



MARIA EDUARDA CARVALHO MONTEIRO

**MAPEAMENTO DIGITAL DE SOLOS DA REGIÃO SUL DE
MG USANDO NOVAS COVARIÁVEIS AMBIENTAIS
PREDITORAS**

LAVRAS – MG

2021

MARIA EDUARDA CARVALHO MONTEIRO

**MAPEAMENTO DIGITAL DE SOLOS DA REGIÃO SUL DE MG USANDO NOVAS
COVARIÁVEIS AMBIENTAIS PREDITORAS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Ciência do Solo, área de concentração em Recursos Ambientais e Uso da Terra, para a obtenção do título de Mestre.

Profa. Dra. Michele Duarte de Menezes
Orientadora

Prof. Dr. Nilton Curi e Prof. e Prof. Dr. Fausto Weimar Acerbi Júnior
Coorientadores

LAVRAS - MG

2021

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Monteiro, Maria Eduarda Carvalho.

Mapeamento digital de solos da região Sul de MG usando
novas covariáveis ambientais preditoras / Maria Eduarda Carvalho
Monteiro. - 2021.

77 p. : il.

Orientador(a): Michele Duarte de Menezes.

Coorientador(a): Nilton Curi, Fausto Weimar Acerbi Júnior.

Dissertação (mestrado acadêmico) - Universidade Federal de
Lavras, 2021.

Bibliografia.

1. Pedometria. 2. Mapeamento digital de solos. 3. Modelagem.
I. Menezes, Michele Duarte de. II. Curi, Nilton. III. Acerbi Júnior,
Fausto Weimar. IV. Título.

MARIA EDUARDA CARVALHO MONTEIRO

**MAPEAMENTO DIGITAL DE SOLOS DA REGIÃO SUL DE MG USANDO NOVAS
COVARIÁVEIS AMBIENTAIS PREDITORAS**

**DIGITAL MAPPING OF SOILS IN THE SOUTH OF MINAS GERAIS USING NEW
PREDICTIVE ENVIRONMENTAL COVARIABLES**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Ciência do Solo, área de concentração em Recursos Ambientais e Uso da Terra, para a obtenção do título de Mestre.

APROVADA em 30 de agosto de 2021.

Dra. Michele Duarte de Menezes - UFLA

Dr. Sérgio Henrique Godinho Silva - UFLA

Dr. Élvio Giasson - UFRGS

Prof. Dra. Michele Duarte de Menezes

Orientadora

Prof. Dr. Nilton Curi e Prof. e Prof. Dr. Fausto Weimar Acerbi Júnior

Coorientadores

LAVRAS – MG

2021

Aos que acreditam na ciência, na tecnologia e na educação
como elementos fundamentais para a constituição de uma sociedade livre e democrática.

Dedico.

AGRADECIMENTOS

À Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG) pelo financiamento da bolsa de mestrado.

À CAPES e CNPq pelo apoio para a realização deste trabalho.

À UFLA, universidade PÚBLICA, GRATUITA e de QUALIDADE.

Ao Departamento de Ciência do Solo pelo apoio técnico e ambiente de contínua troca de saberes.

Ao Professor Nilton Curi, pela confiança depositada e oportunidades oferecidas ao meu crescimento acadêmico.

À minha orientadora, professora Michele Duarte de Menezes, foi um privilégio encarar esse desafio com sua ajuda.

Ao Professor Fausto Weimar Acerbi Júnior, pela atenção oferecida ao desenvolvimento do projeto e por disponibilizar a infraestrutura do DCF para o processamento de dados.

Aos pesquisadores do DCS-UFLA que gentilmente ofereceram seus dados para esta pesquisa: Raísa Brito Vilela, Fábio Arnaldo Pomar Avalos, Bárbara Pereira Christofaro Silva, Mariana Gabriele Marcolino Gonçalves, Luana Sousa Costa.

À Dirce, sempre tão atenciosa.

Aos colegas do Departamento de Solos, pelo apoio e companhia durante esta caminhada, especialmente Marcelo Pelegrino.

Aos colegas do Departamento de Ciências Florestais, Inácio, Maria Cleidivane e July pela ajuda durante os primeiros passos do trabalho.

Ao Tom, pelo amor, amizade e companheirismo, sempre.

Agradeço.

RESUMO

O planejamento adequado do uso do solo é imprescindível para atender aos desafios globais impostos pela demanda de alimentos e mitigação das mudanças climáticas. Para superar esses desafios, os mapas de solos são ferramentas fundamentais à interpretação da paisagem e adequação das atividades produtivas. O mapeamento digital de solos (MDS) tem sido empregado para acelerar a produção de mapas de áreas extensas, em escalas mais detalhadas a um menor custo. O aprimoramento das técnicas de aprendizado de máquinas e a experimentação de novos produtos digitais consistem em demandas urgentes ao desenvolvimento do MDS com os quais esse trabalho buscou contribuir. Informações construídas durante décadas de pesquisa pedológica foram reunidas para orientar a produção de um mapa de solos a partir de uma abordagem quantitativa. O legado das informações foi aplicado à modelagem preditiva de classes de solos com uso do algoritmo *Random Forest*. Dados de aeroespectrometria de raios gama e modelos paleoclimáticos, com aplicações ainda inéditas em MDS no Brasil foram empregados para o mapeamento de 52.982 km², juntamente com variáveis ambientais derivadas do Modelo Digital de Elevação (MDE) e de modelos do clima atual. O nível de extrapolação do modelo foi calculado a partir da Superfície de Similaridade Ambiental Multivariada e a Entropia da predição foi calculada segundo a fórmula da Entropia de Shannon. A acurácia global da predição foi de 89% para o mapeamento de 19 classes de solo em nível de subordem segundo o Sistema Brasileiro de Classificação de Solos. As covariáveis ambientais mais importantes na predição espacial foram os dados de aeroespectrometria de raios gama, o modelo paleoclimático da precipitação total anual estimada há 20.000 anos e a distância vertical da rede de drenagem derivada do MDE.

Palavras-chave: Pedometria. Espectrometria de raios gama. Paleoclima.

ABSTRACT

Proper land use planning is essential to meet the global challenges posed by food demand and climate change mitigation. To overcome these challenges, soil maps are fundamental tools for landscape interpretation and adaptation of productive activities. Digital soil mapping (DSM) has been used to accelerate the production of maps of large areas, at more detailed scales with lower costs. The improvement of machine learning techniques and the experimentation of new digital products are urgent demands for the development of the DSM, in which this work sought to contribute. Information built up over decades of pedological research was gathered to guide the production of a soil map from a quantitative approach. The legacy information was applied to predictive modeling of soil classes using the Random Forest algorithm. Gamma-ray aerospectrometry data and paleoclimatic models, whose application is unprecedented in the MDS in Brazil, were applied for the mapping of 52,982 km², together with environmental variables derived from the Digital Elevation Model (DEM) and current climate models. The extrapolation level of the model was calculated from the Multivariate Environmental Similarity Surface, and the prediction Entropy was calculated according to the Shannon Entropy formula. The overall accuracy of the prediction was 89% for the mapping of 19 soil classes at suborder level according to the Brazilian Soil Classification System. The most important environmental covariables in the spatial prediction were the gamma-ray aerospectrometry data, the paleoclimatic model of the total annual precipitation estimated 20,000 years ago, and the vertical distance from the drainage network derived from the DEM.

Keywords: Pedometry. Airborne gamma ray spectrometry. Paleoclimate.

