be retained, and Kaiser’s eigenvalue rule, according to which the last component to be retained has eigenvalue equal to or greater than 1 was adopted (KAISER, 1958). With the eigenvalues and eigenvectors data generated by PCA, the contribution values (loadings) of each variable in each component were obtained.

2.4 Image segmentation procedures

In the segmentation of principal components (PC), the eCognition Developer 9 computer application and the Multiresolution Segmentation (MRS) algorithm (BENZ *et al.*, 2004) were used. The Multiresolution Segmentation algorithm is a bottom up region-merging technique that groups the pixels of the input variables for homogeneous and contiguous objects at different levels (KARAKIŞ; MARANGOZ; BÜYÜKSALIH, 2006). The objects have additional attributes compared to unit pixels, i.e., they contain statistical information of the set of pixels of which they are composed (average values per variable, median, variance, minimum and maximum values) and additional spatial information (area size, perimeter, compactness index, among others) (BLASCHKE, 2010). Therefore, it was necessary to define homogeneity criteria to produce these contiguous objects.

Trial-and-error process was used to set the composition homogeneity criterion by shape and compactness parameters and to set the input segmentation scale parameter (SP). The shape parameter is based on the deviation of a smooth shape and compactness can be defined by the product of the width and the length over pixels numbers (EL-NAGGAR, 2018). After the trial-and-error process, both shape and compactness parameters chosen were 0.1. The SP determines the maximum heterogeneity allowed for the objects, i.e., SP is a user defined threshold of homogeneity (KARAKIŞ; MARANGOZ; BÜYÜKSALIH, 2006). Thus, larger values result in larger objects, with greater heterogeneity, while smaller values result in smaller and more homogeneous objects. After the trial-and-error process, the SP values chosen were 1 (SP01), 2 (SP02), 5 (SP05) and 10 (SP10) resulting in four levels of segmentations. In addition to the four levels of segmentations, a fifth procedure was performed by the Cheesboard segmentation (CS) algorithm, in an eCognition Developer environment, for pixel-based classification. This algorithm divides the image into square objects of fixed size with predefined sizes. A size 1 object was used, which divided the image according to the original pixels.

2.5 Machine learning procedures

A sample dataset created in ArcGIS 10.4 for Desktop environment from 163 soil profiles (legacy data) and the polygons resulting from the segmentation processes was loaded in R environment using the R-ArcGIS Bridge resources. In R environment, the dataset was subdivided into training (75%) and test (25%) subsets with the use of the proportionate stratification sampling technique to build the ANN's models.

The ANN's were built with the use of the h2o.deeplearning function of the H2O package (CANDEL; PARMAR, 2014) in RStudio v. 1.2.5001 environment. The Multilayer Perceptron was used and the network topology was 4-5-6 being four inputs (number of PCs used), five hidden neurons in one hidden layer (mean of the neurons of the input and output layers) and six output neurons (number of soil orders). Random distribution of the weights occurred through the feedforward process and the update of the weights with the use of the backpropagation method with 100,000 iterations in the training dataset. The activation function used was the hyperbolic tangent, which is a rescheduled and displaced logistic function, with symmetry of approximately 0 and variation from -1 to 1, as follows:

$$f(a) = \frac{e^{a}-e^{-a}}{e^{a}+e^{-a}} \quad \text{Equation4}$$

Where α is the combination of neuron values, weights and bias; xi and ωi represent, respectively, the neuron input values and its weights, and *b*, the bias value as follows:

$$a = \sum_{i} \omega_i x_i + b \qquad \text{Equation5}$$

Five ANN's structures were implemented: four for the segmented data (SP01, SP02, SP05 and SP10) and one for the pixel-oriented approach (CS). To compare the accuracy of classifications using the different levels of segmentation, 16 repetitions of each ANN were performed, resulting in 80 models. The number of repetitions was obtained by calculating the statistical power of the test with a target value of 95%. The results of the models were submitted to analysis of variance (ANOVA) at a significance level of 5%. The Gedeon (1997) method was used for determining the significance of the predictor variables. It is a functional measure based on a dynamic process, on the analysis of the network that works better than the use of sensitivity analysis (GEDEON, 1997) and was implemented in H2O package.

3. Results and discussion

3.1 Principal Components Analysis

Based on the Kaiser’s eigenvalue rule (KAISER, 1958) (**Figure 12**), according to which components with eigenvalues greater than 1 should be retained, the first four (4) components were retained, which explain more than 90% of the variance of the set of original variables. The other PC's explained small proportions of the residual variance of the original variables and were not considered in this study. The contributions of the variables to the composition of each PC are shown in **Figure 13**. Based on these contributions, we attempted to make a physiographic characterization of the PCs and relate them to soil formation factors.

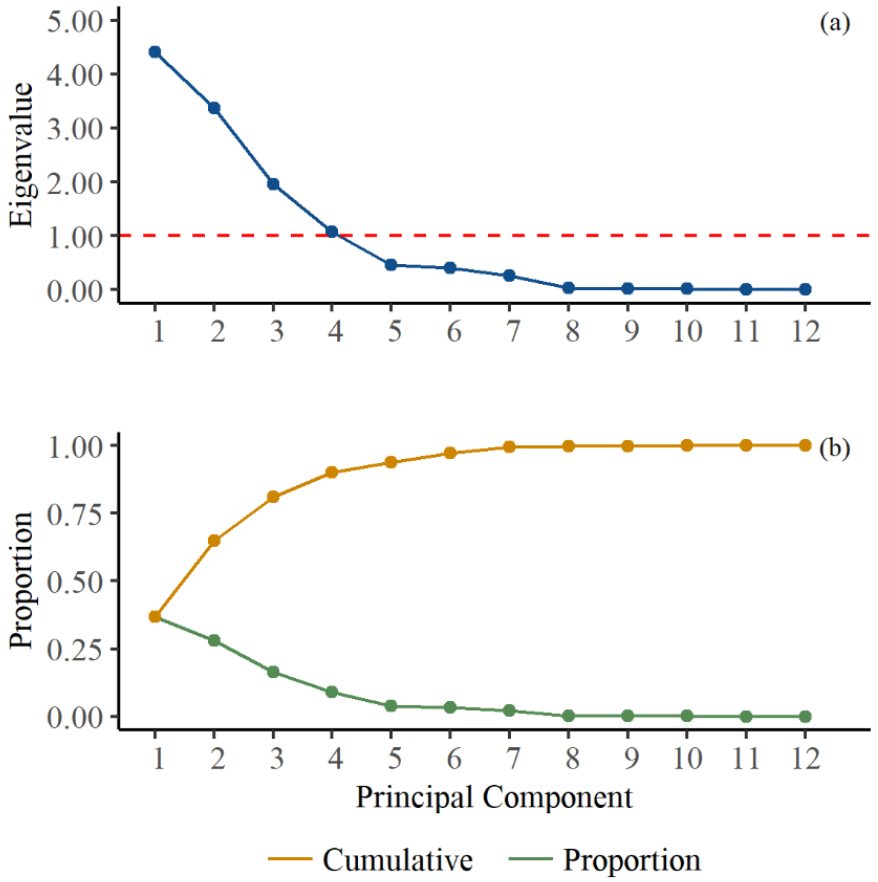

Figure 12. Scree plot (a) and the proportion of Eigenvalues obtained in PCA (b).

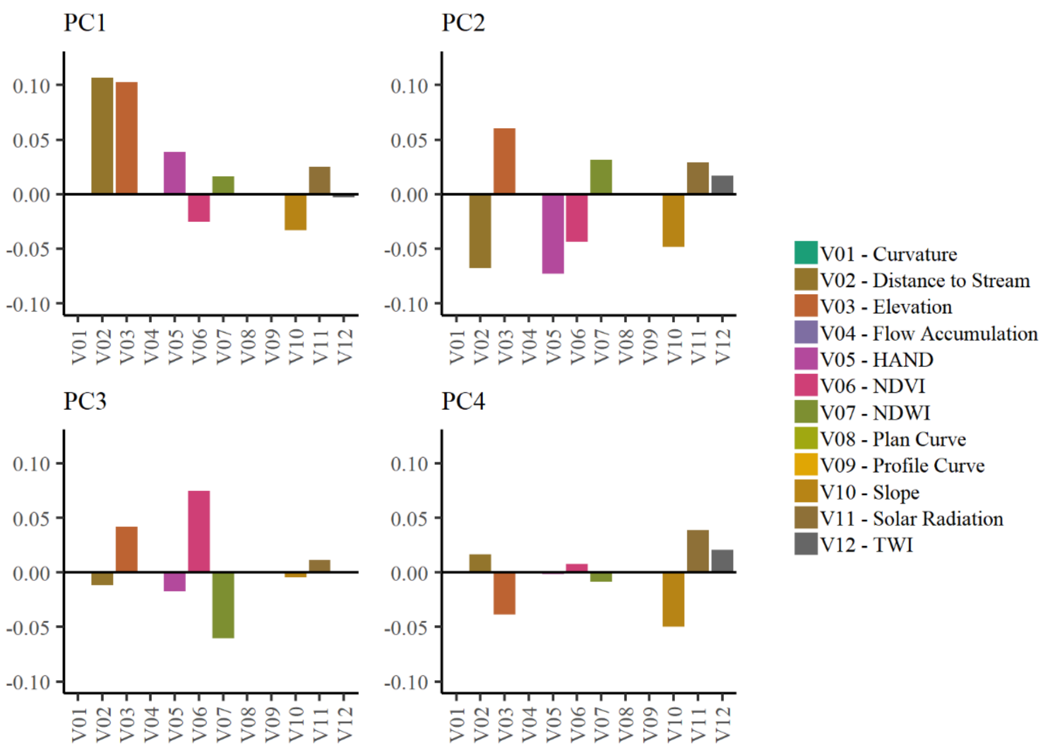

Figure 13. Contribution (loadings) of the original variables in the first 4 PCs.

