



Mapeamento pedológico digital com base em objetos a partir de unidades elementares de relevo

Digital soil mapping based on objects from elementary relief units

Alan Pessoa Valadares*¹ ; Ricardo Marques Coelho ;
Carlos Tadeu de Carvalho Gamba 

¹Instituto Agrônomo, Campinas, SP, Brasil.

²Instituto de Pesquisas Tecnológicas (IPT), São Paulo, SP, Brasil.

E-mail: ricardo.coelho@sp.gov.br (RMC); carlosgamba@ipt.br (CMCG).

*Email para correspondência: alanpeval@gmail.com

Recebido (Received): 23/03/2023

Aceito (Accepted): 29/05/2024

Resumo: Diante das dificuldades em se realizar levantamentos pedológicos por métodos convencionais, o mapeamento pedológico digital (MPD) tem potencial de apoiar a retomada de levantamentos pedológicos sistemáticos. Para isso, é desejável estratégias que possam utilizar informações legadas de levantamentos convencionais. Contudo, é preciso entender quais informações podem ser preditas no MPD de novas áreas, com qualidade satisfatória. Estratégias de MPD com base em objetos têm superado os resultados de abordagens com base em pixels, embora não esteja claro se isso vale para qualquer contexto. Este estudo objetivou avaliar o uso de informações legadas de levantamentos pedológicos sistemáticos, comparando abordagens de MPD com base em pixels e em objetos. A área de estudo compreende os municípios paulistas de Cordeirópolis, Santa Gertrudes, Iracemápolis, Rio Claro e Ipeúna. Utilizou-se unidades elementares de relevo para delimitação de polígonos na construção de bases de dados para MPD com base em objetos. Os resultados foram avaliados considerando a acurácia global, o índice Kappa, além das taxas de verdadeiros positivos e a precisão observadas nas classes preditas. As unidades elementares de relevo foram eficazes quando utilizadas em uma abordagem de MPD com base em objetos. Os testes empregados apresentaram resultados satisfatórios para a extrapolação de informações legadas dos levantamentos pedológicos sistemáticos do estado de São Paulo em abordagens de MPD por pixels e por objetos, alcançando acurácia de 61 % e índice Kappa de 0,54. O baixo desempenho na predição de classes específicas não esteve diretamente relacionado à sua maior ou menor ocorrência na área de estudo.

Palavras-chave: Pedometria; Random Forest; Mapeamento geomorfológico.

Abstract: *Given the difficulties in carrying out pedological surveys using conventional methods, digital mapping of soil classes (DMSC) has the potential to support the resumption of systematic pedological surveys. For this, strategies that use legacy information from conventional surveys are desirable. However, it is necessary to understand what information can be predicted in the DMSC of new areas with satisfactory quality. Object-based DSMC strategies have outperformed pixel-based approaches, although it is unclear whether this is true in any context. This study aimed to evaluate the use of legacy information from systematic soil surveys, comparing pixel-based and object-based DMSC approaches. The study area comprises the São Paulo municipalities of Cordeirópolis, Santa Gertrudes, Iracemápolis, Rio Claro and Ipeúna. Elementary relief units were used to delimit polygons in the construction of databases for DMSC based on objects. The results were evaluated considering the global accuracy, Kappa index, in addition to the true positive rates and the precision observed in the predicted classes. Elementary relief units were effective when used in an object-based DMSC approach. The tests employed showed satisfactory results for the extrapolation of legacy information from systematic pedological surveys in the state of São Paulo in pixel and object DMSC approaches, reaching an accuracy of 61 % and a Kappa index of 0.54. The low performance in predicting specific classes was not directly related to their greater or lesser occurrence in the study area.*

Keywords: Soil survey; Pedometrics, Random forest; Geomorphological mapping.

1. Introdução

A falta de informações sobre solos em nível adequado ao planejamento regional é uma lacuna no conhecimento das características do meio físico, inviabilizando que as especificidades dos diferentes tipos de solos sejam consideradas no planejamento e na tomada de decisão, o que gera prejuízos econômicos e ao meio ambiente, a médio e longo prazo (SARTORI *et al.*, 2019). Essa é uma realidade que ocorre em todas as unidades federativas do Brasil e que se observa também em outros países em desenvolvimento.

O mapeamento pedológico sistemático em nível de semidetalhe no estado de São Paulo cobre apenas cerca de 15% do seu território. Esses levantamentos tiveram início em 1975 com os trabalhos realizados na Seção de Pedologia do Instituto Agrônomo (IAC), que resultaram na publicação da folha de Campinas (OLIVEIRA *et al.*, 1977), na escala de 1:100.000. Como resultado, foram elaborados ao todo 14 mapas pedológicos de nível semidetalhado (1:100.000), seguindo a articulação do Sistema Cartográfico Nacional (IBGE, 2011), até o ano 2000 com a publicação da folha de Marília (COELHO *et al.*, 2000).

Restabelecer a execução de mapeamentos pedológicos, pelo menos em nível de reconhecimento de alta intensidade (IBGE, 2015), é essencial para que existam informações sobre a distribuição de classes de solos com detalhe e qualidade suficientes para apoiar o planejamento regional que possa levar em consideração as características dos diferentes tipos de solo. Os levantamentos pedológicos que existem hoje com cobertura em todo o estado de São Paulo, como os de Oliveira *et al.* (1999), 1:500.000 e Rossi (2017), 1:250.000 e 1:100.000 não apresentam nível de detalhe adequado em todo o estado para análises em nível municipal ou em microbacias hidrográficas, por exemplo.

Diante da elevada exigência de recursos financeiros, logísticos, materiais e humanos para realização de levantamentos pedológicos por métodos exclusivamente convencionais, as metodologias de mapeamento digital de solos (MDS) (MCBRATNEY *et al.*, 2003) são vistas como complementares e trazem uma expectativa de redução do tempo e do custo de elaboração de mapas pedológicos, permitindo também que extensas regiões sejam previamente mapeadas (TEN CATEN *et al.*, 2012), orientando expedições de campo no mapeamento convencional.

Alguns estudos recentes utilizaram técnicas de MDS para mapeamento digital de classes de solo, ou mapeamento pedológico digital (MPD), em contextos brasileiros (CARVALHO JUNIOR *et al.*, 2020). O MPD lança mão de métodos e tecnologias relativamente recentes, que envolvem o processamento digital de dados provenientes de sensoriamento remoto e de mapeamentos temáticos, utilizando também procedimentos de Mineração de Dados e de Aprendizagem de Máquina (WITTEN *et al.*, 2016).

O uso das técnicas de MPD permite a predição de classes de solo de forma satisfatória para áreas não mapeadas (SILVA *et al.*, 2013; CARVALHO JUNIOR *et al.*, 2020), a partir do conhecimento adquirido em áreas de referência, como mostram Wolski *et al.* (2017). Em alguns casos, os resultados do MPD apresentaram mapas com acurácia superior a 75 %, confrontando os mapas com observações obtidas em campo (HÖFIG *et al.*, 2014).

É desejável implementar estratégias que possam ampliar o uso de informações legadas de levantamentos convencionais. Estas informações estão presentes nos mapas, boletins técnicos e nas bases de dados de amostragem do solo em campo e de análises realizadas em laboratório. Essas informações têm potencial de apoiar e até mesmo viabilizar a retomada de levantamentos pedológicos sistemáticos (MCBRATNEY *et al.*, 2003).

Contudo, algumas questões em relação às metodologias de MPD permanecem pouco investigadas. Wolski *et al.* (2017) perceberam, por exemplo, que há limitações na extrapolação de classes de solo com baixa ocorrência nas áreas de referência, ou seja, nem toda a diversidade de solos presente nos levantamentos legados será predita com qualidade no mapeamento de novas áreas. Portanto, é importante entender se existem e quais são as informações disponíveis nestes levantamentos que podem ser utilizadas no MPD de novas áreas, com qualidade satisfatória.