SUMÁRIO

PRIMEIRA PARTE	8
1 INTRODUÇÃO	8
2 REFERENCIAL TEÓRICO	10
2.1 Pedologia e mapeamento de solos	10
2.1.1 Mapeamento de solos no Brasil	11
2.1.2 Mapeamento digital de solos	12
2.1.3 Covariáveis ambientais	13
2.2 Geomorfologia e solos da região Sul de Minas Gerais	16
3 CONSIDERAÇÕES GERAIS	19
REFERÊNCIAS	20
SEGUNDA PARTE	33
ARTIGO: MAPEAMENTO DIGITAL DE SOLOS NO SUL DE MINAS GERAIS: PREDIÇÃO, INCERTEZA E IMPORTÂNCIA DAS COVARIÁVEIS AMBIENTAIS	33
1 INTRODUÇÃO	34
2 MATERIAL E MÉTODOS	37
2.1 Área de estudo	37
2.2 Conjunto de dados	38
2.2.1 Estratégia para lidar com a base de dados desbalanceada	39
2.3 Mapeamento Digital de Solos (MDS)	41
2.3.1 Covariáveis ambientais SCORPAN	41
2.3.2 Predição espacial de solos e validação	44
2.3.3 Parametrização do algoritmo <i>Random Forest</i> e seleção de covariáveis ambientais 45	
2.4 Quantificação da incerteza da predição espacial	46
2.4.1 Avaliação da extrapolação espacial a partir da Superfície de Similaridade Ambiental Multivariada (SSAM)	46
2.4.2 Entropia	47
3 Resultados e discussão	49
3.1 Parametrização do algoritmo <i>Random Forest</i>	49
3.2 Acurácia dos modelos preditivos e seleção de covariáveis ambientais	49
3.3 Avaliação da extrapolação espacial (SSAM)	50

3.4	Entropia.....	52
3.5	Probabilidade de ocorrência.....	54
3.6	Predição espacial de solos	57
3.7	Importância das covariáveis ambientais	60
4	Conclusões.....	64

PRIMEIRA PARTE

1 INTRODUÇÃO

Numerosos estudos têm defendido que o solo seja uma das prioridades da agenda da política ambiental (HARTEMINK; McBRATNEY, 2008; FAO, 2015; HARTEMINK, 2015). Essa postura condiz com a importância central do solo frente aos principais desafios globais como segurança alimentar e hídrica, mitigação e adaptação às mudanças climáticas e proteção da biodiversidade (McBRATNEY; FIELD; KOCH, 2014).

O desenvolvimento de soluções e a construção de cenários requerem informações e conhecimento das propriedades e funções do solo (GAN-LIN et al., 2017). Nesse aspecto, o levantamento de solos é a ferramenta tradicional mais utilizada para se obter informações a respeito de uma área. Através dele é possível fazer a estratificação de ambientes em nível local, visando a caracterizar o solo em relação a sua posição na paisagem frente aos seus diferentes potenciais de uso (RESENDE et al., 2014).

Os mapas de solos, tradicionalmente, consistem em um modelo de representação da distribuição espacial dos solos construído a partir de extensos trabalhos de campo e interpretação das relações solo-paisagem por pedólogos que possuem conhecimento empírico suficiente para espacializar a distribuição dos solos na paisagem (MACHADO et al., 2019a). A materialização desse conhecimento em forma de mapas permite apresentar informações de forma acessível para os agricultores e gestores públicos, disponibilizando conhecimento técnico especializado para diferentes usos e interpretações.

Em nível regional, o adequado planejamento do uso do solo demanda informações detalhadas e para grandes extensões de terras. Como a execução de levantamentos de solos de forma convencional demanda mais tempo, existe a necessidade da aplicação de novas abordagens que atendam, mais rapidamente a elaboração de mapas de solos em escala adequada aos diferentes usos e para áreas mais extensas (MENDONÇA-SANTOS; SANTOS, 2007; MENEZES et al., 2013). O reconhecimento dessa demanda incentivou o início das pesquisas que tem testado técnicas de MDS. Recentemente, no Brasil iniciou-se o Programa Nacional de Solos (PronaSolos), instituído pelo Decreto de Lei nº 9414, de 19 de junho de 2018, cujo principal objetivo é realizar, de forma sistemática, o mapeamento de solos em escalas compatíveis para a gestão de uso das terras, em níveis mais detalhados que os existentes, de todo o território brasileiro utilizando técnicas de MDS.

O MDS é uma releitura da função clássica de formação do solo baseada na interpretação da paisagem, concebida por Dokuchaev e formulada por Jenny (1941), segundo a qual as características do solo são resultado da interação entre os fatores de formação: clima, organismos, relevo e material de origem interagindo ao longo do tempo.

Nessa nova abordagem, o MDS usa dados existentes sobre os solos, sua posição geográfica e variáveis ambientais representadas por dados digitais, que assumem o papel dos fatores de formação dos solos, são as covariáveis ambientais (MCBRATNEY; MENDONÇA SANTOS; MINASNY, 2003). O conhecimento tácito do pedólogo é aproveitado na forma de exemplos reais, classificações de perfis de solos, que orientam o aprendizado de máquinas e a construção de modelos preditivos para a ocorrência de classes de solos em toda uma paisagem (BUI, 2004).

Para a implementação prática em larga escala do MDS existem diversos desafios, dentre eles, a disponibilidade de dados digitais com resolução mais detalhada (MINASNY; MCBRATNEY, 2016), o estabelecimento de relações entre esses dados e a distribuição espacial dos solos (ARROUAYS et al., 2020b) e a escolha da metodologia mais adequada para cada paisagem.

Nesse sentido, este trabalho propõe a utilização de dados legados e novas covariáveis para obter um mapa preditivo de classes de solo para a região Sul de Minas Gerais, uma das dez regiões de planejamento estabelecidas oficialmente pelo governo do Estado Minas Gerais, que congrega 155 municípios em uma extensão de 52.982 km².

Além das questões metodológicas, o produto gerado tem uma aplicação prática para a gestão dos serviços ecossistêmicos, pois os recursos hídricos existentes na região abastecem grandes reservatórios para geração de energia elétrica, dentre eles, o reservatório das usinas hidrelétricas de Furnas e do Funil, responsáveis pela geração de parte da energia consumida na região sudeste. O prolongamento da vida útil dos reservatórios está relacionado ao manejo adequado dos solos nas áreas de captação da bacia. A poluição difusa gerada pela agropecuária, a erosão e o assoreamento da rede de drenagem são alguns dos problemas relatados no plano integrado de recursos hídricos da bacia (ANA, 2017), com os quais o mapeamento de solos pode contribuir para orientar ações e programas direcionados à conservação do solo e da água, embasando o planejamento estratégico, auxiliando gestores ambientais e produtores rurais nas decisões para o melhor manejo e uso do solo na região Sul de Minas Gerais.

2 REFERENCIAL TEÓRICO

2.1 Pedologia e mapeamento de solos

O reconhecimento do solo como corpo natural organizado, objeto de estudo da pedologia é recente, tendo ocorrido na segunda metade do século XIX. Avanços em química, mineralogia e a teoria evolucionista de Darwin influenciaram o desenvolvimento dessa nova concepção (KÄMPF; CURI, 2012a). Os avanços no conhecimento sobre o solo desta época são registrados nos trabalhos de Dokuchaev e colaboradores quando, estimulados pelo interesse no desenvolvimento da agricultura, desenvolveram um modelo para explicar as relações entre os fatores ambientais e a distribuição geográfica dos diferentes tipos de solos. O reconhecimento de que cinco fatores: material de origem, clima, vegetação, relevo e tempo, influem na formação do solo. Teve como consequência uma orientação ao processo de mapeamento de solos (LEGROS, 2006).

No início do século XX, a concepção determinística de Dokuchaev foi adaptada para a noção de probabilidade, pela qual certa combinação de fatores tem como resultado um solo mais provável. Nesta época, a teoria dos fatores de formação do solo da escola russa foi traduzida na equação dos fatores de formação de solo, $S = f(c, o, r, p, t)$, S=solo; c=clima, o=organismos, r=relevo, p=material de origem, t=tempo, divulgada mundialmente (JENNY, 1941).