PC1 is related to the higher and less hilly physiographic regions, which can be inferred both by the negative correlation with slope and by the positive correlation with the distance to the stream network, which indicates the interfluvial dimension, and greater distances represent a relief with geomorphological characteristics less susceptible to dissection, according to the proposition of Ross (1994). PC 2 is also related to higher areas and flatter areas, but near rivers, both horizontally and vertically (HAND), with higher moisture content (characterized by the variables NDWI and TWI). PC 3 is mostly related to the areas with higher vegetation coverage and reduced water depth (NDWI). PC 4 is related to lower and flatter areas, with higher moisture (TWI) and greater contribution of the solar radiation variable. Thus, PC 1 seems to be more related to the soil formation factor "Relief" and PC 2 to “Relief” and “Climate” factors, as it represents characteristics that influence soil temperature and soil moisture. PC 3, in turn, is more related to the “Organisms” factor, as it represents characteristics of more developed vegetation cover, and PC 4 is more related to the “Climate” factor, as it has characteristics that impact soil temperature and soil moisture. A color composition with the first three principal components is presented in **Figure 14**.

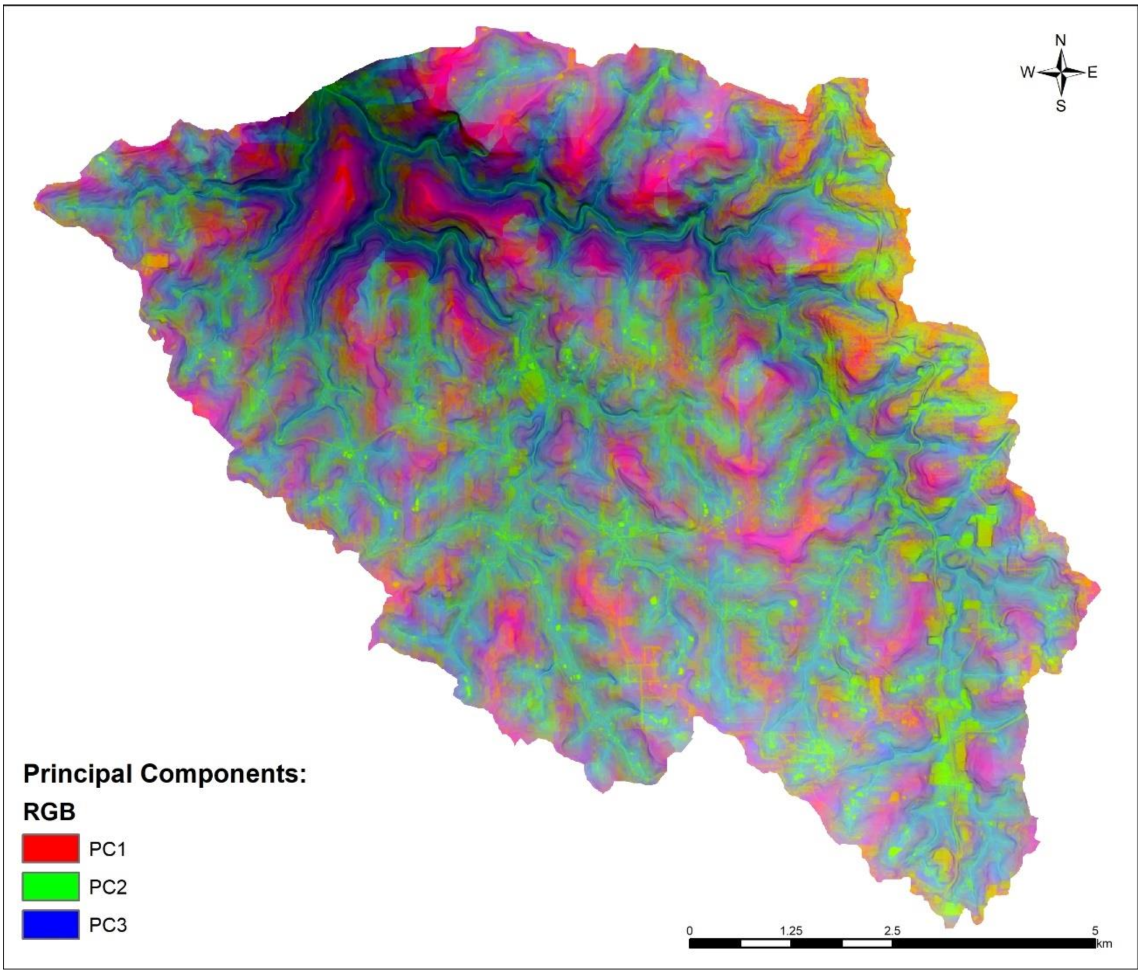

Figure 14. An RGB color composition with the first three principal components: PC1 (red); PC2 (green); PC3 (blue).

Using PCA in digital soil mapping studies, ten Caten et al. (2011) attempted to reduce and avoid redundancy in a set of geomorphometric variables, obtaining more than 65% of the variance of the original data in the first 3 components. According to the authors, the first three PC's were more related to altimetry (PC1), hydrology (PC2) and curvature (PC3) components. Thus, PC 1 and PC 3 were more associated with the relief factor and PC 2 with the soil climate factor. Machado et al. (2019) did not relate the PC's to the soil formation factor. However, they obtained better accuracy in the prediction of soil types with PCA, with the use of all original variables as control.

3.2 Image segmentation and soil map prediction

With the segmentation of PCs, in a spatial resolution of 5m, there was a reduction of more than 98% in the number of original pixels for the new objects

(**Table 6**). In the pixel-based approach (CS), a dimensionless point of soil profile is extrapolated to a square area, which, in this study, corresponds to an area of 25m² (original pixel area). In the GEOBIA approach, the profile was extrapolated to areas on average 60 times larger in the first segmentation level (SP01) (**Figure 15**). A significant reduction in the number of original pixels has a positive impact on the computer costs associated with ANN's modeling, but may have a negative impact on the results depending on the level of generalization of the objects as noted in the overall accuracies (OA) of the predicted maps.

Table 6. Segmentation of PCs.

	Number of objects	**Reduction in the number of objects (%)**	**Object Area (m²) (Average)**
CS (pixels)	3,235,310	0.00%	25
SP01	54,040	98.33%	1,496
SP02	22,176	99.31%	3,660
SP05	7,304	99.77%	11,080
SP10	2,451	99.92%	32,953

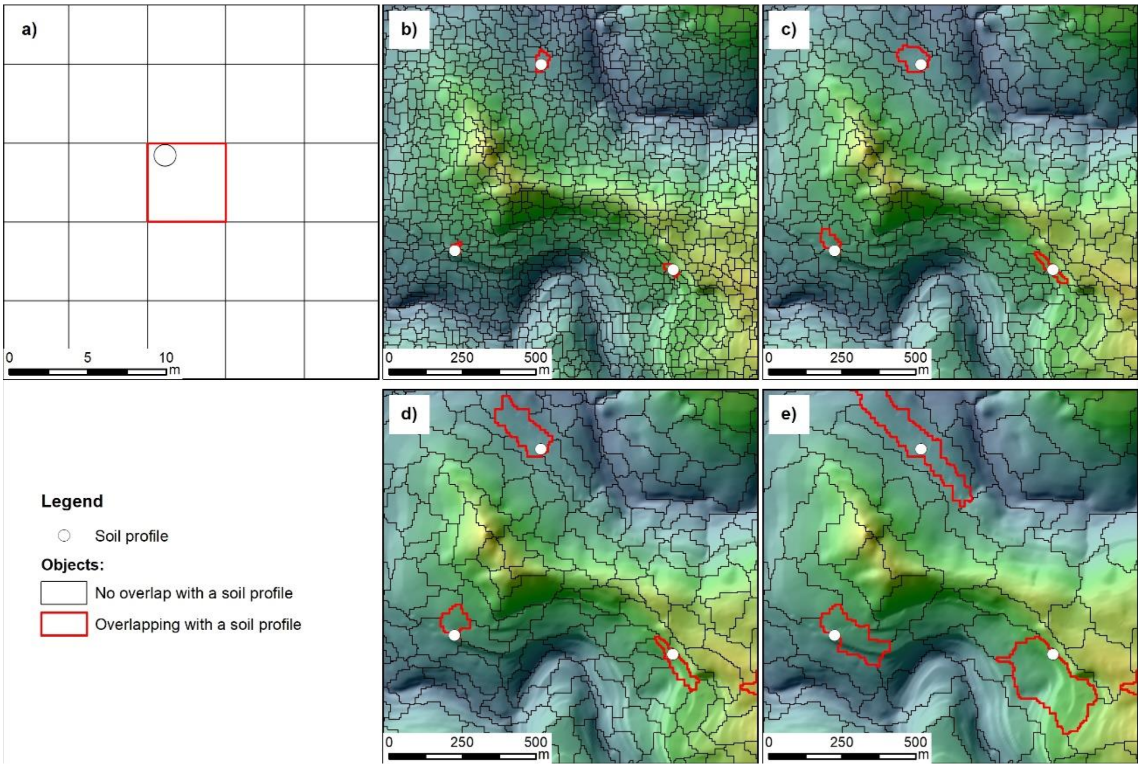

Figure 15. Segmentation variables. a) CS (pixels), b) SP01, c) SP02, d) SP05 and d) SP10.