Recentemente, estratégias de classificação supervisionada com base em objetos têm alcançado melhores resultados, comparada à classificação com base em pixels. Dornik *et al.* (2018) e Cremon *et al.* (2021) melhoraram a *performance* dos mapas pedológicos digitais utilizando classificação com base em objetos. Drăguț e Blaschke (2006) utilizaram a segmentação de imagens para automatizar a classificação das formas elementares de relevo por meio de análise orientada a objetos (GEOBIA – Geographic Object-based Image Analysis) e identificaram grandes vantagens na utilização desta abordagem, dentre elas a boa reprodutibilidade.

No conceito de unidades elementares de relevo, apresentado por Colangelo (1996), as feições são definidas pela homogeneidade dos valores das propriedades morfométricas fundamentais e são delimitados

pela descontinuidade nos valores destas propriedades. Este autor exemplificou as nove unidades elementares derivadas da combinação entre os três tipos de curvaturas em perfil (convexa, retilínea e côncava) e em planta (divergente, retilínea e convergente), classificando as formas combinadas em três domínios (dispersão, transição e concentração), de acordo com o comportamento dos fluxos hídricos superficiais e nove formas (hiperdispensor, mesodispensor longitudinal, hipodispensor, mesodispensor radial, transição, mesoconcentrador radial, hipocentrador, mesoconcentrador longitudinal e hiperconcentrador).

Contudo, é preciso investigar se o resultado encontrado por Drăguț e Blaschke (2006) pode ser reproduzido em outros contextos, no que tange ao nível de classificação de solos, ao tamanho do conjunto de dados, às covariáveis preditivas utilizadas e às classes de solos preditas, uma vez que essas condições variam bastante de um estudo a outro (CARVALHO JUNIOR *et al.*, 2020).

Além disso, a própria metodologia de delimitação dos polígonos na classificação por objetos para fins de MPD foi pouco investigada. Unidades geomorfológicas, por exemplo, apresentam padrões de relevo relativamente homogêneos e estão intrinsecamente associadas à ocorrência de unidades litológicas (ROSS, 1992). Ambos são fatores com grande influência na pedogênese. Unidades elementares de relevo poderiam ser utilizadas para gerar unidades cartográficas e de amostragem na construção de bases de dados de treinamento no MPD, em estratégias de classificação supervisionada com base em objetos.

Diante ao exposto, este estudo teve como objetivo avaliar o uso de informações legadas dos levantamentos pedológicos sistemáticos em nível de semidetalhe no estado de São Paulo para gerar novas informações sobre a distribuição de solos, por meio da classificação supervisionada com duas abordagens: classificação com base em pixels e classificação com base em objetos, utilizando unidades elementares de relevo obtidas automaticamente por meio de análise orientada a objetos (GEOBIA).

2. Material e métodos

2.1. Contexto geomorfológico e geológico da área de estudo

A área de estudo situa-se no polo minero-cerâmico de Santa Gertrudes, que compreende os municípios paulistas de Cordeirópolis, Santa Gertrudes, Iracemápolis, Rio Claro e Ipeúna, como mostra a **Figura 1**.

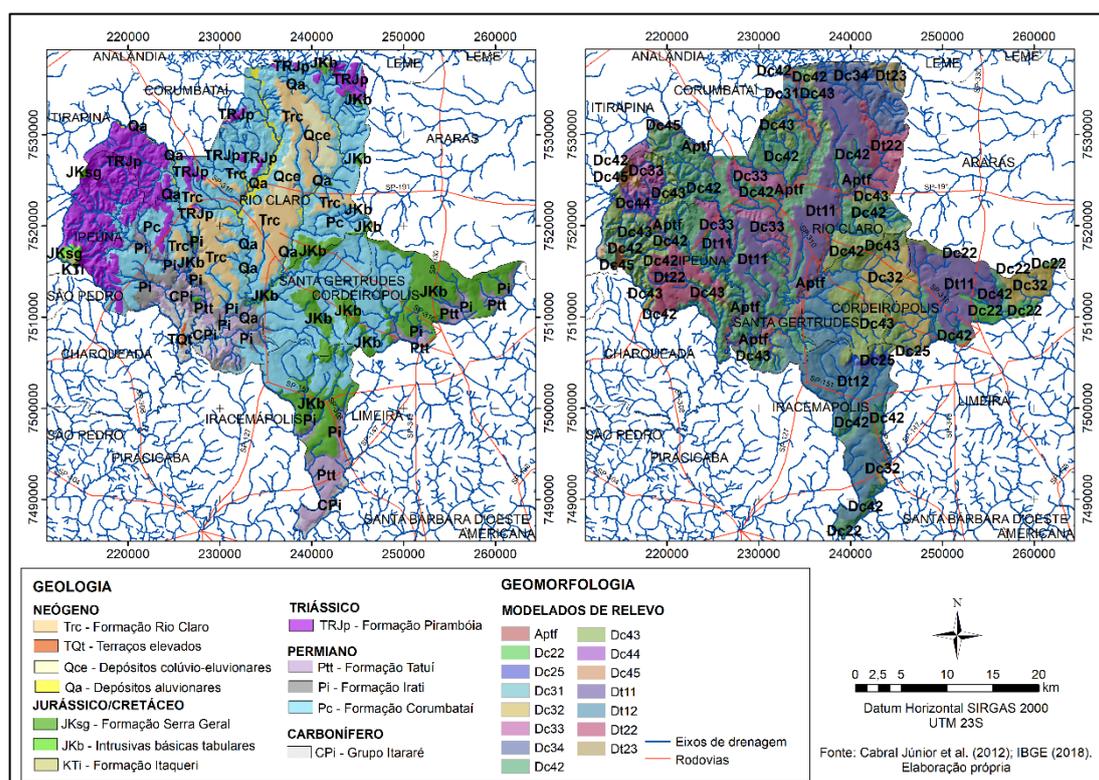


Figura 1: Localização da área de estudo, unidades geológicas (esquerda), escala 1:50.000 (CABRAL JÚNIOR *et al.*, 2012) e modelados de relevo (direita) (IBGE, 2018).

Existe na área de estudo um mapa geológico elaborado por Cabral Júnior *et al.* (2012), que apresenta bom nível de detalhamento (1:50.000) para o objetivo deste estudo. A área está inserida na divisa entre quatro folhas dos levantamentos pedológicos semidetalhados em escala 1:100.000 do estado de São Paulo, são elas: Campinas (OLIVEIRA *et al.*, 1977), São Carlos (PRADO *et al.*, 1981), Araras (OLIVEIRA *et al.*, 1982) e Piracicaba (OLIVEIRA; PRADO, 1989).

Em relação aos materiais de origem, seguindo uma sequência cronológica do mais antigo ao mais recente, encontramos na área de estudo o Grupo Itararé (CPi) na porção sul dos municípios de Ipeúna e Rio Claro, seguido das Formações Tatuí (Ptt) e Irati (Pi), que ocorrem também nos municípios de Iracemápolis e Cordeirópolis, como mostra a **Figura 1**. Na região, o Grupo Itararé é composto por lamitos com seixos, arenitos, folhelhos e siltitos; a Formação Tatuí por siltitos e siltitos arenosos; e a Formação Irati por folhelhos e argilitos cinza escuros, folhelhos betuminosos e calcários associados.

Em grande parte da área está presente a Formação Corumbataí (Pc), que ocorre na porção oeste de Cordeirópolis, norte de Iracemápolis, grande parte do território de Santa Gertrudes e boa parte dos municípios de Rio Claro e Ipeúna, conforme apresentado na **Figura 1**. Esta formação é composta por argilitos, folhelhos e siltitos, associados a arenitos finos quando em sua seção superior. A Formação Pirambóia (TRJp), composta por arenitos médios e finos, ocorre em grande parte do município de Ipeúna e no extremo norte de Rio Claro.