O paradigma de Dokuchaev e a equação de Jenny são amplamente aplicados nos estudos de gênese e mapeamento de solos, pois o pedólogo, para classificar tipos de solos e determinar limites de unidades de mapeamento, se utiliza de um modelo mental intuitivo que assume a existência de forte correlação entre os tipos de solos e o ambiente onde eles se encontram (MENDONÇA-SANTOS; SANTOS, 2003). Trata-se de uma abordagem qualitativa, com base na análise da fisiografia da paisagem, na interpretação de fotografias aéreas, formulação de hipóteses e verificações em campo e em laboratório (LEGROS, 2006).

A metodologia tradicional de mapeamento de solos tem como estratégia o entendimento dos conceitos dos fatores de formação do solo juntamente com as relações solo-paisagem (HUDSON, 1992). Nesse sentido, os mapas de solos são expressão gráfica do conhecimento estruturado do pedólogo (BUI, 2004), que descreve características dos solos de uma determinada área, classifica-os de acordo com um sistema taxonômico, estabelece limites entre

unidades de mapeamento de solos e permite fazer inferências sobre o comportamento dos solos quanto ao uso e ao manejo (MENDONÇA-SANTOS; SANTOS, 2003).

2.1.1 Mapeamento de solos no Brasil

Os mapas de solos são instrumentos estratégicos aos poderes públicos para planejar e gerenciar os recursos naturais da nação (CARVALHO; NUNES; ANTUNES, 2013). Constituem uma ferramenta importante para o planejamento do uso das terras, possibilitando a identificação de diversas classes de solos e seus atributos (químicos, mineralógicos e físicos), gerando informações capazes de auxiliar na definição do uso mais adequado do solo.

O primeiro mapa de solos publicado no Brasil foi o “Esboço Agrogeológico do Estado de São Paulo”, publicado em 1943, com o reconhecimento de 22 tipos de solos no estado, com informações sobre propriedades físicas e químicas (KER; MOTTA; OLIVEIRA, 2017). Estudos com abrangência nacional iniciaram com a criação da Comissão de Solos, do Ministério da Agricultura, em 1947, com a padronização de métodos e terminologias. Mas, apenas na década de 1980, foi publicado o primeiro mapa de solos do Brasil, em escala de 1:5.000.000 (KER; MOTTA; OLIVEIRA, 2017). Atualmente, quase quatro décadas após a publicação do primeiro mapa de solos para o território brasileiro, esta ainda é a escala de informação disponível em abrangência nacional, observando apenas atualizações para o Sistema Brasileiro de Classificação de Solos (SANTOS et al., 2013). Para o estado de Minas Gerais, em 1982 foi publicado um mapa de solos na escala 1:1.000.000 (AMARAL et al., 2004) e, apenas em 2010, a escala do mapa de solos foi melhorada para todo o estado, com a publicação de um mapa com escala de 1:650.000 (CETEC et al., 2010)

São poucos levantamentos de solos realizados em escalas de maior detalhe, que forneçam informações suficientes para uso e direcionamento de atividades em propriedades agrícolas, ou mesmo em âmbito municipal (EMBRAPA, 2016). Levantamentos de solo ou esquemas de amostragem em escala mais detalhada são comuns apenas em pequenas áreas, geralmente para atender projetos específicos (MENDONÇA-SANTOS; SANTOS, 2007; MENEZES et al., 2013).

Apesar da utilidade inquestionável de mapas detalhados de solos, o levantamento convencional de solos, principal método utilizado no país, demanda longos prazos e altas cifras de recursos financeiros. Depende de cartas topográficas e fotos aéreas em escala adequada, de uma boa infraestrutura de rodovias para as prospecções de campo, além de recursos humanos

especializados, pessoal com elevado conhecimento técnico das relações solo-paisagem. Essas exigências, associadas à falta de investimentos em pesquisa, comum em países em desenvolvimento, têm contribuído para a situação de escassez de informações mais detalhadas sobre os solos (McBRATNEY; MENDONÇA SANTOS; MINASNY, 2003; SANTOS et al., 2013; EMBRAPA, 2016).

O Programa Nacional de Solos (PronaSolos), delineado em 2015, e instituído pelo Decreto de Lei nº 9414, em 19 de junho de 2018, tem como objetivo aumentar o nível de conhecimento sobre os solos brasileiros, porém, as dificuldades de execução dos levantamentos convencionais podem ser confirmadas pela estimativa do prazo de até 30 anos para mapear 80% do território brasileiro na escala 1:100.000, 12% na escala 1:50.000 e 3% na escala 1:25.000 (EMBRAPA, 2016).

O cenário atual e as perspectivas futuras para disponibilização de informações sobre os solos do país, mesmo contando que todo o investimento previsto para o programa PronaSolos seja realmente aplicado, indica a necessidade de novas abordagens que atendam, em menor tempo, as demandas atuais dessas informações (MENDONÇA-SANTOS; SANTOS, 2007; MENEZES et al., 2013; FLACH; CORRÊA, 2018). Uma alternativa que se mostra promissora para impulsionar o levantamento pedológico é o Mapeamento Digital de Solos (MDS), que se propõe como método capaz de fornecer mapas de classes ou atributos de solos em maior escala de detalhes, para áreas mais extensas e em menor tempo de execução (McBRATNEY; MENDONÇA SANTOS; MINASNY, 2003).

2.1.2 Mapeamento digital de solos

O MDS foi formalizado no modelo SCORPAN, onde $\text{solo} = f(\text{solo}, \text{clima}, \text{organismos}, \text{relevo}, \text{material de origem}, \text{idade}, \text{localização})$, por meio do qual a ocorrência de diferentes tipos de solo pode ser prevista utilizando covariáveis ambientais preditoras que representam os fatores de formação dos solos, acrescentando informações disponíveis sobre os solos e/ou mapas preexistentes e a localização geográfica do solo que se pretende mapear. As vantagens da utilização deste modelo são a redução da subjetividade da predição e a possibilidade de reportar a incerteza das previsões através das medidas de acurácia (McBRATNEY; MENDONÇA SANTOS; MINASNY, 2003, MENEZES et al., 2013). A utilização deste modelo permite a abordagem estatística da ocorrência de tipos de solos e a variação de seus atributos de solos.

A metodologia padrão para estabelecer estas funções é tomar observações no campo em locais com coordenadas conhecidas e relacioná-las com um conjunto de variáveis preditoras pedologicamente significativas por meio de um modelo (MINASNY; McBRATNEY, 2016). Uma vez que o modelo é treinado para os pontos observados, as previsões podem ser extrapoladas para toda extensão da área representada pela amostragem inicial, oferecendo um mapa digital dos solos e/ou de variação dos atributos. A eficiência do método está no fato de que a variação dos tipos de solos e atributos de solo é muito mais difícil e caro de representar do que as variáveis capazes de prevê-los (LAGACHERIE; McBRATNEY, 2007).

A utilização do modelo SCORPAN no MDS permite descrever, classificar e estudar os padrões de distribuição dos solos na paisagem por meio da aplicação de métodos matemáticos e estatísticos (métodos preditivos ou de predição espacial) para a resolução de incertezas e complexidades inerentes ao sistema solo. Essas aplicações constituem o ramo da Ciência do Solo chamado Pedometria, que se utiliza das abordagens numéricas para o mapeamento de solos, considerando a existência de uma variação determinística (McBRATNEY et al., 2000).

2.1.3 Covariáveis ambientais

As variáveis ambientais representam a influência dos fatores de formação (JENNY, 1941) na ocorrência de diferentes tipos de solos e, portanto, podem ser usadas para a predição na área a ser mapeada. Diversos dados podem servir como covariáveis para a modelagem preditiva de solos: mapas geológicos, modelos climáticos, domínios fitogeográficos, imagens de satélite, entre outros. Esses dados, quando trabalhados em Sistemas de Informações Geográficas (SIG) tornam-se camadas de dados que permitem os modelos matemáticos identificar condições semelhantes para a predição de cada tipo de solo (HENGL, 2006). Os dados que irão compor o conjunto de variáveis preditoras dependem da disponibilidade de informações para a área de interesse, das variáveis necessárias para caracterizar cada fator de formação do solo e da importância de cada variável no mapeamento de classe ou atributo de solo na paisagem (McBRATNEY; MENDONÇA SANTOS; MINASNY, 2003).