The results of OA of the validation of the modeling with the soil profiles (subset of test samples) are shown in **Figure 16**. There was no difference between the groups with ANOVA (p = 0.14). Even so, of the best models in each group, the CS model obtained the worst result (OA = 48.72%). The best model was obtained in SP05, with 61.54% of OA.

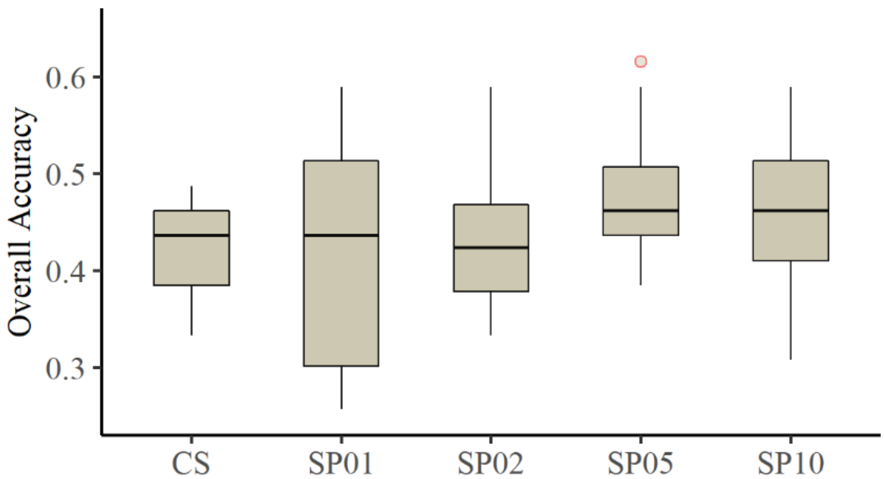

Figure 16. Results of OA in test samples (accuracy).

The relative importance of the variables of each best model at each segmentation level are showed in **Figure 17**. At all levels, the magnitude importance of PC1 and PC2 exceeded more than 50%. It seems to indicate that the soils of the Vale dos Vinhedos, that presents great variability in slope and altitude, were more controlled by formation factors related to the relief and climate. Similar results to those found by Chagas et al. (2017) in a hillslope area and by Meier et al. (2018) in a mountainous area that from 73 variables, 10 were selected, four related to relief and two to climate.

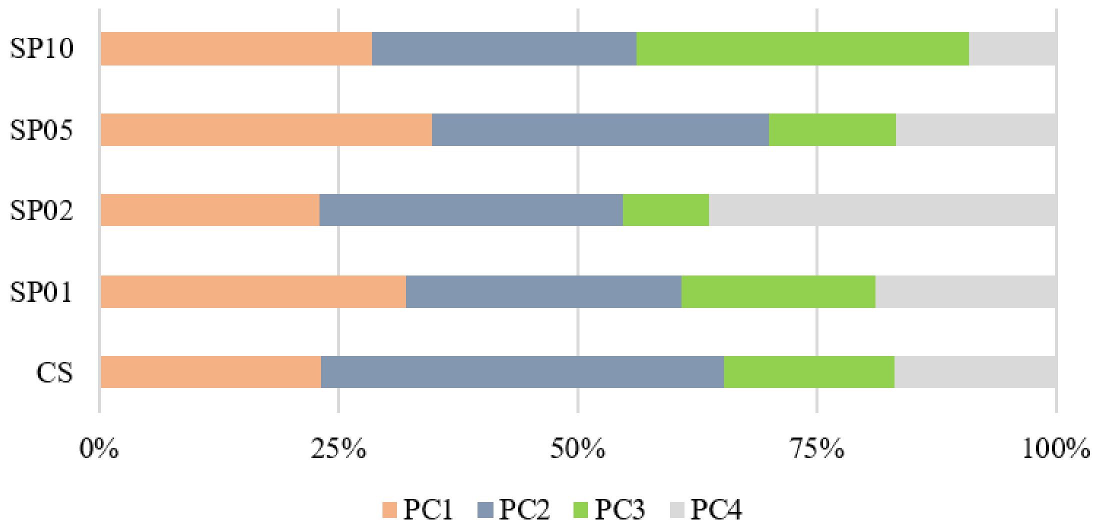

Figure 17. Relative importance of the variable in each segmentation level.

With the best models for each segmentation level, extrapolations of soil classes at each level were performed for the entire study area (**Figure 18**). From the elaboration of error matrices, OA values that indicate the degree of reproducibility (agreement) of the predicted maps with the legacy map were obtained, as shown in the summary of accuracies in **Table 7** and error matrices in **Table 8**. The best result in the spatial prediction of soil types found in this study (OA = 45.68%) was obtained with the GEOBIA approach. Segmentation was performed with a scale parameter of 1 (SP1), and the average unit area of these objects was 1.496 m². This result was approximately 9 percentage points more accurate than the pixel-based procedure (CS). It was found that larger objects, on average, generated less accurate soil maps.

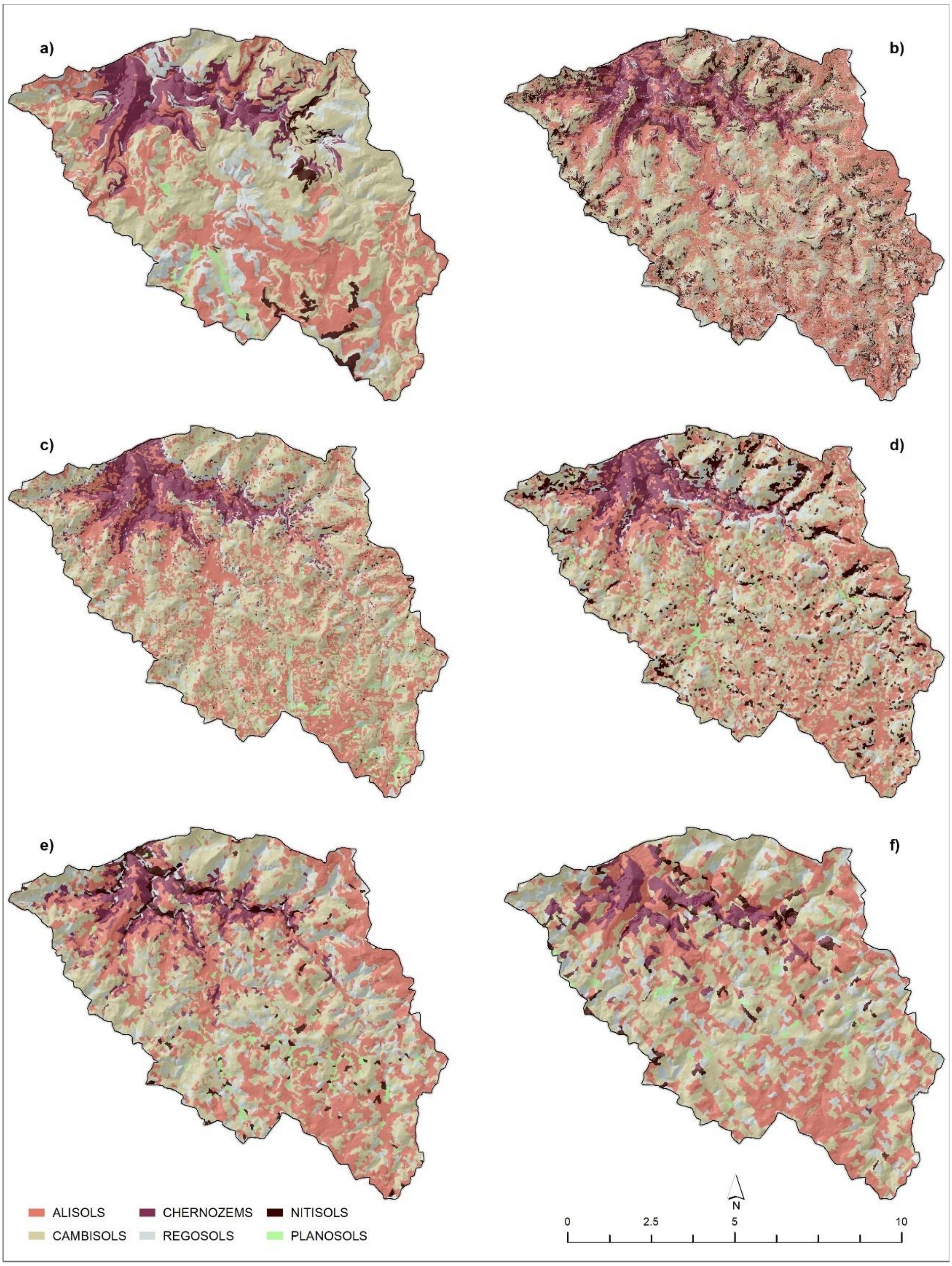

Figure 18. Conventional map and map predicted by ANN's. a) Conventional map; b) CS (pixel-based classification); c) SP01, d) SP02, and) SP05 and f) SP10.