Também se observa na área de estudo rochas Intrusivas Básicas tabulares (JKb), que são comuns nos municípios de Cordeirópolis e Santa Gertrudes, ocorrendo também na porção central de Iracemápolis, são equivalentes intrusivos das rochas vulcânicas da Formação Serra Geral. A Formação Serra Geral (JKsg) é composta por basaltos tholeiíticos e andesi-basaltos tholeiíticos, com augita e pigeonita e ocorre apenas na linha da cuesta, localizada na porção sudoeste de Ipeúna. Acima da linha da cuesta, no município de Ipeúna encontramos também um pequeno trecho da Formação Itaqueri (KTi), como apresentado na **Figura 1**, constituída por membros alternados de arenitos com cimento argiloso, folhelhos e conglomerados.

A Formação Rio Claro (Trc) ocorre em boa parte do platô de Rio Claro, no município de mesmo nome, é composta por arenitos, arenitos conglomeráticos, arenitos argilosos e argilitos intercalados. Também está presente um pequeno trecho de Terraços Elevados (TQt), Depósitos Colúvio-eluvionares (Qce) entre manchas da Formação Rio Claro e Depósitos Aluvionares (Qa) ao longo da calha do rio nas porções central e norte do município de Rio Claro, como mostra a **Figura 1**, compostos por sedimentos de granulometria variada, com predomínio de arenosos e arenoargilosos.

Exceto pelo pequeno trecho na linha de cuesta, compreendendo a escarpa e o reverso da cuesta, onde ocorrem as Formações Serra Geral e Itaqueri, toda a área de estudo encontra-se na província geomorfológica da Depressão Periférica Paulista (ROSS; MOROZ, 1997). Ocorrem, predominantemente, modelados de relevo sob processos de dissecação e topo convexo (Dc), com variados níveis de densidade de drenagem e de graus de entalhamento. Observam-se também trechos de modelados de topo tabular (Dt), associados à área de ocorrência da Formação Rio Claro (Trc) e às manchas contínuas e de maiores dimensões das Intrusivas básicas (JKb), como mostra a **Figura 1**.

Os modelados de relevo onde predominam processos de acumulação de sedimentos ocorrem associados às calhas dos rios de maior fluxo acumulado, associados ou não a Depósitos aluvionares (Qa), como mostram a **Figura 1**.

2.2. *Covariáveis preditivas de classes de solo*

As covariáveis preditivas são essenciais para o MPD por classificação supervisionada. Elas contêm as informações com as quais o modelo preditivo é construído e a partir das quais a ocorrência de classes de solos é predita. Utilizam-se variáveis relacionadas a um ou mais fatores de formação dos solos (clima, organismos, topografia, material parental e tempo) (JENNY, 1941). Além dos fatores de formação, o MPD acrescenta à função preditiva os fatores “vizinhança”, com características de variabilidade espacial e de posicionamento, além do próprio “solo”, como função preditiva do próprio objeto de predição (MCBRATNEY *et al.*, 2003).

Elementos associados ao relevo são os mais frequentemente utilizados no MPD. No mapeamento pedológico convencional, a relação solo-paisagem é traduzida nos delineamentos das unidades de mapeamento (HUDSON, 1992). O relevo também é um dos condicionantes dos processos de remoção, transporte e deposição de materiais que ocorrem na superfície e para caracterizar o relevo considerou-se a morfometria e características geomorfológicas.

As variáveis morfométricas utilizadas neste estudo foram obtidas a partir de modelos digitais de elevação hidrológicamente condicionados (MDEHC), com resolução espacial de 30 metros. Os MDEHC foram elaborados a partir de cartas topográficas na escala 1:50.000, utilizando o método de interpolação ANUDEM (HUTCHINSON, 1988), que melhor se aproxima da representação natural da superfície de drenagem, preservando tanto os topos dos interflúvios quanto os eixos de drenagem. As covariáveis morfométricas primárias foram a elevação, declividade, curvatura em perfil, curvatura em planta, curvatura combinada e direção de fluxo. As covariáveis morfométricas secundárias calculadas foram a área de contribuição e índice de umidade topográfica (BEVEN; KIRKBY, 1979).

As características geomorfológicas utilizadas foram as unidades morfoesculturais, modelados de relevo e pelo conceito de formas ou unidades elementares de relevo (feições mínimas) conforme Colangelo (1996). Como fontes de informações sobre unidades morfoesculturais utilizou-se os mapas geomorfológicos do estado de São Paulo em escala 1:500.000 (ROSS; MOROZ, 1997) e 1:250.000 (IBGE, 2018). Também foram utilizados os modelados de relevo do mapa geomorfológico do IBGE (2018).

Os métodos para mapeamento das formas elementares feitos por meio de interpretação visual são muito subjetivos e demandam bastante tempo de trabalho manual, podendo ser inviável em áreas muito extensas. Neste estudo, a segmentação foi realizada com base nas variáveis morfométricas elevação, declividade, curvatura em perfil e curvatura em planta, calculadas a partir dos MDEHC. Estas quatro variáveis foram utilizadas como dados de entrada para o algoritmo de segmentação de imagens SNIC (Simple Non-interactive Clustering) (ACHANTA; SÜSTRUNK, 2017) e implementado na plataforma Google Earth Engine (GEE) (GORELICK, 2017).

Para caracterização do material parental, utilizou-se as unidades geológicas presentes no mapa geológico de Cabral Júnior *et al.* (2012) para a área de estudo. Para caracterização da vizinhança, foram calculadas a variabilidade espacial considerando a entropia da elevação e da declividade, além da distância até o eixo de drenagem e da distância até topo do interflúvio, que servem como informações sobre o posicionamento. A distância até o eixo de drenagem foi calculada considerando os canais presentes nas cartas topográficas em escala 1:50.000. Os interflúvios foram definidos a partir dos limites entre microbacias hidrográficas elaboradas utilizando um limiar de 5.000 pixels (~ 450 ha) de área de contribuição para definição dos pixels pertencentes à rede de drenagem no MDEHC.

Utilizou-se também como covariáveis preditivas as três principais componentes calculadas a partir da análise de componentes principais que considerou as variáveis morfométricas, as unidades morfoesculturais, o modelado de relevo e o material de origem. A análise de componentes principais é uma técnica da estatística multivariada utilizada para transformar linearmente um conjunto variáveis correlacionadas, gerando um conjunto menor de variáveis não correlacionadas e que explica a maior parte da variabilidade dos dados originais, presente em suas componentes principais (JOHNSON; WICHERN, 1998).

Neste estudo, a variabilidade dos fatores pedogenéticos clima, organismos e tempo foi considerada desprezível entre as classes de solo da área de estudo. Portanto, não foram utilizadas covariáveis preditivas para caracterizar estes fatores de formação dos solos.

2.3. Bases de dados para treinamento e validação

As informações de referência sobre a ocorrência de classes de solo são essenciais para o MPD. Sem estas informações, é impossível realizar o treinamento de modelos para a classificação supervisionada, pois, a parametrização dos modelos se dá a partir dos exemplos fornecidos *a priori* no conjunto de dados de treinamento. A taxonomia de solos utilizada para identificar as classes de solos na base de dados de treinamento define quais serão os níveis de classificação associados às classes de solos que o modelo poderá prever.