Existem diversas variáveis que podem representar cada um dos fatores de formação do solo do modelo SCORPAN. O fator solo, diz respeito às informações de tipos ou atributos de solo capazes de treinar modelos de predição. Um exemplo onde esse fator configurou uma das variáveis mais importantes para a predição da classe de solo foi na Dinamarca (ADHIKARI et

al., 2014), onde o teor de argila, previamente interpolado espacialmente, foi utilizado como covariável ambiental.

O fator clima pode ser representado, por exemplo, por superfícies climáticas produzidas pela interpolação de dados de estações climáticas ou por variáveis derivadas do MDE. Valores de temperatura mínima, máxima, precipitação, incidência da radiação solar, altitude, evapotranspiração potencial, são exemplos de variáveis ambientais representativas do clima (COELHO et al., 2021). As relações entre as variáveis climáticas e a ocorrência de determinados tipos de solos é relatada com frequência na literatura, desde os primeiros trabalhos de pedologia. A relação precipitação/evapotranspiração, por exemplo, ofereceu subsídios para separação de solos ricos em ferro e solos ricos em cálcio (Pedalfers e Pedocals) da antiga classificação de solos norte americana (SCHAETZL; ANDERSON, 2005). Este trabalho foi registrado em 1935 nos Estados Unidos (SHAW, 2010). Essa mesma relação é capaz de indicar climas semiáridos e possibilidade de precipitação de carbonatos em profundidade (McBRATNEY; MENDONÇA SANTOS; MINASNY, 2003).

Atualmente, a utilização das variáveis climáticas para inferências sobre os solos conta com dados referentes à temperatura e precipitação para o período de 1970 até o ano 2000, disponíveis para o globo terrestre, com resolução de 1 km² na plataforma *Worldclim – World Climate Data* (FICK; HIJMANS, 2017), fornecendo uma base de informações importante para a modelagem de atributos e tipos de solos. Além da interpolação dos dados sobre o clima atual, estão disponíveis nesta mesma plataforma modelos climáticos para o passado, simulando condições há 120.000 anos, 22.000 anos e 6.000 anos (TAYLOR; STOUFFER; MEEHL, 2012). Nestes dados, para cada período, são apresentados produtos desenvolvidos por diferentes instituições de pesquisa que, utilizando submodelos de representação do comportamento das correntes oceânicas, da concentração de gases da atmosfera e da dinâmica das placas glaciais, obtiveram estimativas sobre o clima em épocas passadas. Dados paleoclimáticos têm potencial de utilização principalmente em solos tropicais, uma vez que, muitas das paisagens podem não estar em equilíbrio com as condições climáticas atuais (WILFORD, 2012). E ainda, segundo Wilford (2012), a análise das paisagens deve levar em consideração o tempo de exposição aos agentes do intemperismo, uma vez que tais situações são utilizadas para inferir sobre as taxas de erosão e acumulação. Deste modo, sendo o processo de formação do solo multivariado, as condições atuais e pretéritas, uma vez atuantes, devem ser consideradas como potenciais covariáveis ambientais.

Outra variável que apresenta relação explícita do clima com o solo é a vegetação nativa, demonstrada em diversos trabalhos (HARIDASAN, 1982; FELFILI; DA SILVA, 1993; SKORUPA et al., 2012). O clima tem forte influência na distribuição dos domínios. As estações seca e chuvosa bem definidas favorecem as fitofisionomias com características de savana, já climas mais equilibrados e úmidos favorecem formações florestais (RIBEIRO; WALTER, 2008). Considerando que o clima apresenta uma variação mais suave que outros fatores de formação, nas zonas de transição, os solos são ao menos parcial e localmente responsáveis pela variação dos tipos vegetacionais, determinando ambientes preferenciais de espécies (COELHO; ROSSI; MATTOS, 2017). A disponibilidade de nutrientes, o teor de alumínio trocável, a drenagem são características que influenciam a variação das fitofisionomias (MENEZES et al., 2017) e o acúmulo de carbono no subsolo. Dados contínuos sobre o estoque de carbono dos solos são provenientes de modelagens preditivas e encontram-se disponíveis para o Brasil a partir dos trabalhos coordenados pela Embrapa Solos (VASQUES et al., 2017) e Universidade Federal de Viçosa (GOMES et al., 2019).

Os modelos digitais de elevação (MDE), são os dados mais utilizados para mapeamento preditivo (GRIMM et al., 2008). Os modelos podem ser utilizados em diferentes resoluções, derivados a partir de imagens de satélite ou de cartas topográficas. O MDE serve de base para o processamento de atributos do terreno como altitude, declividade, plano de curvatura, índices de umidade (ten CATEN et al., 2012). Muitos trabalhos aplicaram esses atributos para classificação de solos, considerando o relevo como principal direcionador da variação (JAFARI et al., 2014; VAYSSE; LAGACHERIE, 2015; SILVA et al., 2016).

A utilização da relação entre a classe de solo e o fator material de origem no mapeamento de solos está tradicionalmente ligada à escola russa inaugurada por Dokuchaev (BREVIK et al., 2016). A fonte de dados mais utilizada nesse caso são os mapas geológicos em formato de polígonos e georreferenciados, geralmente disponíveis em escalas menos detalhadas que a demanda das informações sobre os solos, o que implica em uma dificuldade de sua utilização no MDS. Além disso, a característica categórica dos dados inviabiliza algumas abordagens metodológicas de mapeamento. Outro tipos de dado que tem sido utilizado como alternativa ao mapas geológicos são os dados de espectrometria de raios gama captados em levantamentos aerogeofísicos. Eles quantificam o fluxo de radiação gama naturalmente emitido pelos solos e rochas até uma profundidade de aproximadamente 40 cm. Essa radiação resulta do decaimento radioativo de K, U e Th e é capaz de fornecer informações quantitativas sobre o material de origem dos solos e a intensidade do intemperismo (WILFORD, 2012). O potencial