Table 7. Accuracy of map predicted by ANN.

Predict map	Overall accuracy	User's accuracy	Producer's accuracy	Soil
CS	36.54%	37.61%	47.80%	ALISOLS
		50.01%	33.84%	CAMBISOLS
		58.05%	48.22%	CHERNOZEMS
		18.42%	18.55%	REGOSOLS
		1.15%	5.64%	NITISOLS
		7.03%	7.44%	PLANOSOLS
SP 01	45.68%	40.74%	46.48%	ALISOLS
		52.83%	55.02%	CAMBISOLS
		64.12%	54.46%	CHERNOZEMS
		24.30%	15.98%	REGOSOLS
		7.33%	1.37%	NITISOLS
		4.14%	14.26%	PLANOSOLS
SP 02	40.47%	40.42%	47.18%	ALISOLS
		50.59%	43.00%	CAMBISOLS
		63.70%	43.69%	CHERNOZEMS
		20.83%	20.70%	REGOSOLS
		3.79%	11.17%	NITISOLS
		12.87%	31.88%	PLANOSOLS
SP 05	42.04%	38.34%	48.50%	ALISOLS
		52.00%	45.01%	CAMBISOLS
		60.04%	39.96%	CHERNOZEMS
		26.74%	26.75%	REGOSOLS
		4.27%	5.61%	NITISOLS
		14.39%	35.98%	PLANOSOLS
SP 10	41.80%	39.33%	45.71%	ALISOLS
		50.89%	50.64%	CAMBISOLS

	60.13%	38.18%	CHERNOZEMS
	17.65%	16.34%	REGOSOLS
	3.43%	4.24%	NITISOLS
	9.17%	11.61%	PLANOSOLS

Table 8. Error matrices.

PREDICTED MAP (CS)

		S1	S2	S3	S4	S5	S6	Total (ha)	PU (%)
ORIGINAL MAP	**S1**	1108	658	73	264	173	41	2317	47.80
	S2	1293	1189	112	546	348	26	3514	33.84
	S3	177	115	443	70	113	1	918	48.22
	S4	248	361	120	206	171	3	1109	18.55
	S5	80	33	14	29	9	2	168	5.64
	S6	40	21	0	3	5	6	75	7.44
	Total (ha)	2945	2378	763	1117	819	79	8100	
	UA (%)	37.61	50.01	58.05	18.42	1.15	7.03		
	OA (%)	36.54%							

PREDICTED MAP (SP 01)

		S1	S2	S3	S4	S5	S6	Total (ha)	PU (%)
ORIGINAL MAP	**S1**	1077	912	69	122	5	132	2317	46.48
	S2	1077	1933	113	288	10	93	3514	55.02
	S3	154	141	500	120	3	0	918	54.46
	S4	229	587	89	177	10	17	1109	15.98
	S5	80	56	9	15	2	5	167	1.37
	S6	26	30	0	7	0	11	75	14.26
	Total (ha)	2643	3659	780	729	31	257	8100	
	UA (%)	40.74	52.83	64.12	24.30	7.33	4.14		
	OA (%)	45.68%							

PREDICTED MAP (SP 02)

		S1	S2	S3	S4	S5	S6	Total (ha)	PU (%)
ORIGINAL MAP	**S1**	1094	843	73	165	84	60	2320	47.18
	S2	1151	1512	69	499	197	88	3516	43.00
	S3	164	103	401	186	61	3	917	43.69
	S4	204	456	80	229	128	9	1107	20.70
	S5	67	57	5	19	19	1	167	11.17
	S6	28	17	1	2	3	24	74	31.88
	Total (ha)	2708	2988	629	1100	491	184	8100	
	UA (%)	40.42	50.59	63.70	20.83	3.79	12.87		
	OA (%)	40.47%							

PREDICTED MAP (SP 05)

		S1	S2	S3	S4	S5	S6	Total (ha)	PU (%)
ORIGINAL MAP	**S1**	1124	798	57	163	61	114	2318	48.50
	S2	1250	1581	117	503	31	32	3514	45.01
	S3	223	119	367	125	83	1	918	39.96
	S4	218	487	63	297	34	10	1109	26.75
	S5	98	33	7	17	9	2	167	5.61
	S6	19	22	0	4	2	27	75	35.98
	Total (ha)	2932	3041	611	1109	220	186	8100	
	UA (%)	38.34	52.00	60.04	26.74	4.27	14.39		
	OA (%)	42.04%							

PREDICTED MAP (SP 10)

		S1	S2	S3	S4	S5	S6	Total (ha)	PU (%)
ORIGINAL MAP	**S1**	1058	879	45	287	14	31	2314	45.71
	S2	1011	1780	132	494	56	42	3515	50.64
	S3	242	220	351	45	61	1	919	38.18
	S4	265	528	54	181	69	11	1110	16.34
	S5	103	44	1	12	7	0	168	4.24
	S6	11	47	0	8	0	9	75	11.61

Total (ha)	2689	3498	584	1027	207	95	8100
UA (%)	39.33	50.89	60.13	17.65	3.43	9.17	
OA (%)	41.80%						

S1 (Alisols); S2 (Cambisols); S3 (Chernozems); S4 (Regosols); S5 (Nitisols); S6 (Planosols); UA (User's accuracy); PU (Producer's accuracy); OA (Overall accuracy).

Similar results were found in the first study of digital soil mapping using an GEOBIA approach (DORNIK; DRÂGUŢ; URDEA, 2017). The authors used nine levels of scale parameters and obtained better results than the pixel-based approach. The authors found OA results of 58%, 10 percentage points higher than the pixel-based strategy, corroborating the present study. Studies of soil mapping using the GEOBIA approach are still scarce. However, Prudente et al. (2017) considered equivalent the land use classification by GEOBIA and the per-pixel approach using Landsat-8 satellite images with 30 m pixel size. Baker et al. (2013) compared the results of classification of aerial images by GEOBIA and per-pixel classification with 1, 2 ,5, 15 and 30 m pixel sizes in a study case of mapping forest cleanings. The authors found that GEOBIA classification was significantly more accurate than the per-pixel classification at the very finest resolution of 1 m; for coarser pixel size, GEOBIA and per-pixel methods are not statistically different. Similar results to this study that no statistically difference was found between the classifications with 5 m of pixel size.

The other literature studies on strategies to extrapolate a pixel information of georeferenced points of soil profiles to a region did not use the segmentation strategy, opting for the strategy of buffers around each site. Using buffers with radii of 25m, 50m, 75m and 100m around the samples, Pelegrino et al. (2016) found that the best results were obtained with the use of 25m buffers, which corresponds to an area of approximately 1,960 m². The OA were of 50% and 30% in two study areas with four and five mapping units, respectively. These results were similar to those found in the present study (OA = 45.68%), though this one had 6 mapping units. Also, in a similar way, as larger areas were used around the sampling points, the degree of accuracy decreased, as it was found in this study. The same pattern was found in two different study areas in the study conducted by Campos et al. (2019) that used buffer sizes between 50m and 250m around points of the georeferenced soil profiles. When the areas were larger, the accuracy decreased. In the study conducted by Machado et al. (2019) there was no comparison of different sizes of buffers, and only the buffer with a radius of 30m was used, which corresponds to an area of approximately 2,827m² around the sampling point. Of the strategies used, the buffer strategy obtained the best accuracy, reaching values of 70% of OA.

Analysis of studies that used buffers provided insight on the behavior of the extrapolation of pixel information of sampling soil points to a region. However, it is difficult to compare those studies with the present study, because the other authors used the values of all pixels in the buffer areas, that is, a one-to-many (1: N) relationship (one soil profile for many pixels). In the present study, a one-to-one ratio (1: 1) was used, i.e., one soil profile for one object. Moreover, the differences in the environmental conditions of the study areas, and differences in modeling, such as classifiers and number of samples and variables used, make comparisons difficult.