Os levantamentos pedológicos convencionais presentes na área de estudo utilizaram o conceito de “Unidade de Solo”. A Unidade de Solo corresponde a uma consociação de solos (Soil Science Division Staff, 2017) que pode apresentar atributos diferenciais não só pedológicos, mas também de ordem geomorfológica e geológica (OLIVEIRA *et al.*, 1982). O conceito de Unidade de Solo é regionalizado e está associado ao material de origem, ao relevo e aos processos de ocorrência regional. Com isso, foi realizada a conversão das Unidades de Solo presentes na área de estudo para a nomenclatura adotada pelo Sistema Brasileiro de Classificação do Solo (SiBCS) (SANTOS *et al.*, 2018). Isso permite maior coerência na extrapolação das informações contidas nas Unidades de Solo para novas áreas, que não necessariamente apresentarão os mesmos atributos diferenciais encontrados nas áreas de referência.

A conversão para a nomenclatura adotada pelo SiBCS foi feita avaliando o conteúdo descritivo e analítico dos memoriais e boletins que acompanham os mapas pedológicos convencionais da área de estudo. A conversão foi realizada até o 4º nível categórico (Subgrupo) do SiBCS, acrescido de informações sobre o horizonte superficial (horizonte A) e do grupamento textural do solo, conforme IBGE (2015).

Como área de referência, foram utilizadas as unidades de mapeamento simples presentes nos mapas pedológicos convencionais semidetalhados, em escala 1:100.000, que cobrem a área de estudo: Campinas (OLIVEIRA, *et al.*, 1979), São Carlos (PRADO *et al.*, 1981), Araras (OLIVEIRA, *et al.*, 1982) e Piracicaba (OLIVEIRA; PRADO, 1989). Nos levantamentos pedológicos semidetalhados, as classes descritas na legenda das unidades de mapeamento simples devem ocorrer em pelo menos 85 % da área da respectiva unidade de mapeamento (IBGE, 2015).

As matrizes de dados de treinamento para a classificação supervisionada com base em pixels foram formadas pela sobreposição das covariáveis preditivas com as áreas das unidades de mapeamento simples dos levantamentos convencionais presentes na área de estudo. Utilizou-se também os polígonos representantes das unidades elementares de relevo para a construção da base de dados de treinamento em uma abordagem de classificação supervisionada com base em objetos (Geographic Object-based Image Analysis – GEOBIA).

As unidades elementares de relevo foram criadas a partir da segmentação do MDEHC utilizando variáveis morfométricas, por meio do algoritmo SNIC (Simple Non-interactive Cluster) (ACHANTA; SÜSTRUNK, 2017). As covariáveis preditivas e as classes de solos foram reamostradas para cada polígono de unidade elementar de relevo, sendo calculada a mediana dos valores dos pixels cujos centroides interceptavam os respectivos polígonos. No caso das covariáveis categóricas e das Unidades de Solo, a reamostragem foi realizada considerando o valor mais frequente (moda) entre os pixels cujos centroides interceptavam os respectivos polígonos.

As bases de dados passaram por uma etapa de testes para a escolha dos tipos de variáveis (categórica ou contínua), identificação de informações inconsistentes e seleção de covariáveis preditivas mais relevantes para os modelos. Os seguintes procedimentos de pré-processamento foram considerados: estratificação, seleção de dados, discretização e seleção de variáveis (WEISS; ZHANG, 2003).

Para testagem da qualidade dos modelos e dos mapas pedológicos digitais foram utilizados 136 pontos de observação em campo presentes nas bases de dados provenientes dos levantamentos convencionais de solos na área de estudo, que não faziam intersecção com as unidades de mapeamento simples utilizadas como área de referência. A **Tabela 1** indica a quantidade de pontos de observação em campo utilizados na etapa de validação dos mapas pedológicos digitais, por Unidade de Solo e sua correspondente consociação na classificação do SiBCS.

Tabela 1: Quantidade de observações utilizadas para validação dos mapas pedológicos digitais produzidos para a área de estudo, por classe de solo classificada no 4º nível do SiBCS, acrescida do tipo de horizonte A e grupamento textural.

Unidade de Solo	Classificação segundo o SiBCS	n
Olaria	NXd típico, A moderado, textura argilosa e muito argilosa	33
Ribeirão Preto	LVef típico, A moderado, textura argilosa e muito argilosa	25
Limeira	LVd típico, A moderado, textura argilosa e muito argilosa	23
Barão Geraldo	LVdf típico, A moderado, textura argilosa e muito argilosa	22
Serrinha	PVAd PV Ae PAd ou PAe arênico abrupto, A moderado e fraco, textura arenosa/média	7
Estruturada	NVef e NVdf típico, A moderado, textura argilosa e muito argilosa	7
Alva	PVAd e PV Ae abrupto, A moderado, textura arenosa/média	5
Laranja Azeda	LVAd típico, A moderado, textura média	4
Campestre	PVe nitossólico e NVe típico, A moderado, textura argilosa/muito argilosa	3
Coqueiro	LVAd psamítico e típico, A moderado e fraco, textura média	3
Hortolândia	LVd típico, A moderado, textura média	2
Camarguinho	LVAd espesso-húmico, textura média, argilosa ou muito argilosa	2

L = Latossolo, N = Nitossolo, P = Argissolo, X = Háptico, A = Amarelo, VA = Vermelho Amarelo, V = Vermelho, L = Litólico, Q = Quartzarênico, d = distrófico, e = eutrófico, f = férrico, ef = eutrófico, df = distroférrico.

2.6. Classificação supervisionada e avaliação dos resultados

Como opção de algoritmo para classificação supervisionada adotou-se o Random Forest (BREIMAN, 2001) do tipo comitê (Ensemble Methods for Prediction) (WEISS; ZHANG, 2003). O Random Forest pode ser considerado uma das técnicas disponíveis mais promissoras para o MPD (CHAGAS *et al.*, 2016; RUDIYANTO *et al.*, 2016; HENGL *et al.*, 2015; HEUNG *et al.*, 2016; HEUNG *et al.*, 2017; SOUZA *et al.*, 2016 *apud* MACHADO *et al.*, 2019). Este classificador avalia várias árvores de decisão aleatórias e diferentes umas das outras, alcançando desempenho equivalente ou superior a algoritmos que utilizam outras abordagens na classificação supervisionada, como árvores de decisão, redes neurais e redes bayesianas (VALADARES *et al.*, 2019).

Os mapas pedológicos digitais foram criados utilizando a plataforma Google Earth Engine (GEE) (GORELICK, 2017). As bases de dados que geraram os melhores ajustes na classificação pelo Random Forest foram transferidas para o GEE, onde a implementação do modelo preditivo e a classificação das áreas a mapear podem ser processados de maneira mais eficaz em uma única plataforma. As informações preditas foram exportadas do GEE e vetorizadas no software ArcGIS 10.1 (ESRI, 2012), onde foram criados os mapas pedológicos digitais.

O desempenho dos classificadores em cada classe de predição foi avaliado pela precisão e pelas taxas de verdadeiros positivos (WEISS; ZHANG, 2003). Para avaliar o desempenho global dos classificadores, utilizou-se a acurácia global e o índice Kappa (COHEN, 1960), calculados confrontando-se as classes de solo presentes nos mapas pedológicos digitais com aquelas observadas em campo. Os intervalos para interpretação do índice Kappa foram os de Landis e Koch (1977).

Considerou-se também uma avaliação dos resultados utilizando uma matriz de confusão ponderada, semelhante ao proposto por Höfig *et al.* (2014). No presente estudo, a matriz de confusão foi ponderada pela proporção da área das unidades de mapeamento simples de cada Unidade de Solo em relação à área total das unidades de mapeamento simples na área de referência.

3. Resultados

Os mapas pedológicos elaborados para a área de estudo utilizando as abordagens de classificação com base em pixels e com base em unidades elementares de relevo (GEOBIA) apresentaram resultados muito próximos, quando comparados pelos valores de acurácia global e índice Kappa. A acurácia foi ligeiramente maior na abordagem por pixels (46,32 %) do que na abordagem por GEOBIA utilizando unidades elementares de relevo (45,93 %), como mostra a **Tabela 2**. Neste caso, a pequena diferença deve-se a um único acerto a mais observado na matriz de confusão do mapa pedológico elaborado por classificação com base em pixels.