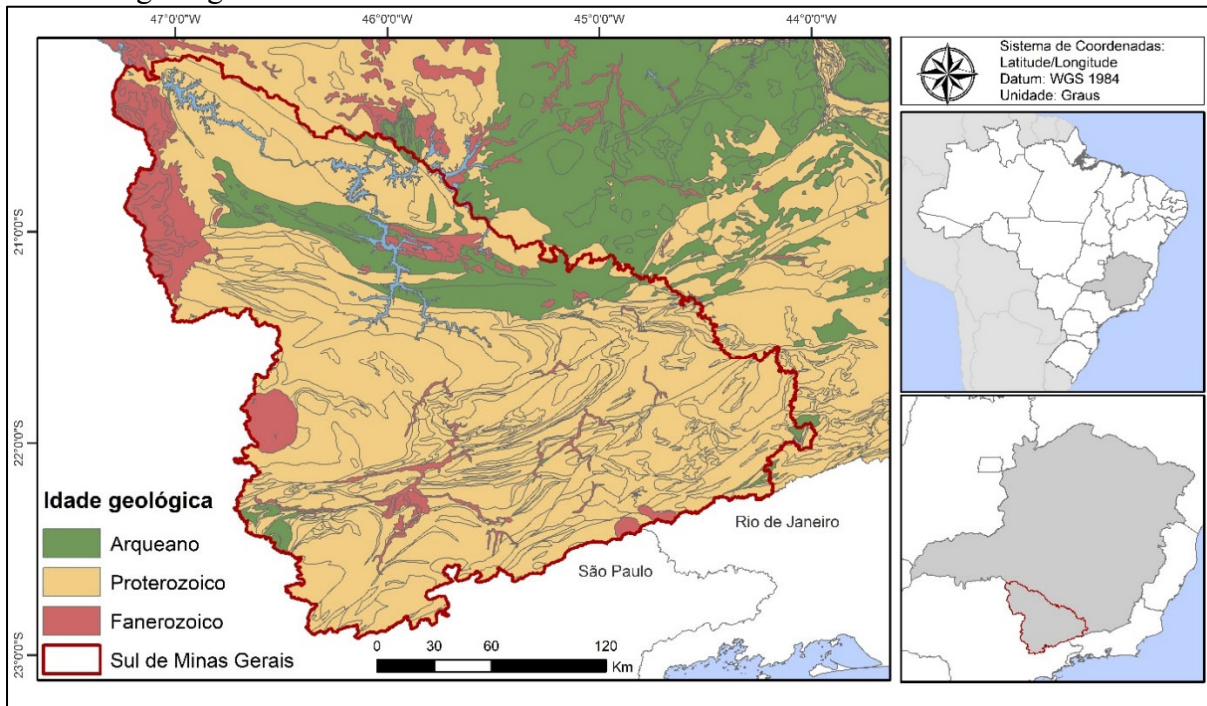
de uso desses dados tem sido explorado em outros países para pesquisas na área de MDS, obtendo boas respostas para construção de mapas de solos (ELLILI et al., 2019) e de intensidade do intemperismo (WILFORD, 2012).

O tempo é um dos fatores mais complexos de representar. Estabelecer limites dentro dos quais a pedogênese ocorre desconsidera a formação poligenética da maioria dos solos (JOHNSON; WATSON-STEGNER, 1987). Em condições tropicais os processos de pedogênese estão intimamente ligados às taxas de erosão (RESENDE et al., 2014) Considerando esse aspecto, o fator tempo pode ser representado de maneira implícita em outras variáveis usadas no MDS, como o relevo, os modelos paleoclimáticos e a espectrometria de raios gama.

2.2 Geomorfologia e solos da região Sul de Minas Gerais

A geologia observada na região Sul de Minas Gerais pode ser compreendida a partir da divisão do território em dois tipos de terreno: os antigos, constituídos de rochas de idade arqueana e proterozóica; e os jovens, que compõem a cobertura formada por rochas de idade fanerozóica (ALMEIDA, 2000; ALMEIDA et al., 1981) (Figura 1). Nos terrenos antigos predominam as rochas metamórficas e ígneas que englobam os crátons, terrenos mais estáveis antigos, e os sistemas orogênicos, formados por cadeias de montanhas soerguidas em consequência de colisões entre placas litosféricas durante o chamado Evento Brasileiro. Os terrenos jovens são formados por injeções magmáticas alcalinas, carbonáticas que se manifestaram em superfície na forma de vulcões (ALMEIDA et al., 1981; BRITO NEVES; CAMPOS-NETO; FUCK, 1999; CAMPOS-NETO, 2000; ALMEIDA, 2000).

Figura 1- Mapa de idade geológica do Sul de Minas, adaptado do mapa geológico de Minas Gerais (CPRM, 2014) destacando a região Sul de Minas e a idade das formações geológicas.



Na parte Oeste da região, destaca-se a Serra da Canastra, soerguida durante o Evento Brasiliano e constituída por rochas metassedimentares de origem marinha, predominantemente filitos e quartzitos (DIAS et al., 2013). Na parte Leste, as formações orogênicas brasileiras deram origem a um quadro geológico de alta complexidade, que constituem o conjunto de expressões morfológicas do Campo das Vertentes e da Serra da Mantiqueira. As unidades litológicas estão em grau metamórfico elevado e refletem a história evolutiva resultante das múltiplas colisões durante o Evento Brasiliano (HEILBRON et al., 2017).

As áreas cobertas por camadas de rochas de idade fanerozóica ficam restritas a uma pequena parte no oeste da região, próxima a região do Triângulo Mineiro, cuja principal unidade de preenchimento é a formação chamada Serra Geral, constituída essencialmente por basaltos de granulação fina, que ocorrem nos grandes vales da região como uma sucessão de lavas extravasadas a partir de fraturas abertas durante o período fanerozóico (BARBOSA et al., 1970). Dentre as intrusões de rochas ígneas, se destaca o Complexo de Poços de Caldas localizado junto à fronteira com São Paulo, que forma uma estrutura circular de 30 km de diâmetro, constituída por grandes massas de depósitos alcalinos ricos em bauxita, zircônio e molibdênio (ULBRICH; GOMES, 1981).

A maioria dos sistemas brasileiros corresponde às terras altas, ao passo que os crátons, parcial ou totalmente cobertos por rochas sedimentares fanerozóicas e hospedeiros das grandes bacias hidrográficas atuais, se expressam na forma das terras baixas do território. Assim, apesar de passados mais de 500 milhões de anos da sua geração, essas formações mantêm as suas posições relativas no relevo da região (ALKMIM, 2020) e são determinantes para a constituição dos solos do Sul de Minas.

Até 2010, os solos do Sul de Minas contavam com um mapa na mesma escala que todo o estado de Minas Gerais, 1:1.000.000. Contudo, embora de forma esparsa, várias regiões foram estudadas durante as décadas de 1950-1960 pelos órgãos especializados do Ministério da Agricultura, tais como a EPFS (Equipe de Pedologia e Fertilidade dos Solos), DPP (Divisão de Pesquisa Pedológica), CNEPA (Centro Nacional de Ensino e Pesquisas Agronômicas) e SNLCS (Serviço Nacional de Levantamento e Conservação dos Solos), atual CNPS (Centro Nacional de Pesquisa de Solos) da EMBRAPA. Das regiões abrangidas por esses estudos, apenas a “Área de Influência do Reservatório de Furnas” (CNEPA, 1962) está localizada no Sul de Minas.

Três anos mais tarde, no ano de 2007, o CETEC elaborou o Mapa de Solos do Estado de Minas Gerais na escala 1:650.000 (CETEC et al., 2010), obtido por compilação dos levantamentos já realizados e complementado por um levantamento de maior detalhamento amostral das porções mineiras das bacias hidrográficas dos rios Doce, Grande e Paraíba do Sul, obtido através de levantamentos de solos de média intensidade. Destes locais, parte da bacia do rio Grande está em território sul mineiro e contou com amostras adicionais. Até o momento, este mapa é o mais detalhado disponível para a região Sul de Minas. Nele, os solos são classificados até o quarto nível categórico segundo o SiBCS. Neste levantamento foram identificadas 13 Subordens de solos.

No mapa de solos da CETEC (CETEC et al., 2010) para o Sul de Minas Gerais, a proporção das diferentes classes de solo na região Sul de Minas Gerais reflete a característica dos solos de ambientes tropicais, desenvolvidos sob marcantes e prolongadas condições de altas temperaturas e umidade. Essa combinação proporciona maior ocorrência de solos em estágio avançado de intemperismo-lixiviação, com baixa atividade das argilas, intensa remoção de sílica e bases e concentração residual de óxidos (RESENDE et al., 2014). Com essas características e seguindo uma lógica de evolução pedogenética, se enquadram os LATOSSOLOS que ocorrem em 46% da área do Sul de Minas Gerais. São os solos mais intemperizados e profundos, com estrutura granular bem desenvolvida. Os ARGISSOLOS,

27% da área, com a característica de acúmulo de argila no horizonte subsuperficial. As proporções de ocorrência destes solos no mapa somam 73% da área, demonstrando a predominância dessas classes da paisagem do Sul de Minas. Os CAMBISSOLOS, considerados mais jovens em uma escala de evolução pedogenética, ocorrem em 18% da área. São solos pouco desenvolvidos, apresentando maiores proporções relativas de minerais primários facilmente intemperizáveis e maiores relações silte/argila. Ocupam geralmente posições de relevo mais acidentado na paisagem, porém, eventos e ambientes pretéritos não expressos no relevo atual tornam mais complexas sua identificação e diferenciação (SILVA et al., 2016; COELHO; ROSSI; MATTOS, 2017).