The use of ANN's for digital soil mapping has been promising (Behrens et al. 2005; Chagas et al. 2013), however it presents variation of results according to the specifics of each study. Bagheri Bodaghabadi et al. (2015) found errors training below 11% in a large number of ANN models for prediction of soil series, however the validation data (OA) of the best models ranged from 12 to 34%; the authors suggest that the high prediction errors were due to using only terrain-related attributes. Calderano Filho et al. (2014) used ANN's with spectral variables from remote sensing products, geomorphometric variables and geology in an area with a high degree of lithological diversity and the results provided consistent data similar to maps produced by conventional methods with OA ranged from 94 to 95%, though with more spatial detail. Similar results were found by Behrens et al. (2005) with mean OA above 92% for prediction soil maps. Chagas et al. (2017) made a comparison between the predicted map obtained from ANN models and a conventional soil map and found 63% of OA, it is better agreement that the comparisons with sampling locations (56%).

4. Conclusions

In this study we have outlined a novel digital soil mapping strategy using only georeferenced points of soil profiles to delineate detailed polygons of soil classes by the GEOBIA approach and ANN's models. The results showed low accuracies, perhaps due to using only sampling points and a small sample size. This study found no evidence to suggest statistically difference between GEOBIA and per-pixel classification for soil mapping. Nevertheless, the best models of ANN's were obtained with the use of GEOBIA approach. Based on the best models for each segmentation level, the agreement of reproducibility of the maps predicted with the GEOBIA approach was higher than that of the pixel-based approach.

GEOBIA studies for soil mapping are scarce. Despite the low accuracies found in this study, GEOBIA for the digital mapping of soil classes is apparently a promising approach for delineating detailed polygons of soil classes once the segmentation groups pixels of variables that describe natural characteristics related to soil formation factors into homogeneous polygons. In addition, the GEOBIA approach has shown better results in detailed mappings that require small pixel sizes.

CAPÍTULO V – CONCLUSÕES GERAIS

A estratégia de uma revisão sistemática dos artigos brasileiros de mapeamento digital de classes de solos (CAPÍTULO III) foi eficiente para apresentar informações quantitativas sem o viés subjetivo de escolha de literatura como em uma revisão tradicional. Os critérios de inclusão e exclusão foram explícitos. Assim foi possível obter uma visão geral das abordagens utilizadas desde o início do mapeamento de classes de solos no Brasil e indicar boas práticas a serem utilizadas. Dentre as boas práticas, é possível elencar a utilização de tamanho de pixel compatível com a escala de estudo, uma melhor distribuição das variáveis preditoras em relação aos fatores de formação do solo e a não necessariamente de maior quantidade de variáveis. Também proveu uma informação do cenário ambiental na qual é mais difícil de realizar a separação de classes de solos: aquelas situações com maiores densidades de unidades de mapeamento por unidade de área. Além disso, evidenciou que, dentre os classificadores utilizados, os que deram melhores resultados foram os métodos de decisão em árvore e métodos de redes neurais artificiais.

Além dessas informações, também ficou evidenciado quais técnicas ainda não foram testadas em âmbito nacional como, por exemplo, a classificação orientada a objetos, que foi empregada no estudo apresentado no CAPÍTULO IV. Apesar de não ter se obtido evidência estatística que os modelos obtidos pela classificação orientada a objetos, os valores mais altos foram obtidos com a segmentação das variáveis e os mapas reproduzidos com maior concordância com o mapa existente também foi aquele realizado com classificação orientada a objetos em relação aos métodos baseados em pixels.

Para futuros trabalhos, sugere-se que os procedimentos metodológicos apresentados no CAPÍTULO IV sejam repetidos em outras áreas de estudo para verificar a existência de melhorias significativas nos resultados de mapeamento de classes de solos. Além disso, também se sugere que sejam explorados recursos que são intrínsecos da segmentação de variáveis para o mapeamento de solos. A abordagem GEOBIA possibilita que sejam elaborados modelos preditivos a partir de descritores (atributos) relativos ao tamanho e forma dos objetos, como tamanho dos poligonos, dimensão de borda, proporção entre comprimento e largura, compacidade e também elaborar regras de classificação a partir das relações topológicas contextuais entre os objetos. Esses descritores podem ser utilizados em conjunto com os dados das variáveis preditoras geomorfométricas e espectrais. Até o presente momento não foi encontrado estudo de mapeamento de classes de solos que tenha utilizado as informações de contexto e forma dos objetos para discriminação das unidades de mapeamento.

REFERÊNCIAS BIBLIOGRÁFICAS

AGGARWAL, C. C. **Neural networks and deep learning**. Cham: Springer International Publishing, 2018. Disponível em: https://doi.org/10.1007/978-3-319-94463-0. Acesso em: 20 ago. 2019.

ARROUAYS, D. *et al.* Impressions of digital soil maps: the good, the not so good, and making them ever better. **Geoderma Regional**, Netherlands, v. 20, p. e00255, 2020. Disponível em: https://doi.org/10.1016/j.geodrs.2020.e00255. Acesso em: 6 mar. 2020.

ARROUAYS, D.; LAGACHERIE, P.; HARTEMINK, A. E. Digital soil mapping across the globe. **Geoderma Regional**, Netherlands, v. 9, p. 1–4, 2017. Disponível em: https://doi.org/10.1016/j.geodrs.2017.03.002. Acesso em: 6 set. 2017.

ARRUDA, G. P. *et al.* Digital soil mapping using reference area and artificial neural networks. **Scientia Agricola**, Piracicaba, v. 73, n. 3, p. 266–273, 2016. Disponível em: https://doi.org/10.1590/0103-9016-2015-0131. Acesso em: 18 set. 2017.

ARRUDA, G. P. ; DEMATTÊ, J. A. M.; CHAGAS, C. S. Mapeamento digital de solos por redes neurais artificiais com base na relação solo-paisagem. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 37, n. 2, p. 327–338, 2013. Disponível em: https://doi.org/10.1590/S0100-06832013000200004. Acesso em: 12 jul. 2017.

BAGATINI, T.; GIASSON, E.; TESKE, R. Expansão de mapas pedológicos para áreas fisiograficamente semelhantes por meio de mapeamento digital de solos. **Pesquisa Agropecuária Brasileira**, Brasília, v. 51, n. 9, p. 1317–1325, 2016. Disponível em: https://doi.org/10.1590/s0100-204x2016000900031. Acesso em: 14 set. 2017.

BAGATINI, T.; GIASSON, E.; TESKE, R. Seleção de densidade de amostragem com base em dados de áreas já mapeadas para treinamento de modelos de árvore de decisão em mapeamento digital de solos. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 39, n. 4, p. 960–967, 2015. Disponível em: https://doi.org/10.1590/01000683rbcs20140289. Acesso em: 14 jul. 2017.

BAGHERI BODAGHABADI, M. *et al.* Digital Soil Mapping Using Artificial Neural Networks and Terrain-Related Attributes. **Pedosphere**, China, v. 25, n. 4, p. 580–591, 2015. Disponível em: https://doi.org/10.1016/S1002-0160(15)30038-2. Acesso em: 12 dez. 2016.

BAKER, B. A. *et al.* Does spatial resolution matter? a multi-scale comparison of object-based and pixel-based methods for detecting change associated with gas well drilling operations. **International Journal of Remote Sensing**, United Kingdom, v. 34, n. 5, p. 1633–1651, 2013. Disponível em: https://doi.org/10.1080/01431161.2012.724540. Acesso em: 14 maio 2020.

BEHRENS, T.; SCHOLTEN, T. A comparison of data-mining techniques in predictive soil mapping. *In*: LAGACHERIE, P.; MCBRATNEY, A. B.; VOLTZ, M. (Ed.). **Digital soil mapping**: an introductory perspective. Netherlands: Elsevier, 2007. v. 31, Chapter 25, p. 353–364. Disponível em: https://doi.org/10.1016/S0166-2481(06)31025-2. Acesso em: 28 dez. 2017.

BEHRENS, Thorsten *et al.* Digital soil mapping using artificial neural networks. **Journal of Plant Nutrition and Soil Science**, Germany, v. 168, n. 1, p. 21–33, 2005. Disponível em: https://doi.org/10.1002/jpln.200421414. Acesso em: 7 set. 2017.

BENZ, U. C. *et al.* Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. **ISPRS Journal of Photogrammetry and Remote Sensing**, Netherlands, v. 58, n. 3–4, p. 239–258, 2004. Disponível em: https://doi.org/10.1016/j.isprsjprs.2003.10.002. Acesso em: 5 mar. 2019.

BLASCHKE, T. Object based image analysis for remote sensing. **ISPRS Journal of Photogrammetry and Remote Sensing**, Netherlands, v. 65, n. 1, p. 2–16, 2010. Disponível em: https://doi.org/10.1016/j.isprsjprs.2009.06.004. Acesso em: 7 set. 2017.

BOX, G.; COX, D. An analysis of transformations: applying the box-cox transformation. **Journal of the Royal Satistical Society**, United Kingdom, v. Series B, n. 26, p. 211–252, 1964.

BRASIL. Ministério da Defesa. Exército Brasileiro. **Especificação técnica para a aquisição de dados geoespaciais vetoriais (ET-ADGV).** [Brasília]: Diretoria de Serviço Geográfico do Exército Brasileiro, [2009]. 225 p. Disponível em: http://esteio.com.br/downloads/legislacao/ADGV_FINAL_18_02_09_final_2.pdf. Acesso em: 13 fev. 2020.