Tabela 2: Acurácia global e índice Kappa dos mapas pedológicos digitais produzidos por classificação supervisionada com base em pixels e em unidades elementares de relevo (GEOBIA) utilizando o Random Forest na área piloto.

Classificação	Validação convencional		Validação ponderada ¹	
	OA ² (%)	Kappa	OA ² (%)	Kappa
Pixels	46,32	0,3664	61,00	0,5396
GEOBIA	45,93	0,3668	60,51	0,5375

¹Considerou-se o peso atribuído a cada classe equivalente à proporção da área das observações presente na base de dados de treinamento como fator de ponderação. ²OA = Acurácia global.

Considerando a validação ponderada pela frequência da Unidade de Solo na área de referência, a qualidade dos mapas pedológicos digitais produzidas por ambas abordagens, por pixels e por GEOBIA, pode ser considerada boa (LANDIS; KOCH, 1977). Os valores de acurácia global foram bastante próximos, aproximadamente 61 %, como apresentado na **Tabela 2**.

Apesar da semelhança entre a qualidade dos resultados globais dos mapas produzidos pelas duas abordagens, há diferenças significativas nos resultados observados por classe. Em três Unidades de Solo (PVe nitossólico e NVe típico, A moderado, textura argilosa/muito argilosa; LVd típico, A moderado, textura média; LVAd típico, A moderado, textura média) os resultados com a abordagem por unidades elementares de relevo (GEOBIA) superou a classificação por pixels, em relação à taxa de verdadeiros positivos. Em três outros casos a abordagem por pixels foi superior à classificação por unidades elementares

de relevo (GEOBIA) (LVd típico, A moderado, textura argilosa e muito argilosa; NXd típico, A moderado, textura argilosa e muito argilosa; LVef típico, A moderado, textura argilosa e muito argilosa), como mostra a **Tabela 3**.

Tabela 3: Taxa de verdadeiros positivos (exatidão do usuário) e precisão (exatidão do produtor) para as Unidades de Solo nos mapas pedológicos digitais elaborados por classificação com base em pixels e em unidades elementares de relevo (GEOBIA).

Unidade de Solo	SiBCS	TPR ¹	Precisão ²		Peso (%)	
		Pixels (%)	GEOBIA (%)	Pixels (%)		GEOBIA (%)
Serrinha	PVAd PV Ae PAd ou PAe arênico abrupto, A moderado e fraco, textura arenosa/média	85,7	85,7	54,5	42,8	27,8
Limeira	LVd típico, A moderado, textura argilosa e muito argilosa	60,9	56,5	28,5	30,2	20,5
Olaria	NXd típico, A moderado, textura argilosa e muito argilosa	45,5	42,4	60,0	50,0	15,5
Barão Geraldo	LVdf típico, A moderado, textura argilosa e muito argilosa	18,2	18,2	50,0	57,1	11,1
Ribeirão Preto	LVef típico, A moderado, textura argilosa e muito argilosa	44,0	37,5	52,3	56,2	8,3
Alva	PVAd e PV Ae abrupto, A moderado, textura arenosa/média	100,0	100,0	83,3	100,0	6,9
Estruturada	NVef e NVdf típico, A moderado, textura argilosa e muito argilosa	57,1	57,1	80,0	66,6	4,0
Campestre	PVe nitossólico e NVe típico, A moderado, textura argilosa/muito argilosa	66,7	100,0	33,3	33,3	2,5
Camarguinho	LVAd espesso-húmico, textura média, argilosa ou muito argilosa	50,0	50,0	100,0	50,0	1,2
Laranja Azeda	LVAd típico, A moderado, textura média	0,0	25,0	0,0	50,0	0,9
Hortolândia	LVd típico, A moderado, textura média	50,0	100,0	50,0	66,6	0,6
Coqueiro	LVAd psamítico e típico, A moderado e fraco, textura média	0,0	0,0	-	-	0,2

¹TPR = Taxa de verdadeiros positivos ou exatidão do usuário (número de observações classificadas corretamente em relação ao total de observações reais em determinada classe). ²Precisão = exatidão do produtor (número de observações classificadas corretamente em relação ao total de observações preditas em determinada classe).

As abordagens foram equivalentes em relação as taxas de verdadeiros positivos para seis classes (PVAd e PV Ae abrupto, A moderado, textura arenosa/média; PVAd PV Ae PAd ou PAe arênico abrupto, A moderado e fraco, textura arenosa/média; NVef e NVdf típico, A moderado, textura argilosa e muito argilosa; LVAd espesso-húmico, textura média, argilosa ou muito argilosa; LVdf típico, A moderado, textura argilosa e muito argilosa; LVAd psamítico e típico, A moderado e fraco, textura média), conforme a **Tabela 3**.

A única Unidade de Solo cujas taxas de verdadeiros positivos foram nulas para ambas as abordagens foi a dos LVAd psamítico e típico, A moderado e fraco, textura média. A taxa de verdadeiros positivos foi nula para os LVAd típico, A moderado, textura média, na abordagem de classificação por pixels, como observa-se na **Tabela 3**.

A abordagem de classificação com base em pixels teve melhor desempenho quanto às taxas de verdadeiros positivos em classes com maior ocorrência na área de referência (LVd típico, A moderado, textura argilosa e muito argilosa; NXd típico, A moderado, textura argilosa e muito argilosa; LVef típico, A moderado, textura argilosa e muito argilosa), como apresenta a **Tabela 3**.

As diferenças entre as abordagens de classificação por pixels e por unidades elementares de relevo são ainda maiores considerando a precisão obtida nas classes. Neste caso, apenas o resultado de uma classe (PVe nitossólico e NVe típico, A moderado, textura argilosa/muito argilosa) foi equivalente entre as abordagens. Os LVAd psamítico e típico, A moderado e fraco, textura média não foram preditos pelos modelos de ambas as abordagens nos locais com pontos de validação, portanto, não foi possível avaliar a precisão nesta classe, conforme apresentado na **Tabela 3**.

A classificação com base em unidades elementares de relevo superou os resultados da classificação com base em pixels, em relação à precisão, em seis classes (PVAd e PVAe abrupto, A moderado, textura arenosa/média; LVd típico, A moderado, textura média; LVdf típico, A moderado, textura argilosa e muito argilosa; LVef típico, A moderado, textura argilosa e muito argilosa; LVAd típico, A moderado, textura média; LVd típico, A moderado, textura argilosa e muito argilosa). Houve quatro classes onde a precisão foi maior na abordagem de classificação supervisionada com base em pixels (LVAd espesso-húmico, textura média, argilosa ou muito argilosa; NVef e NVdf típico, A moderado, textura argilosa e muito argilosa; NXd típico, A moderado, textura argilosa e muito argilosa; PVAd PVAe PAD ou PAe arênico abrupto, A moderado e fraco, textura arenosa/média), como mostra a **Tabela 3**.

Apesar de haver mais classes nas quais a precisão foi maior na classificação com base em unidades elementares de relevo, estas classes somadas correspondem a 48,5 % da área de referência. Além disso, a precisão foi maior utilizando a classificação com base em pixels em duas classes com grande ocorrência na área de referência, os PVAd PVAe PAD ou PAe arênico abrupto, A moderado e fraco, textura arenosa/média (Serrinha – PVA3) que ocupam 27,9 % área de referência e os NXd típico, A moderado, textura argilosa e muito argilosa (Olaria – NX1) que ocupam 15,5 % da área de referência, conforme apresentado na **Tabela 3**, na coluna “Peso”.