3 CONSIDERAÇÕES GERAIS

Novas covariáveis ambientais inéditas em estudos de solos no Brasil se mostraram promissoras para aplicação no MDS. Os dados de espectrometria gama, já reconhecidamente importantes para a predição de classes e atributos de solo em estudos internacionais, foram aplicados com sucesso na predição e configuraram como as principais covariáveis preditoras. Outro destaque, foi o resultado obtido para a variável Paleoclima, representada pela estimativa de precipitação para 22.000 anos atrás. Este dado, disponível para todo o globo, ainda não havia sido aplicado ao MDS no Brasil e indica um caminho para aprofundar estudos sobre a influência de climas pretéritos na formação dos solos tropicais. As variáveis de relevo, já consagradas no MDS completaram o conjunto dos dados mais importantes para construir o mapa de solos da região Sul de Minas Gerais com 89% de acurácia.

REFERÊNCIAS

- 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**, v. 214–215, n. 2014, p. 101–113, 2014. doi: 10.1016/j.geoderma.2013.09.023.
- ALKMIM, F. F. **História Geológica de Minas Gerais**. Acesso em: 26 jul. 2021.
- ALMEIDA, F. F. M. de. Origem e evolução da plataforma brasileira. In: **Boletim 241**. Rio de Janeiro: DNPM/Divisão de Geologia e Mineralogia, 2000. p. 36. 2000.
- AMARAL, F. C. S. do; SANTOS, H. G. dos; ÁGILO, M. L. D.; DUARTE, M. N.; OLIVEIRA, R. P. de; CARVALHO JÚNIOR, W. **Mapeamento de Solos e Aptidão Agrícola das terras do estado de Minas Gerais**. Rio de Janeiro: Embrapa, 2004.
- ANA, A. N. de Á. **Plano Integrado de Recursos Hídricos da Bacia Hidrográfica do Rio Grande**, 2017.
- ARAÚJO, A. R.; OLIVEIRA, J. M.; PEREIRA, P.; CURI, N.; MARQUES, J. J. G. M. **Geomorfologia , solos e aptidão agrícola das terras da bacia do Alto Rio Grande , Minas Gerais**. Lavras: Ed: UFLA, 2018.
- ARROUAYS, D.; MCBRATNEY, A.; BOUMA, J.; LIBOHOVA, Z.; RICHER-DE-FORGES, A. C.; MORGAN, C. L. S.; ROUDIER, P.; POGGIO, L.; MULDER, V. L. Impressions of digital soil maps: The good, the not so good, and making them ever better. **Geoderma Regional**, v. 20, p. e00255, 2020a. doi: 10.1016/j.geodrs.2020.e00255.
- ARROUAYS, D.; POGGIO, L.; SALAZAR GUERRERO, O. A.; MULDER, V. L. Digital soil mapping and GlobalSoilMap. Main advances and ways forward. **Geoderma Regional**, v. 21, p. e00265, 2020b. doi: 10.1016/j.geodrs.2020.e00265.
- BARBOSA, O.; BRAUN, O. P. G.; DYER, R. C.; CUNHA, C. A. B. R. Geologia da região do Triângulo Mineiro. **Boletim DNPM/DFPM**, 136, p. 140, 1970.
- BEAMISH, D. Gamma ray attenuation in the soils of Northern Ireland, with special reference to peat. **Journal of Environmental Radioactivity**, v. 115, p. 13–27, 2013. doi: 10.1016/j.jenvrad.2012.05.031.
- BERTOLDO, M. A. **CARACTERIZAÇÃO EDAFOAMBIENTAL DA CAFEICULTURA NA REGIÃO DE TRÊS PONTAS, MINAS GERAIS**. 2008. Universidade Federal de Lavras - UFLA, 2008.
- BRASIL, M. das M. e E. **Projeto Radam Brasil. Levantamento de Recursos Naturais. Volume 32. Folhas SF 23/24 Rio de Janeiro/Vitória**. Rio de Janeiro: MME, 1983.

BREIMAN, L. Random Forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001. doi: 10.1023/A:1010933404324.

BREVIK, E. C.; CALZOLARI, C.; MILLER, B. A.; PEREIRA, P.; KABALA, C.; BAUMGARTEN, A.; JORDÁN, A. Soil mapping, classification, and pedologic modeling: History and future directions. **Geoderma**, v. 264, p. 256–274, 2016. doi: 10.1016/j.geoderma.2015.05.017.

BRITO NEVES, B. B.; CAMPOS-NETO, M. C.; FUCK, R. A. From Rodinia to Western Gondwana: An approach to the Brasiliano-Pan-Africa. v. 22, n. 3, p. 155–166, 1999.

BRUNGARD, C. W.; BOETTINGER, J. L.; DUNIWAY, M. C.; WILLS, S. A.; EDWARDS, T. C. Machine learning for predicting soil classes in three semi-arid landscapes. **Geoderma**, v. 239–240, p. 68–83, 1 fev. 2015. doi: 10.1016/j.geoderma.2014.09.019.

BUI, E. N. Soil survey as a knowledge system. **Geoderma**, v. 120, n. 1–2, p. 17–26, 2004. doi: 10.1016/j.geoderma.2003.07.006.

CAIRES, S.; FONTES, M. P. F.; KER, J. C. Dados para o mapa de solos de Minas Gerais. In: UFV (Ed.). **Infraestrutura de Dados Espaciais do Sistema Estadual de Meio Ambiente e Recursos Hídricos**. Belo Horizonte: SISEMA, 2007. 2007.

CAMERA, C.; ZOMENI, Z.; NOLLER, J. S.; ZISSIMOS, A. M.; CHRISTOFOROU, I. C.; BRUGGEMAN, A. Geoderma A high resolution map of soil types and physical properties for Cyprus : A digital soil mapping optimization. **Geoderma**, v. 285, p. 35–49, 2017. doi: 10.1016/j.geoderma.2016.09.019.

CAMPOS-NETO, M. C. **Orogenic Systems from Southwestern Gondwana: An approach to Brasiliano-PanAfrican Cycle and Orogenic Collage in Southeastern Brazil**. Rio de Janeiro: 31º International Geological Congress, 2000.

CARVALHO, C. C.; NUNES, F. C.; ANTUNES, A. M. H. Histórico do levantamento de solos no Brasil: Da industrialização brasileira à era da informação. **Revista brasileira de cartografia**, v. 65, p. 997–1013, 2013.

CARVALHO, A. P. DE; LARACH, J. O. I.; CAMARGO, M. N.; OLIVEIRA, J. B. de; SANTOS, H. G. dos; WITTERN, E. P.; CONCEIÇÃO, M. da; TAVARES, N. P.; LOUZADA, E. S.; RAMOS, D. P.; PADRO, H.; SOUZA, J. L. R. de; MONIZ, A. C.; ALMEIDA, C. L. F.; DURIEZ, M. A. de; JOHAS, R. A. L.; MELO, M. E. C. C. de; ARAUJO, W. S. de; BLOISE, R. M.; MOREIRA, G. N. C.; PAULA, J. L. de; LIMA, T. da C.; RODRIGUES, E. M.; ANTONELLO, L. L.; LIMA, P. C.; PINTO, J. da S. **Conjunto de dados dos Anais da II Reunião de Classificação, Correlação de Solo e Interpretação de Aptidão Agrícola**. 1º ed. Rio de Janeiro: Embrapa Solos, 1983a.