BRUNGARD, C. W. *et al.* Machine learning for predicting soil classes in three semi-arid landscapes. **Geoderma**, Netherlands, v. 239–240, p. 68–83, 2015. Disponível em: https://doi.org/10.1016/j.geoderma.2014.09.019. Acesso em: 13 fev. 2020.

BURROUGH, P. A. Principles of geographical information systems for land resources assessment. **Geocarto International**, China, v. 1, n. 3, p. 54–54, 1986. Disponível em: https://doi.org/10.1080/10106048609354060. Acesso em: 29 nov. 2018.

CALDERANO FILHO, B. *et al.* Artificial neural networks applied for soil class prediction in mountainous landscape of the Serra Do Mar. **Revista Brasileira De Ciencia Do Solo**, Viçosa, v. 38, n. 6, p. 1681–1693, 2014. Disponível em: https://doi.org/101590S01000683201400060003. Acesso em: 6 set. 2017.

CAMPOS, A. R. *et al.* Prediction of soil classes with data collected in buffers delimited pixels in georeferenced soil profiles. **Revista Brasileira de Ciencias Agrarias**, Recife, v. 14, n. 2, p. 1–9, 2019a. Disponível em: https://doi.org/10.5039/agraria.v14i2a5653. Acesso em: 27 mar. 2019.

CAMPOS, A. R. *et al.* Selection of Environmental covariates for classifier training applied in digital soil mapping. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 42, p. 1–15, 2019b. Disponível em: https://doi.org/10.1590/18069657rbcs20170414. Acesso em: 12 dez. 2019.

CANCIAN, L. C.; DALMOLIN, R. S. D.; CATEN, A. T. Bibliometric analysis for pattern exploration in worldwide digital soil mapping publications. **Anais da Academia Brasileira de Ciências**, Rio de Janeiro, v. 90, n. 4, p. 3911–3923, 2018. Disponível em: https://doi.org/10.1590/0001-3765201820180423. Acesso em: 13 fev. 2020.

CANDEL, A.; PARMAR, V. **Deep learning with H2O**. Mountain View, CA: H2O.ai, Inc., 2014.

CARVALHO JÚNIOR, W. *et al.* Digital soilscape mapping of tropical hillslope areas by neural networks. **Scientia Agricola**, Piracicaba, v. 68, n. 6, p. 691–696, 2011. Disponível em: https://doi.org/10.1590/S0103-90162011000600014. Acesso em: 13 dez. 2017.

CHAGAS, C. da S. *et al.* Atributos topográficos e dados do Landsat7 no mapeamento digital de solos com uso de redes neurais. **Pesquisa Agropecuária Brasileira**, Brasília, v. 45, n. 5, p. 497–507, 2010. Disponível em: https://doi.org/10.1590/S0100-204X2010000500009. Acesso em: 12 set. 2017.

CHAGAS, C. da S. *et al.* Data mining methods applied to map soil units on tropical hillslopes in Rio de Janeiro, Brazil. **Geoderma Regional**, Netherlands, v. 9, p. 47–55, 2017. Disponível em: https://doi.org/10.1016/j.geodrs.2017.03.004. Acesso em: 6 set. 2017.

CHAGAS, C. ; OLIVEIRA, C. A.; FERNANDES, E. I. Comparison between artificial neural networks and maximum likelihood classification in digital soil mapping. **Revista Brasileira de Ciência do solo**, Viçosa, v. 37, n. 2, p. 339–351, 2013. Disponível em: http://www.scielo.br/pdf/rbcs/v37n2/a05v37n2.pdf. Acesso em: 14 set. 2017.

CHAGAS, C. S.; CARVALHO JÚNIOR, W.; BHERING, S. B. Integration of Quickbird data and terrain attributes for digital soil mapping by artificial neural networks. **Revista Brasileira de Ciencia do Solo**, Viçosa, v. 35, n. 3, p. 693–704, 2011. Disponível em: https://www.scielo.br/scielo.php?pid=S0100-06832011000300004&script=sci_abstract. Acesso em: 14 set. 2017.

CHEN, G. *et al.* Geographic object-based image analysis (GEOBIA): emerging trends and future opportunities. **GIScience & Remote Sensing**, United States, v. 55, n. 2, p. 159–182, 2018. Disponível em: https://doi.org/10.1080/15481603.2018.1426092. Acesso em: 13 dez. 2019.

COELHO, F. F.; GIASSON, E. Comparação de métodos para mapeamento digital de solos com utilização de sistema de informação geográfica. **Ciência Rural**, Santa Maria, v. 40, n. 10, p. 2099–2106, 2010. Disponível em: https://doi.org/10.1590/S0103-84782010005000156. Acesso em: 7 set. 2017.

COLLARD, F. *et al.* Refining a reconnaissance soil map by calibrating regression models with data from the same map (Normandy, France). **Geoderma Regional**, Netherlands, v. 1, n. C, p. 21–30, 2014. Disponível em:

https://doi.org/10.1016/j.geodrs.2014.07.001. Acesso em: 6 mar. 2020.

COSTA, E. M.; SAMUEL-ROSA, A.; ANJOS, L. H. C. Digital elevation model quality on digital soil mapping prediction accuracy. **Ciência e Agrotecnologia**, Lavras, v. 42, n. 6, p. 608–622, 2018. Disponível em: https://doi.org/10.1590/1413-70542018426027418. Acesso em: 12 dez. 2019.

COSTA, H. R. do N. **Aplicação de técnicas de inteligência artificial em processos de fabricação de vidro**. 2006. Tese (Doutorado em Engenharia) - Escola Politécnica, Universidade de São Paulo, São Paulo, 2006.

CRIVELENTI, R. C. *et al.* Data mining to infer soil-landscape relationships in digital soil mapping. **Pesquisa Agropecuária Brasileira**, Brasília, v. 44, n. 12, p. 1707–1715, 2009. Disponível em: https://doi.org/10.1590/S0100-204X2009001200021. Acesso em: 19 set. 2017.

DEMATTÊ, J. A. M. *et al.* Remote sensing from ground to space platforms associated with terrain attributes as a hybrid strategy on the development of a pedological map. **Remote Sensing**, Switzerland, v. 8, n. 10, p. 826, 2016. Disponível em: https://doi.org/10.3390/rs8100826. Acesso em: 13 dez. 2017.

DEMATTÊ, J. A. M. *et al.* Soil class and attribute dynamics and their relationship with natural vegetation based on satellite remote sensing. **Geoderma**, Netherlands, v. 302, n. March 2016, p. 39–51, 2017. Disponível em: https://doi.org/10.1016/j.geoderma.2017.04.019. Acesso em: 12 set. 2017.

DESCLÉE, B.; BOGAERT, P.; DEFOURNY, P. Forest change detection by statistical object-based method. **Remote Sensing of Environment**, United States, v. 102, n. 1–2, p. 1–11, 2006. Disponível em: https://doi.org/10.1016/j.rse.2006.01.013. Acesso em: 29 nov. 2018.

DIAS, L. M. S. *et al.* Soil class prediction by data mining in an area of the sedimentary São Francisco basin. **Pesquisa Agropecuária Brasileira**, Brasília, v. 51, n. 9, p. 1396–1404, 2016. Disponível em: https://doi.org/10.1590/s0100-204x2016000900038. Acesso em: 18 set. 2017.

DORNIK, A.; DRÂGUŢ, L.; URDEA, P. Classification of soil types using geographic object-based image analysis and Random Forest. **Pedosphere**, China, v. 28, n. 6, p. 913–925, 2017. Disponível em: https://doi.org/10.1016/S1002-0160(17)60377-1. Acesso em: 18 dez. 2018.

DRĂGUŢ, L.; BLASCHKE, T. Automated classification of landform elements using object-based image analysis. **Geomorphology**, Netherlands, v. 81, n. 3–4, p. 330–344, 2006. Disponível em: https://doi.org/10.1016/j.geomorph.2006.04.013. Acesso em: 6 set. 2017.

DRĂGUŢ, L.; EISANK, C. Automated object-based classification of topography from SRTM data. **Geomorphology**, Netherlands, v. 141–142, p. 21–33, 2012. Disponível em: https://doi.org/10.1016/j.geomorph.2011.12.001. Acesso em: 13 dez. 2018.

EL-NAGGAR, A. M. Determination of optimum segmentation parameter values for extracting building from remote sensing images. **Alexandria Engineering**

Journal, Egypt, v. 57, n. 4, p. 3089–3097, 2018. Disponível em: https://doi.org/10.1016/j.aej.2018.10.001. Acesso em: 14 maio 2020.

FERNANDES, R. R.; NUNES, G. M.; DRESCHER, R. Monitoramento de propriedades rurais através de dados multisensores em nível orbital. **Revista Ambiência**, Guarapuava, v. 7, n. 1, p. 75–87, 2011. Disponível em: https://doi.org/10.5777/ambiencia.2011.01.06. Acesso em: 29 nov. 2018.