Os PVAd PVAe PAD ou PAe arênico abrupto, A moderado e fraco, textura arenosa/média (Serrinha – PVA3) foram a Unidade de Solo mais predita por ambas abordagens de classificação. A segunda classe mais predita foi a dos NXd típico, A moderado, textura argilosa e muito argilosa (Olaria – NX1), ultrapassando os LVd típico, A moderado, textura argilosa e muito argilosa (Limeira – LV1) que tinham a segunda maior ocorrência na área de referência.

Ao todo, foram preditas 22 Unidades de Solo, ou seja, consociações de solos (Soil Science Division Staff, 2017) presentes nas bases de dados de treinamento construídas para a área de estudo. As 22 classes foram preditas utilizando tanto a abordagem de classificação supervisionada com base em pixels quanto por unidades elementares de relevo. Os mapas pedológicos digitais elaborados por classificação com base em pixels e com base em unidades elementares de relevo (GEOBIA) são apresentados na **Figura 2**.

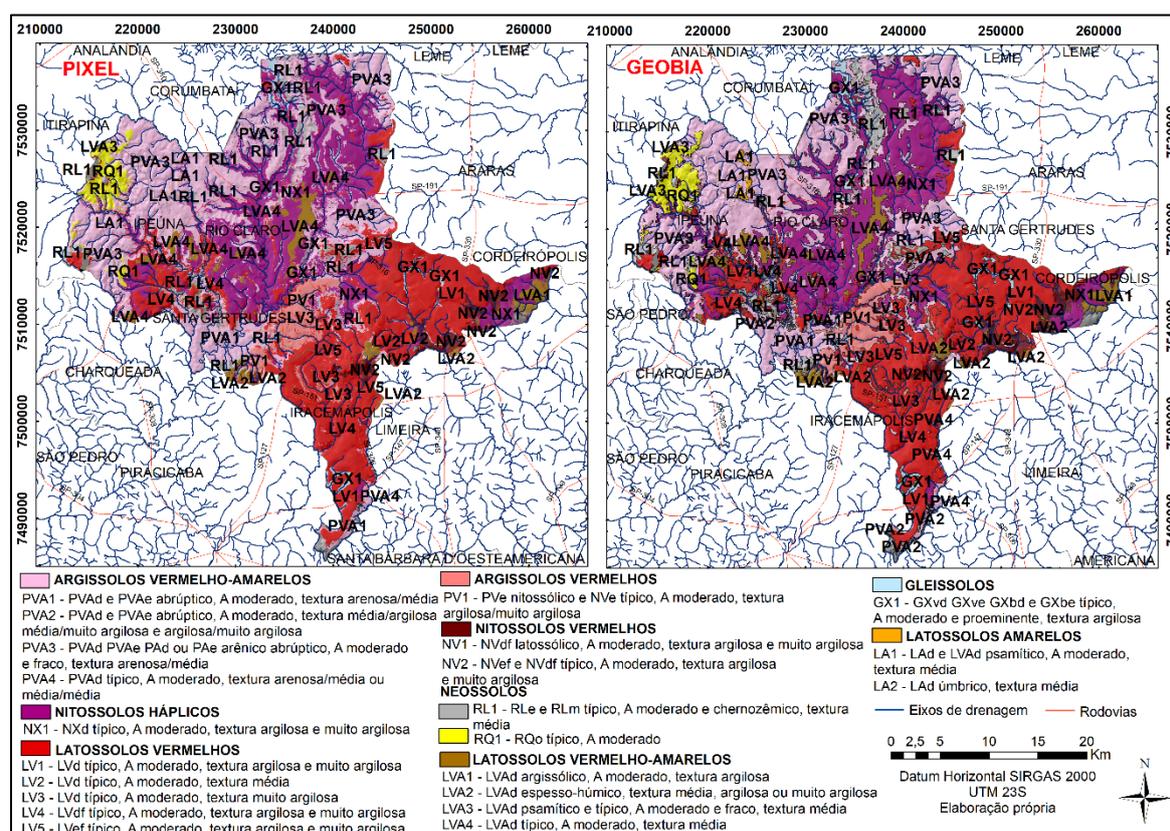


Figura 1: Mapas pedológicos digitais elaborados por classificação supervisionada com base em pixels (esquerda) e com base em unidades elementares de relevo (direita), por meio de classificação supervisionada utilizando o Random Forest.

O aspecto geral dos mapas para a classificação taxonômica de solos no 4º nível do SiBCS (Subordem) é, de maneira geral, bastante semelhante entre as abordagens por pixels e por unidades elementares de relevo. O mapa pedológico produzido pela classificação por pixels é mais fragmentado, quando visto com mais detalhe, apresentando transições entre delineamentos mais abruptas, como pode ser observado em algumas áreas da **Figura 2**. As transições entre os delineamentos apresentam-se mais uniformes e bem definidas no mapa pedológico produzido por classificação supervisionada com base em unidades elementares de relevo, como mostra a **Figura 2**.

4. Discussão

A qualidade dos mapas pedológicos gerados para a área de estudo aplicando as abordagens de classificação supervisionada com base em pixels e em unidades elementares de relevo (GEOBIA) foi compatível com outros trabalhos em MPD, avaliando os resultados entre 0,37 e 0,54 alcançados para o índice Kappa. Estes valores estão próximos ao valor médio de 0,48, calculado para os estudos em MPD realizados no Brasil (TEN CATEN *et al.*, 2012). Os resultados de acurácia global entre 45,9 % e 61,0 % são semelhantes aos encontrados por Criveleni *et al.* (2009) (60,8 %) e Silva *et al.* (2013) (52 %) que utilizaram unidades de mapeamento de levantamentos convencionais como fonte de informações sobre a ocorrência de classes de solo.

Neste estudo, a qualidade dos mapas pedológicos digitais elaborados por classificação supervisionada com base em pixels e com base em unidades elementares de relevo (GEOBIA), avaliada pela acurácia global e índice Kappa foi bastante semelhante. A literatura recente (DORNIK *et al.*, 2018; CREMON *et al.*, 2021) indica a classificação utilizando GEOBIA como uma técnica promissora para o mapeamento de classes de solos, superando as abordagens por pixels. Cremon *et al.* (2021) encontraram acurácia de 70 %, sendo 5 % maior que na abordagem por objetos, nos testes realizados. No trabalho de Dornik *et al.* (2018), que alcançaram acurácia global de 58 %, a vantagem foi de 10 % em relação à abordagem de classificação por pixels.

Foi importante investigar os resultados obtidos por Unidade de Solo em cada uma das abordagens, em busca das vantagens e desvantagens de cada uma. Nos LVef típico, A moderado, textura argilosa e muito argilosa (Ribeirão Preto – LV5) a taxa de verdadeiros positivos foi melhor na classificação com base em pixels. Os solos dessa consociação apresentam grande importância agronômica, devido à sua estrutura bem desenvolvida e, principalmente, à boa fertilidade natural, relativamente rara em ambientes tropicais úmidos (OLIVEIRA *et al.*, 1982; OLIVEIRA; PRADO, 1984).

A taxa de verdadeiros positivos também foi melhor na classificação com base em pixels nos NXd típico, A moderado, textura argilosa e muito argilosa (Olaria – NX1), que também apresentam grande importância agronômica e que merecem recomendações específicas de manejo do solo (OLIVEIRA *et al.*, 1982; OLIVEIRA; PRADO, 1984; BERTOL *et al.*, 2007) pensando a melhor conservação das características desejáveis, presentes naturalmente nestes solos.