CARVALHO, A. P. de; LARACH, J. O. I.; CAMARGO, M. N.; OLIVEIRA, J. B. de; SANTOS, H. G. dos; WITTERN, E. P. M. K. P.; CONCEIÇÃO, M. da; TAVARES, N. P.;

LOUZADA, E. S.; RAMOS, D. P.; PADRO, H.; SOUZA, J. L. R. de; MONIZ, A. C.; ALMEIDA, C. L. F. de; DURIEZ, M. A. de M.; JOHAS, R. A. L.; MELO, M. E. C. C. de M.; ARAUJO, W. S. de; BLOISE, R. M.; MOREIRA, G. N. C.; PAULA, J. L. de; LIMA, T. da C.; RODRIGUES, E. M.; ANTONELLO, L. L.; LIMA, P. C.; PINTO, J. da S. **Conjunto de dados dos Anais da II Reunião de Classificação, Correlação de Solo e Interpretação de Aptidão Agrícola**. Rio de Janeiro: Embrapa Solos, 1983b.

CATTLE, S. R.; MEAKIN, S. N.; RUSZKOWSKI, P.; CAMERON, R. G. Using radiometric data to identify aeolian dust additions to topsoil of the Hillston district, western NSW. **Soil Research**, v. 41, n. 8, p. 1439–1456, 2003. doi: 10.1071/SR03026.

CETEC; FEAM; UFV; UFLA. **Mapa de solos do Estado de Minas Gerais**. 1º ed. Belo Horizonte: Fundação Estadual do Meio Ambiente - FEAM, 2010.

CHAKRABORTY, S.; MAN, T.; PAULETTE, L.; DEB, S.; LI, B.; WEINDORF, D. C.; FRAZIER, M. Rapid assessment of smelter/mining soil contamination via portable X-ray fluorescence spectrometry and indicator kriging. **Geoderma**, v. 306, n. April, p. 108–119, 2017. doi: 10.1016/j.geoderma.2017.07.003.

CNEPA, C. N. de P. A. **Levantamento de reconhecimento dos solos da região sob influência do reservatório de Furnas**. 1º ed. Rio de Janeiro: Ministério da Agricultura, 1962.

COELHO, F. F.; GIASSON, E.; CAMPOS, A. R.; TIECHER, T.; COSTA, J. J. F.; COBLINSKI, J. A. Digital soil class mapping in Brazil: A systematic review. **Scientia Agricola**, v. 78, n. 5, p. 1–11, 2021. doi: 10.1590/1678-992x-2019-0227.

COELHO, R. M.; ROSSI, M.; MATTOS, I. F. de A. Solos da Mata Atlântica. In: **Pedologia: Solos dos Biomas Brasileiros**. 1º ed. Viçosa, MG: Sociedade Brasileira de Ciência do Solo, 2017. p. 261–302. 2017.

COHEN, J. A coefficient of agreement for nominal scales. **Educational and Psychological Measurement**, v. 20, n. 1, p. 37- 46 ST- A coefficient of agreement for nominal, 1960.

CONAB. Acompanhamento da safra brasileira de café - Primeiro Levantamento. **Companhia Nacional de Abastecimento**, v. 6, 2019.

CONGALTON, R. G.; GREEN, K. **Assessing the Accuracy of Remotely Sensed Data Principles and Practices**. 3. ed. London: CRC Press, 2019. v. 53

CONRAD, O.; BECHTEL, B.; BOCK, M.; DIETRICH, H.; FISCHER, E.; GERLITZ, L.; WEHBERG, J.; WICHMANN, V.; BÖHNER, J. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. **Geoscientific Model Development**, v. 8, n. 7, p. 1991–2007, 2015. doi: 10.5194/gmd-8-1991-2015.

COOK, S. E.; CORNER, R. J.; GROVES, P. R.; GREALISH, G. J. Use of airborne gamma

radiometric data for soil mapping. **Australian Journal of Soil Research**, v. 34, n. 1, p. 183–194, 1996. doi: 10.1071/SR9960183.

CUNHA, S. B. da. Geomorfologia fluvial. In: CUNHA, S. B. C.; GUERRA, A. J. T. (Ed.). **Geomorfologia: Exercícios, técnicas e aplicações**. Rio de Janeiro: Bertrand Brasil, 2013. p. 348. 2013.

CURI, N.; RESENDE, M.; SANTANA, D. P. Solos de várzeas de Minas Gerais. **Informe Agropecuário**, v. 13, n. 152, p. 3–10, 1985.

DE ALMEIDA, F. F. M.; HASUI, Y.; DE BRITO NEVES, B. B.; FUCK, R. A. Brazilian structural provinces: An introduction. **Earth Science Reviews**, v. 17, n. 1–2, p. 1–29, 1981. doi: 10.1016/0012-8252(81)90003-9.

DIAS, P. H. A.; NOCE, C. M.; PEDROSA-SOARES, A. C.; SEER, H. J.; DUSSIN, I. A.; VALERIANO, C. de M.; KUCHENBECKER, M. O Grupo Ibiá (Faixa Brasília Meridional): Evidências Isotópicas Sm-Nd E U-Pb De Bacia Colisional Tipo Flysch. **Geonomos**, v. 19, p. 90–99, 2013. doi: 10.18285/geonomos.v19i2.45.

DUDA, B. M.; WEINDORF, D. C.; CHAKRABORTY, S.; LI, B.; MAN, T.; PAULETTE, L.; DEB, S. Soil characterization across catenas via advanced proximal sensors. **Geoderma**, v. 298, p. 78–91, 2017. doi: 10.1016/j.geoderma.2017.03.017.

EITEN, G. Vegetação do Cerrado. In: PINTO, M. N. (Ed.). **Cerrado: caracterização, ocupação e perspectivas**. Brasília: UnB/SEMATEC, 1994. p. 17–73. 1994.

ELITH, J.; KEARNEY, M.; PHILLIPS, S. The art of modelling range-shifting species. **Methods in Ecology and Evolution**, v. 1, n. 4, p. 330–342, 2010. doi: 10.1111/j.2041-210x.2010.00036.x.

ELLILI, Y.; MALONE, B. P.; MICHOT, D.; MINASNY, B.; VINCENT, S.; WALTER, C.; LEMERCIER, B. Comparing three approaches of spatial disaggregation of legacy soil maps based on DSMART algorithm. **Soil Discussions**, n. June, p. 1–39, 2019.

EMBRAPA. **Procedimentos normativos de levantamentos pedológicos**. 1 Ed. ed. Brasília: Embrapa, 1995.

EMBRAPA. Programa Nacional de Solos do Brasil (Pronasolos). In: KER, J. C.; CURI, N.; SCHAEFER, C. E. G. R.; VIDAL-TORRADO, P. (Ed.). **Série Documentos**. 1º ed. Rio de Janeiro: Embrapa Solos, 2016. 183 p. 93. 2016.

ESRI. **ArcGIS Desktop: Release 10.2Redlands CA**, 2013. .

ETTEN, J. Van; SUMNER, M.; CHENG, J.; BASTON, D.; BEVAN, A.; BIVAND, R.; Busetto, L.; CANTY, M.; FASOLI, B.; FORREST, D.; GRAY, J.; GREENBERG, J. A.; HIEMSTRA, P.; KARNEY, C.; MATTIUZZI, M. Package ‘ raster ’ R topics documented :

2021.

FAO, F. and A. O. of U. N. **Our Soils Under Threat** International Year of Soils fao.org, , 2015. .

FELFILI, J. M.; DA SILVA, M. C. A comparative study of cerrado (sensu stricto) vegetation in Central Brazil. **Journal of Tropical Ecology**, v. 9, n. 3, p. 277–289, 1993. doi: 10.1017/S0266467400007306.

FERNANDES, M. M.; MARQUES, F. S. M. e. Dados para o mapa de solos de Minas Gerais. In: CETEC (Ed.). **Infraestrutura de Dados Espaciais do Sistema Estadual de Meio Ambiente e Recursos Hídricos**. Belo Horizonte: SISEMA, 2007. 2007.