FIGUEIREDO, S. R. *et al.* Multiple logistic regression applied to soil survey in Rio Grande do Sul state, Brazil. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 32, p. 2779–2785, 2008.

FLORES, C. A. *et al.* **Os solos do Vale dos Vinhedos**. Brasília - DF: Embrapa, 2012.

GARDNER, M. .; DORLING, S. . Artificial neural networks (the multilayer perceptron): a review of applications in the atmospheric sciences. **Atmospheric Environment**, United Kingdom, v. 32, n. 14–15, p. 2627–2636, 1998. Disponível em: https://doi.org/10.1016/S1352-2310(97)00447-0. Acesso em: 28 set. 2018.

GEDEON, T. D. Data mining of inputs: analysing magnitude and functional measures. **International Journal of Neural Systems**, Singapore, v. 08, n. 02, p. 209–218, 1997. Disponível em: https://doi.org/10.1142/S0129065797000227. Acesso em: 14 maio 2020.

GERÇEK, D.; TOPRAK, V.; STROBL, J. Object-based classification of landforms based on their local geometry and geomorphometric context. **International Journal of Geographical Information Science**, United Kingdom, v. 25, n. 6, p. 1011–1023, 2011. Disponível em: https://doi.org/10.1080/13658816.2011.558845. Acesso em: 14 nov. 2019.

GIASSON, E *et al.* Evaluation of five algorithms of decision trees and three digital elevation models for digital soil mapping at semidetail level at the Lageado Grande Watershed, RS, Brazil. **Ciencia Rural**, Santa Maria, v. 43, n. 11, p. 1967–1973, 2013. Disponível em: https://doi.org/10.1590/S0103-84782013001100008. Acesso em: 12 set. 2017.

GIASSON, Elvio *et al.* Decision trees for digital soil mapping on subtropical basaltic steeplands. **Scientia Agricola**, Piracicaba, v. 68, n. 2, p. 167–174, 2011. Disponível em: https://doi.org/10.1590/S0103-90162011000200006. Acesso em: 14 set. 2017.

GIASSON, Elvio *et al.* Digital soil mapping using multiple logistic regression on terrain parameters in southern Brazil. **Scientia Agricola**, Piracicaba, v. 63, n. 3, p. 262–268, 2006. Disponível em: https://doi.org/10.1590/S0103-90162006000300008. Acesso em: 7 set. 2017.

GIASSON, Elvio *et al.* Instance selection in digital soil mapping: a study case in Rio Grande do Sul, Brazil. **Ciência Rural**, Santa Maria, v. 45, n. 9, p. 1592–1598, 2015. Disponível em: https://doi.org/10.1590/0103-8478cr20140694. Acesso em: 14 set. 2017.

GONZALES, R. C.; WOODS, R. E. **Digital image processing.** 3rd ed. United States: Prentice-Hall, Inc., 2006.

GRINAND, C. *et al.* Extrapolating regional soil landscapes from an existing soil map: Sampling intensity, validation procedures, and integration of spatial context. **Geoderma**, Netherlands, v. 143, n. 1–2, p. 180–190, 2008. Disponível em: https://doi.org/10.1016/j.geoderma.2007.11.004. Acesso em: 13 fev. 2020.

HÖFIG, P.; GIASSON, E.; VENDRAME, P. R. S. Mapeamento digital de solos com base na extrapolação de mapas entre áreas fisiograficamente semelhantes. **Pesquisa Agropecuária Brasileira**, Brasília, v. 49, n. 12, p. 958–966, 2014. Disponível em: https://doi.org/10.1590/S0100-204X2014001200006. Acesso em: 14 jul. 2017.

IPPOLITI R., G. A. *et al.* Análise digital do terreno: ferramenta na identificação de pedoformas em microbacia na região de “Mar de Morros” (MG). **Revista Brasileira de Ciência do Solo**, Viçosa, v. 29, n. 2, p. 269–276, 2005. Disponível em: https://doi.org/10.1590/S0100-06832005000200012. Acesso em: 29 nov. 2018.

JENNY, H. Factors of Soil Formation. **Soil Science**, United States, v. 52, n. 5, p. 415, 1941. Disponível em: https://doi.org/10.1097/00010694-194111000-00009. Acesso em: 11 jul. 2018.

KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. **Psychometrika**, United States, v. 23, n. 3, p. 187–200, 1958. Disponível em: https://doi.org/10.1007/BF02289233. Acesso em: 12 nov. 2019.

KARAKIŞ, S.; MARANGOZ, A. M.; BÜYÜKSALIH, G. Analysis of segmentation parameters in ecognition software. *In*: ISPRS Workshop on Topographic Mapping from Space, 36., 2006, Ankara, Turkey. **Proceedings** […]. Ankara, Turkey: [*s. n.*], 2006. p. 14–16. Disponível em: http://www.isprs.org/proceedings/XXXVI/1-W41/makaleler/Karakis_Segmentation_Parameters.pdf. Acesso em: 29 abr. 2020.

KHALEDIAN, Y.; MILLER, B. A. Selecting appropriate machine learning methods for digital soil mapping. **Applied Mathematical Modelling**, United States, v. 81, p. 401–418, 2020. Disponível em: https://doi.org/10.1016/j.apm.2019.12.016. Acesso em: 13 fev. 2020.

LAGACHERIE, P.; MCBRATNEY, A. B. Soil information systems and spatial soil inference systems: perspectives for digital soil mapping. *In*: LAGACHERIE, P.; MCBRATNEY, A. B.; VOLTZ, M. (Ed.). **Digital soil mapping**: an introductory perspective. Netherlands: Elsevier, 2007. v. 31, Chapter 1, p. 3–22. Disponível em: https://doi.org/10.1016/S0166-2481(06)31001-X. Acesso em: 28 dez. 2017.

MA, Y. *et al.* Pedology and digital soil mapping (DSM). **European Journal of Soil Science**, United Kingdom, v. 70, n. 2, p. 216–235, 2019. Disponível em: https://doi.org/10.1111/ejss.12790. Acesso em: 13 fev. 2020.

MACHADO, D. F. T. *et al.* Soil type spatial prediction from Random Forest:

different training datasets, transferability, accuracy and uncertainty assessment. **Scientia Agricola**, Piracicaba, v. 76, n. 3, p. 243–254, 2019. Disponível em: https://doi.org/10.1590/1678-992x-2017-0300. Acesso em: 5 jun. 2019.

MARTHA, T. R. *et al.* Detection of coastal landforms in a deltaic area using a multi-scale object-based classification method. **Current Science**, India, v. 114, n. 06, p. 1338, 2018. Disponível em: https://doi.org/10.18520/cs/v114/i06/1338-1345. Acesso em: 14 nov. 2019.

MAXWELL, A. E. *et al.* Assessing machine-learning algorithms and image- and lidar-derived variables for GEOBIA classification of mining and mine reclamation. **International Journal of Remote Sensing**, United Kingdom, v. 36, n. 4, p. 954–978, 2015. Disponível em: https://doi.org/10.1080/01431161.2014.1001086. Acesso em: 24 abr. 2020.

MCBRATNEY, A.; MENDONÇA SANTOS, M.; MINASNY, B. On digital soil mapping. **Geoderma**, Netherlands, v. 117, n. 1–2, p. 3–52, 2003. Disponível em: https://doi.org/10.1016/S0016-7061(03)00223-4. Acesso em: 11 jul. 2018.

MEIER, M. *et al.* Digital soil mapping using machine learning algorithms in a tropical mountainous area. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 42, p. 1–22, 2018. Disponível em: https://doi.org/10.1590/18069657rbcs20170421. Acesso em: 12 dez. 2019.

MINASNY, B.; MCBRATNEY, A. B. Digital soil mapping: a brief history and some lessons. **Geoderma**, Netherlands, v. 264, n. August, p. 301–311, 2016. Disponível em: https://doi.org/10.1016/j.geoderma.2015.07.017. Acesso em: 12 set. 2017.

MOURA-BUENO, J. M. *et al.* Prediction of soil classes in a complex landscape in Southern Brazil. **Pesquisa Agropecuária Brasileira**, Brasília, v. 54, 2019. Disponível em: https://doi.org/10.1590/s1678-3921.pab2019.v54.00420. Acesso em: 12 dez. 2019.

MYINT, S. W. *et al.* Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. **Remote Sensing of Environment**, United States, v. 115, n. 5, p. 1145–1161, 2011. Disponível em: https://doi.org/10.1016/j.rse.2010.12.017. Acesso em: 29 nov. 2018.

PELEGRINO, M. H. P. *et al.* Mapping soils in two watersheds using legacy data and extrapolation for similar surrounding areas. **Ciência e Agrotecnologia**, Lavras, v. 40, n. 5, p. 534–546, 2016. Disponível em: https://doi.org/10.1590/1413-70542016405011416. Acesso em: 18 set. 2017.

POLIDORO, J. C. *et al.* **Programa nacional de solos do Brasil (PronaSolos)**. Rio de Janeiro: Embrapa Solos, 2016.