Semelhante ao observado por Wolski *et al.* (2017), observa-se uma limitação na capacidade de extrapolação das Unidades de Solo, porém, essa limitação não parece estar associada diretamente à maior ou menor ocorrência na área de referência deste estudo. Os PVAd e PVAe abrupto, A moderado, textura arenosa/média (Alva – PVA1) que ocorrem em menos de 7 % da área de referência apresentaram bons resultados em todos os testes realizados. Por outro lado, os LVD típico, A moderado, textura argilosa e muito argilosa (Limeira – LV1) com mais de 20 % de ocorrência tiveram precisão baixa, em torno de 30 %. Devido à conceituação das unidades de solo Alva e Limeira e das unidades de solo a elas associadas (OLIVEIRA *et al.*, 1979; 1982; Oliveira, 1999), estes resultados sugerem forte influência da geologia para o desempenho obtido na predição destas duas classes de solo.

4. Conclusões

A rotina semiautomática para identificação das unidades elementares de relevo foi bastante eficaz para definir uma unidade cartográfica para o MPD, podendo ser utilizada na construção de bases de dados de treinamento em uma abordagem de classificação supervisionada utilizando GEOBIA, alcançando desempenho satisfatório. Desta forma, unidades elementares de relevo podem ser uma unidade de predição de classes de solo, assim como os pixels, com a vantagem de apresentar transições menos abruptas e delineamentos consistentes ao relevo.

As informações legadas dos levantamentos convencionais semidetalhados do estado de São Paulo em escala 1:100.000 podem ser utilizadas para produzir mapas pedológicos digitais com qualidade satisfatória

em área que equivale a, aproximadamente, 40 % de uma folha cartográfica em escala 1:100.000, com classificação em nível de Subordem do SiBCS, a partir de mapeamento geológico e geomorfológico com bom detalhamento, alcançando acurácia de 61 % e índice Kappa de 0,54.

O baixo desempenho na predição de Unidades de Solo não esteve diretamente relacionado à sua maior ou menor ocorrência na área de estudo, mas sim à associação específica de Unidades de Solo a determinadas características das covariáveis, sendo evidenciado tanto com a abordagem de classificação com base em pixels quanto naquela com base em unidades elementares de relevo. Ou seja, a qualidade e especificidade das relações entre as covariáveis preditivas em cada classe a ser predita parecem ser mais importantes do que a simples quantidade de observações existentes, sendo desejáveis estudos com maior aprofundamento nesses aspectos.

Referências

ACHANTA, R.; SÜSTRUNK, S. Superpixels and Polygons using Simple Non-Iterative Clustering. In: PROCEEDINGS OF THE IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION. **Anais [...]**. p. 4651-4660. 2017.

ADHIKARI K.; MINASNY B.; GREVE M.B. GREVE M.H. Constructing a soil class map of denmark based on the FAO legend using digital techniques. **Geoderma**, vol. 214-215, p. 101-113. 2014.

BERTOL, I.; LEITE, D.; ENGEL, F.L.; COGO, N.P.; GONZÁLEZ, A.P. 2007. Erodibilidade de um nitossolo háplico aluminico determinada em condições de campo. **Revista Brasileira de Ciência do Solo**, 31(3), p. 541-549.

BEVEN, K.J.; KIRKBY, M.J. A physically based, variable contributing area model of basin hydrology. **Journal Hydrological Sciences Bulletin**, vol. 24, p. 43-69, 1979.

BREIMAN, L. Random Forests. **Journal Machine Learning**, vol. 45, p. 5-32, 2001.

CABRAL JUNIOR, M.; GAMBA, T. DE C.; ALMEIDA, A. S.; FIAMMETTI, I. C. C.; TANNO, L. C.; MONTE, E. D.; MELONI, R. E. **Formulação do Plano Diretor de Mineração dos municípios de Cordeirópolis, Ipeúna, Iracemápolis, Rio Claro e Santa Gertrudes**. São Paulo (Rel. IPT n. 129345-205), 2012.

CAMERA C.; ZOMENI Z.; NOLLER J.S.; ZISSIMOS A.M.; CHRISTOFOROU I.C.; BRUGGEMAN A. A high resolution map of soil types and physical properties for Cyprus: a digital soil mapping optimization. **Geoderma**, vol. 85, p.35-49. 2017.

CARVALHO JUNIOR, W.D E; PEREIRA, N.R.; FERNANDES FILHO, E.I.; CALDERANO FILHO, B.; PINHEIRO, H.S.K; CHAGAS, C.DA.S.; BHERING, S.B.; PEREIRA, V.R.; LAWALL, S. Sample design effects on soil unit prediction with machine: randomness, uncertainty, and majority map. **Revista Brasileira de Ciência do Solo**, vol. 44, e0190120, 2020.

CHAGAS, C.S.; CARVALHO JUNIOR, W.; BHERING, S.B.; CALDERANO FILHO, B. Spatial prediction of soil surface texture in a semiarid region using random forest and multiple linear regressions. **Catena**, vol. 139, p. 232-240. 2016.

COELHO, R.M.; BERTOLANI, F.C.; ROSSI, M.; NASCIMENTO, P.C.; MENK, J.R.F; PRADO, H.; AMORIM, D.A. **Levantamento pedológico semidetalhado do Estado de São Paulo: folha de Marília**. Mapa escala 1:100.000. 2000.

COHEN, J. A Coefficient of Agreement for Nominal Scales. **Educational and Psychological Measurement**, vol. 20, New York, NY, p. 37-46, 1960.

COLANGELO, A.C. O Modelo de Feições Mínimas, ou das Unidades Elementares de Relevo: Um Suporte Cartográfico para Mapeamentos Geoecológicos. **Revista do Departamento de Geografia - USP** 10(1). São Paulo, p.29-40, 1996.

CREMON, É. H.; PEREIRA, A. C.; PAULA, L. D. L. D.; NUNES, E. D. Geological and terrain attributes for predicting soil classes using pixel- and geographic object-based image analysis in the Brazilian Cerrado. **Geoderma**, v. 401, p. 115315, nov. 2021. DOI 10.1016/j.geoderma.2021.115315. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0016706121003955>. Acesso em: 21 fev. 2023.

CRIVELENTI, R.C.; COELHO, R.M.; ADAMI, S.F.; OLIVEIRA, S.R. DE M. Mineração de dados para a inferência de relações solo-paisagem em mapeamentos digitais de solo. **Revista Agropecuária Brasileira**, v.44, n.12, p.1707-1715, 2009.

DORNIK, A.; DRĂGUȚ, L.; URDEA, P. Classification of Soil Types Using Geographic Object-Based Image Analysis and Random Forests. **Pedosphere**, v. 28, n. 6, p. 913–925, dez. 2018. DOI 10.1016/S1002-0160(17)60377-1. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1002016017603771>. Acesso em: 21 fev. 2023.

DRĂGUȚ, L.; BLASCHKE, T. Automated classification of landform elements using object-based image analysis. **Geomorphology**, v. 81, p. 330-344. 2006.

ESRI. Environmental Systems Research. **Introducing ArcGIS 10.1**. Arc News, Spring, 2012. Disponível em: <https://www.esri.com/news/arcnews/spring12articles/introducing-arcgis-101.html> Acesso: 18 de fevereiro de 2023.

GORELICK, N.; HANCHER, M.; DIXON, M.; ILYUSHCHENKO, S.; THAU, D.; MOORE, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. **Remote Sensing of Environment**, 202, p. 18-27. 2017.

HENGL, T.; HEUVELINK, G.B.M.; KEMPEN, B.; LEENAARS, J.G.B.; WALSH, M.G.; SHEPHERD, K.; SILA, A.; MACMILLAN, R.A.; JESUS, J.M.; TAMENE, L.; TONDOH, J.E. Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. **Plos One**, vol. 10, e0125814. 2015.

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

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

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**, v.49, p.958-966, 2014.

HUDSON, B. D. The soil survey as paradigm-based science. **Soil Science Society of American Journal**, v.56, p.836-841. 1992.

HUTCHINSON, M. F. Calculation of hydrologically sound digital elevation models. **Proceedings of the Third International Symposium on Spatial Data Handling**. p. 17-19. 1988.

IBGE. **Mapa Índice Digital, Mapeamento Geral do Brasil**. Instituto Brasileiro de Geografia e Estatística, Coordenação de Cartografia, 4ª ed., Rio de Janeiro, 30p. 2011.

IBGE. **Manual técnico de pedologia**. Instituto Brasileiro de Geografia e Estatística, Coordenação de Recursos Naturais e Estudos Ambientais, 3º ed., Rio de Janeiro, 430 p. 2015.

IBGE. **Mapeamento de Recursos Naturais do Brasil: Geomorfologia**. Instituto Brasileiro de Geografia e Estatística, Coordenação de Recursos Naturais e Estudos Ambientais. Mapa na escala 1:250.000. 2018.

JENNY, H. **Factors of soil formation, a system of quantitative pedology**. McGrawHill: New York, 1941. 281 p.

- JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis**. Madison: Prentice Hall International, 1998. 816p.
- LANDIS, J.R; KOCH, G.G. The Measurement of Observer Agreement for Categorical Data. **Biometrics**, vol. 33, p. 159-174, 1977.
- MACHADO, D.F.T; SILVA, S.H.G; CURTI, N.; MENEZES, M.D. Soil type spatial prediction from Random Forest: different training datasets, transferability, accuracy and uncertainty assessment. **Scientia Agricola**, vol. 76, n.3, p. 243-254. 2019.
- MCBRATNEY, A.B.; SANTOS, M.L.M.; MINASNY, B. On digital soil mapping. **Geoderma**, v.117, p. 3-52, 2003.
- OLIVEIRA, J.B. de; CAMARGO, M.N.de; ROSSI, M.; CALDERANO FILHO, B. **Mapa pedológico do Estado de São Paulo: legenda expandida**. Campinas: Instituto Agrônomo/EMBRAPA Solos, 1999. v. 1. 64 p. (inclui Mapa, escala 1:500.000).
- OLIVEIRA, J.B. de; MENK, J.R.F.; ROTTA, C.L. **Levantamento pedológico semidetalhado dos solos do Estado de São Paulo: folha de Campinas**. Mapa escala 1:100.000. Instituto Agrônomo - IAC, Campinas, 1977.
- OLIVEIRA, J.B.; BARBIERI, J.L.; ROTTA, C.L.; TREMOCOLDI, W. **Levantamento pedológico semidetalhado do Estado de São Paulo: Quadrícula de Araras**. Campinas, Instituto Agrônomo, 1982. 180p. (Boletim Técnico, 72 + mapa 1:100.000).
- OLIVEIRA, J.B.; PRADO, H. **Levantamento pedológico semidetalhado do Estado de São Paulo: quadrícula de São Carlos II**. Memorial descritivo. 1984.
- OLIVEIRA, J.B.; PRADO, H. **Levantamento pedológico semidetalhado do Estado de São Paulo: folha de Piracicaba**. Mapa escala 1:100.000. 1989.
- PÁSZTOR, L.; LABORCZI, A.; BAKACSI, Z.; SZABO, J.; ILLES, G. Compilation of a national soil-type map for Hungary by sequential classification methods. **Geoderma**, vol. 311:p. 93-108. 2018.
- Perrota, M.M.; Salvador, E.D.; Lopes, R.C. **Mapa Geológico do estado de São Paulo, escala 1:750.000**. CPRM, Programa Geologia do Brasil – PGB, São Paulo, 2006.
- PRADO, H. OLIVEIRA, J.B.; ALMEIDA, C.L.F. **Levantamento pedológico semidetalhado do Estado de São Paulo: Quadrícula de São Carlos**. Campinas, EMBRAPA/CPA/IA, 1981. (Mapa, escala 1:100.000)
- ROSS, J. L. S. 1992. O registro cartográfico dos fatos geomorfológicos e a questão da taxonomia do relevo. **Revista do Departamento de Geografia**, 6, 17-29.
- ROSS, J.L.S.; MOROZ, I.C. **Mapa Geomorfológico do Estado de São Paulo**. Laboratório de Geomorfologia. Departamento de Geografia – FFLCH –USP/Laboratório de Cartografia Geotécnica – Geologia Aplicada – IPT/FAPESP (Mapas e Relatórios). São Paulo, SP, 1997.
- ROSSI, M. 2017. **Mapa pedológico do Estado de São Paulo: revisado e ampliado**. São Paulo: Instituto Florestal, 2017. V.1. 118p. (inclui Mapas)
- RUDIYANTO, R.; MINASNY, B.; SETIAWAN, B.I.; ARIF, C.; SAPTOMO, S.K.; CHADIRIN, Y. Digital mapping for cost-effective and accurate prediction of the depth and carbon stocks in Indonesian peatlands. **Geoderma**, vol. 272, p. 20-31. 2016.
- SANTOS, H.G.; JACOMINE, P.K.T.; ANJOS, L.H.C.; OLIVEIRA, V.A.; LUMBRERAS, J.F.; COELHO; M.R. **Sistema Brasileiro de Classificação de Solos**. Embrapa, Brasília, DF, 356 p. 2018.

SARTORI, M.; PHILIPPIDIS, G.; FERRARI, E.; BORRELLI, P.; LUGATO, E.; MONTANARELLA, L.; PANAGOS, P. A linkage between the biophysical and the economic: Assessing the global market impacts of soil erosion. **Land Use Policy**, vol. 86, p. 299-312, 2019.

SILVA, C.C.; COELHO, R.M.; OLIVEIRA, S.R.M. Digital pedological mapping of Botucatu sheet (SF-22-ZB-VI-3): data training on conventional maps and field validation. **Revista Brasileira de Ciência do Solo**, vol. 37, n. 4, p. 846-857, 2013.

SOIL SCIENCE DIVISION STAFF. **Soil survey manual**. United States Department of Agriculture (USDA), Handbook 18. Washington, DC, 638 p. 2017.

SOUZA, E.; FERNANDES FILHO, E.I.; SCHAEFER, C.E.G.R.; BATJES, N.H.; SANTOS, G.R.; PONTES, L.M.. Pedotransfer functions to estimate bulk density from soil properties and environmental covariates: Rio Doce basin. **Scientia Agricola**, 73: 525-534, 2016

TEN CATEN, A.; DALMOLIN, R.S.D.; SANTOS, M.L.M.; GIASSON, E. Mapeamento digital de classes de solo: características da abordagem brasileira. **Ciência Rural**, v.42, p. 1989-1997, 2012.

TESKE, R.; GIASSON, E.; BAGATINI, T. Comparação de esquemas de amostragem para treinamento de modelos preditores no mapeamento digital de classes de solos. **Rev. Bras. Cienc. Solo**. vol. 39, p. 14-20. 2015.

VALADARES, A.P; COELHO, R.M.; OLIVEIRA, S.R.M. Preprocessing procedures and supervised classification applied to a database of systematic soil survey. **Scientia Agricola**, v. 76(5), p.439-447. 2019.

WEISS, S. M.; ZHANG, T. Performance analysis and evaluation. In: Ye, N. **The handbook of Data Mining**. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, vol. 14, p. 425 – 440, 2003.

WITTEN, I.H.; FRANK, E.; HALL, M.A; PAL, C.J. **Data mining: practical machine learning tools and techniques**. 4° ed. Morgan Kaufmann, Burlington, MA, 629 p. 2016.

WOLSKI, M.S.; DALMOLIN, R.S.D.; FLORES, C.A.; MOURA-BUENO, J.M.; TEN CATEN, A.; KAISER, D.R. Digital Soil Mapping and its implications in the extrapolation of soil-landscape relationship in detailed scale. **Pesquisa Agropecuária Brasileira**, v. 52, p. 633-642, 2017.



Este artigo é distribuído nos termos e condições do *Creative Commons Attributions/Atribuição-NãoComercial-CompartilhaIgual* (CC BY-NC-SA).