PONTIUS, R. G.; MILLONES, M. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. **International Journal of Remote Sensing**, United Kingdom, v. 32, n. 15, p. 4407–4429, 2011. Disponível em: https://doi.org/10.1080/01431161.2011.552923. Acesso em: 13 fev. 2020.

POPPIEL, R. R. *et al.* Pedology and soil class mapping from proximal and remote sensed data. **Geoderma**, Netherlands, v. 348, n. October 2018, p. 189–206, 2019. Disponível em: https://doi.org/10.1016/j.geoderma.2019.04.028. Acesso em: 12 dez. 2019.

PRUDENTE, V. H. R. *et al.* Comparative assessment between per-pixel and object-oriented for mapping land cover and use. **Engenharia Agrícola**, Jaboticabal, v. 37, n. 5, p. 1015–1027, 2017. Disponível em: https://doi.org/10.1590/1809-4430-eng.agric.v37n5p1015-1027/2017. Acesso em: 14 maio 2020.

RENNÓ, C. D. *et al.* HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia. **Remote Sensing of Environment**, United States, v. 112, n. 9, p. 3469–3481, 2008. Disponível em: https://doi.org/10.1016/j.rse.2008.03.018. Acesso em: 29 nov. 2018.

RIZZO, R. *et al.* Digital soil mapping at local scale using a multi-depth Vis–NIR spectral library and terrain attributes. **Geoderma**, Netherlands, v. 274, p. 18–27, 2016. Disponível em: https://doi.org/10.1016/j.geoderma.2016.03.019. Acesso em: 13 dez. 2020.

ROSS, J. L. S. Análise empírica da fragilidade dos ambientes naturais antropizados. **Revista do Departamento de Geografia**, São Paulo, v. 8, p. 63–74, 1994. Disponível em: https://doi.org/10.7154/RDG.1994.0008.0006. Acesso em: 5 jun. 2019.

SANTOS, H. G. *et al.* Distribuição espacial dos níveis de levantamento de solos no Brasil. *In*: CONGRESSO BRASILEIRO DE CIÊNCIA DO SOLO, 34., 2013, Florianópolis. **Anais** […]. [*S. l.*]: SBCS, 2013. p. 1–4.

SARMENTO, E. C. *et al.* Prediction of soil orders with high spatial resolution: response of different classifiers to sampling density. **Pesquisa Agropecuária Brasileira**, Brasília, v. 47, n. 9, p. 1395–1403, 2012. Disponível em: https://doi.org/10.1590/S0100-204X2012000900025. Acesso em: 19 set. 2017.

SCHULTZ, B. **Análise de imagens orientada a objetos e amostragem estatística no monitoramento de cana-de-açúcar, milho e soja no Estado de São Paulo**. 2016. Tese (Doutorado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais - INPE, São José dos Campos, 2016.

SILVA, B. P. C. *et al.* Digital soil mapping including additional point sampling in Posses ecosystem services pilot watershed, southeastern Brazil. **Scientific Reports**, United Kingdom, v. 9, n. 1, p. 13763, 2019. Disponível em: https://doi.org/10.1038/s41598-019-50376-w. Acesso em: 12 dez. 2019.

SILVA, C. C. *et al.* Digital pedological mapping of Botucatu sheet (SF-22-Z-B-VI-3): data training on conventional maps and field validation. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 37, n. 4, p. 846–857, 2013. Disponível em: https://doi.org/10.1590/S0100-06832013000400003. Acesso em: 19 set. 2017.

SILVA, S. *et al.* Proximal sensing and digital terrain models applied to digital soil mapping and modeling of Brazilian Latosols (Oxisols). **Remote Sensing**, Switzerland, v. 8, n. 8, p. 614, 2016. Disponível em: https://doi.org/10.3390/rs8080614. Acesso em: 13 dez. 2017.

SILVA, S. H. G. *et al.* Retrieving pedologist's mental model from existing soil map and comparing data mining tools for refining a larger area map under similar environmental conditions in Southeastern Brazil. **Geoderma**, Netherlands, v. 267, p. 65–77, 2016. Disponível em: https://doi.org/10.1016/j.geoderma.2015.12.025. Acesso em: 13 dez. 2017.

SILVERO, N. E. Q. *et al.* Protocol for the use of legacy data and magnetic signature on soil mapping of São Paulo Central West, Brazil. **Science of The Total Environment**, Netherlands, v. 693, p. 133463, 2019. Disponível em: https://doi.org/10.1016/j.scitotenv.2019.07.269. Acesso em: 12 dez. 2019.

TEAM, R. C. **R**: a language and environment for statistical computing. Vienna, Austria: [*s. n.*], 2019.

TEN CATEN, A. *et al.* An appropriate data set size for digital soil. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 37, n. 1, p. 359–366, 2013. Disponível em: https://doi.org/http://dx.doi.org/10.1590/S0100-06832013000200007. Acesso em: 19 set. 2017.

TEN CATEN, A. *et al.* Extrapolation of soil-landscaperelationship from a reference area. **Ciência Rural**, Santa Maria, v. 41, p. 812–816, 2011a. Disponível em: https://doi.org/10.1590/S0103-84782011000500012. Acesso em: 19 set. 2017.

TEN CATEN, A. *et al.* Mapeamento digital de classes de solos: características da abordagem brasileira. **Ciência Rural**, Santa Maria, v. 42, n. 11, p. 1989–1997, 2012. Disponível em: https://doi.org/10.1590/S0103-84782012001100013. Acesso em: 12 set. 2017.

TEN CATEN, A. *et al.* Multiple logistic regressions: controlling factors in applications to soil class prediction. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 35, n. 1, p. 53–62, 2011b. Disponível em: https://doi.org/10.1590/S0100-06832011000100005. Acesso em: 14 set. 2017.

TEN CATEN, A. *et al.* Multivariate analysis applied to reduce the number of predictors in digital soil mapping. **Pesquisa Agropecuária Brasileira**, Brasília, v. 46, n. 5, p. 553–561, 2011c. Disponível em: https://doi.org/10.1590/S0100-204X2011000500014. Acesso em: 19 set. 2017.

TEN CATEN, A. *et al.* Principal components as predictor variables in digital mapping of soil classes. **Ciência Rural**, Santa Maria, v. 41, n. 7, p. 1170–1176, 2011d. Disponível em: https://doi.org/10.1590/S0103-84782011000700011. Acesso em: 12 set. 2017.

TEN CATEN, A.; DALMOLIN, R. S. D.; RUIZ, L. F. C. Digital soil mapping: strategy for data pre-processing. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 36, n. 4, p. 1083–1092, 2012. Disponível em: https://doi.org/10.1590/S0100-06832012000400003. Acesso em: 19 set. 2017.

TESKE, R.; GIASSON, E.; BAGATINI, T. Comparação de esquemas de amostragem para treinamento de modelos preditores no mapeamento digital de classes de solos. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 39, n. 1, p. 14–20, 2015a. Disponível em: https://doi.org/10.1590/01000683rbcs20150344. Acesso em: 14 set. 2017.

TESKE, R.; GIASSON, E.; BAGATINI, T. Comparison of the use of digital elevation models in digital mapping of soils of Dois Irmãos, RS, Brazil. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 38, n. 5, p. 1367–1376, 2014. Disponível em: https://doi.org/10.1590/S0100-06832014000500002. Acesso em: 19 set. 2017.

TESKE, R.; GIASSON, E.; BAGATINI, T. Production of a soil map associating common digital soil mapping techniques with hand delineation of soil mapping units. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 39, n. 4, p. 950–959, 2015b. Disponível em: https://doi.org/10.1590/01000683rbcs20140285. Acesso em: 18 set. 2017.

VASQUES, G. M. *et al.* 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**, United Kingdom, v. 66, n. 4, p. 767–779, 2015. Disponível em: https://doi.org/10.1111/ejss.12255. Acesso em: 13 dez. 2017.

VIEIRA, M. A. *et al.* Object Based Image Analysis and Data Mining applied to a remotely sensed Landsat time-series to map sugarcane over large areas. **Remote Sensing of Environment**, United States, v. 123, p. 553–562, 2012. Disponível em: https://doi.org/10.1016/j.rse.2012.04.011. Acesso em: 29 nov. 2018.

WOLSKI, M. S. *et al.* Digital soil mapping and its implications in the extrapolation of soil-landscape relationships in detailed scale. **Pesquisa Agropecuaria Brasileira**, Brasília, v. 52, n. 8, p. 633–642, 2017. Disponível em: https://doi.org/10.1590/S0100-204X2017000800009

ZHANG, G.; LIU, F.; SONG, X. Recent progress and future prospect of digital soil mapping: A review. **Journal of Integrative Agriculture**, China, v. 16, n. 12, p. 2871–2885, 2017. Disponível em: https://doi.org/10.1016/S2095-3119(17)61762-3. Acesso em: 13 fev. 2018.

ZUUR, A. F.; IENO, E. N.; ELPHICK, C. S. A protocol for data exploration to avoid common statistical problems. **Methods in Ecology and Evolution**, United States, v. 1, n. 1, p. 3–14, 2010. Disponível em: https://doi.org/10.1111/j.2041-210X.2009.00001.X