FICK, S. E.; HIJMANS, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. **International Journal of Climatology**, v. 37, n. 12, p. 4302–4315, 2017. doi: 10.1002/joc.5086.

FLACH, C. W.; CORRÊA, E. A. Levantamento De Solos No Brasil: Métodos, Práticas E Dificuldades. **Geographia Meridionalis**, v. 3, n. 3, p. 420, 2018. doi: 10.15210/gm.v3i3.12237.

FRIEDMAN, J.; HASTIE, T.; TIBSHIRANI, R. *The Elements of Statistical Learning Preface to the Second Edition*. 2008.

GAN-LIN, Z.; FENG, L. I. U.; XIAO-DONG, S.; ZHANG, G. lin; LIU, F.; SONG, X. dong. Recent progress and future prospect of digital soil mapping: A review. **Journal of Integrative Agriculture**, v. 16, n. 12, p. 2871–2885, 2017. doi: 10.1016/S2095-3119(17)61762-3.

GATEHOUSE, R. D.; WILLIAMS, I. S.; PILLANS, B. J. Fingerprint windblow dust in south-eastern Australian soils by uranium-lead dating of detrital zircon. **Australian Journal of Soil Research**, v. 39, n. 1, p. 7–12, 2001. doi: 10.1071/SR99078.

GIAROLA, N. F. B.; CURI, N.; SIQUEIRA, J. O.; CHAGAS, C. da S.; FERREIRA, M. M. **Solos da regio sob influência do reservatório da hidrelétrica de Itutinga/Camargos (MG): perspectiva ambiental**. Lavras, MG: CEMIG, 1997.

GOMES, L. C.; FARIA, R. M.; DE SOUZA, E.; VELOSO, G. V.; SCHAEFER, C. E. G. R.; FILHO, E. I. F. Modelling and mapping soil organic carbon stocks in Brazil. **Geoderma**, v. 340, n. December 2017, p. 337–350, 2019. doi: 10.1016/j.geoderma.2019.01.007.

GONÇALVES, M. G. M. **Soil Survey As Support To Precision Coffee Crop and Winter Wines Development in Southeast Lavras - Mg**. 2020. Universidade Federal de Lavras, 2020.

GRIMM, R.; BEHRENS, T.; MÄRKER, M.; ELSENBEEER, H. Soil organic carbon

concentrations and stocks on Barro Colorado Island - Digital soil mapping using Random Forests analysis. **Geoderma**, 2008. doi: 10.1016/j.geoderma.2008.05.008.

HARIDASAN, M. Aluminium accumulation by some cerrado native species of central Brazil. **Plant and Soil**, v. 65, n. 2, p. 265–273, 1982. doi: 10.1007/BF02374657.

HARTEMINK, A. E. On global soil science and regional solutions. **Geoderma Regional**, v. 5, p. 1–3, 2015. doi: 10.1016/j.geodrs.2015.02.001.

HARTEMINK, A. E.; MCBRATNEY, A. A soil science renaissance. **Geoderma**, v. 148, n. 2, p. 123–129, 2008. doi: 10.1016/j.geoderma.2008.10.006.

HEILBRON, M.; RIBEIRO, A.; VALERIANO, C. N.; PACIULLO, F. V.; ALMEIDA, J. C. H.; TROUW, R. J. A.; TUPINAMBÁ, M.; SILVA, L. G. E. The Ribeira Belt. **Springer**, p. 326, 2017. doi: 10.1007/978-3-319-01715-0_15.

HENGL, T. Finding the right pixel size. **Computers and Geosciences**, v. 32, n. 9, p. 1283–1298, 2006. doi: 10.1016/j.cageo.2005.11.008.

HENGL, T.; DE JESUS, J. M.; HEUVELINK, G. B. M.; GONZALEZ, M. R.; KILIBARDA, M.; BLAGOTIĆ, A.; SHANGGUAN, W.; WRIGHT, M. N.; GENG, X.; BAUER-MARSCHALLINGER, B.; GUEVARA, M. A.; VARGAS, R.; MACMILLAN, R. A.; BATJES, N. H.; LEENAARS, J. G. B.; RIBEIRO, E.; WHEELER, I.; MANTEL, S.; KEMPEN, B. SoilGrids250m: Global gridded soil information based on machine learning. **PLoS ONE**, v. 12, n. 2, p. 1–40, 2017. doi: 10.1371/journal.pone.0169748.

HEUNG, B.; BULMER, C. E.; SCHMIDT, M. G. Predictive soil parent material mapping at a regional-scale: A Random Forest approach. **Geoderma**, v. 214–215, p. 141–154, 2014. doi: 10.1016/j.geoderma.2013.09.016.

HEUNG, B.; HO, H. C.; ZHANG, J.; KNUDBY, A.; BULMER, C. E.; SCHMIDT, M. G. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. **Geoderma**, v. 265, p. 62–77, 2016. doi: 10.1016/j.geoderma.2015.11.014.

HIJMANS, A. R. J.; PHILLIPS, S.; LEATHWICK, J.; ELITH, J.; HIJMANS, M. R. J. Package ‘dismo’. 2020.

HOUNKPATIN, K. O. L.; SCHMIDT, K.; STUMPF, F.; FORKUOR, G.; BEHRENS, T.; SCHOLTEN, T.; AMELUNG, W.; WELP, G. Predicting reference soil groups using legacy data: A data pruning and Random Forest approach for tropical environment (Dano catchment, Burkina Faso). **Scientific Reports**, v. 8, n. 1, p. 1–16, 2018. doi: 10.1038/s41598-018-28244-w.

HUDSON, H. D. Division s-5-soil genesis, morphology & classification. v. 841, p. 836–841, 1992.

IAEA. **Guidelines for Radioelement Mapping Using Gamma-ray Spectrometry Data**. International Atomic Energy Agency, Austria, 2003.

IBGE. Manual técnico de pedologia. In: **Manuais Técnicos em Geociências**. 2. ed. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística - IBGE, 2007. p. 316. 2007.

INDA JUNIOR, A. V.; KÄMPF, N. Variabilidade de goethita e hematita via dissolução redutiva em solos de região tropical e subtropical. **Revista Brasileira de Ciência do Solo**, v. 29, n. 6, p. 851–866, 2005. doi: 10.1590/s0100-06832005000600003.

JAFARI, A.; AYOUBI, S.; KHADEMI, H.; FINKE, P. A.; TOOMANIAN, N. Selection of a taxonomic level for soil mapping using diversity and map purity indices: A case study from an Iranian arid region. **Geomorphology**, v. 201, p. 86–97, 2013. doi: 10.1016/j.geomorph.2013.06.010.

JAFARI, A.; KHADEMI, H.; FINKE, P. A.; VAN DE WAUW, J.; AYOUBI, S. Spatial prediction of soil great groups by boosted regression trees using a limited point dataset in an arid region, southeastern Iran. **Geoderma**, 2014. doi: 10.1016/j.geoderma.2014.04.029.

JENNY, H. **Factors of soil formation: a system of quantitative pedology**. 1. ed. New York: Soil Science, 1941.

JOHNSON, D. L.; WATSON-STEGNER, D. Evolution model of pedogenesis. **Soil Science**, v. 143, n. 5, p. 349–366, 1987. doi: 10.1097/00010694-198705000-00005.

KÄMPF, N.; CURI, N. Conceito de solo e sua evolução histórica. In: KER, J. C.; CURI, N.; SCHAEFER, C. E. G. R.; VIDAL-TORRADO, P. (Ed.). **Pedologia: Fundamentos**. 1. ed. Viçosa: Sociedade Brasileira de Ciência do Solo, 2012a. p. 343. 2012.

KÄMPF, N.; CURI, N. Formação e evolução do solo (Pedogênese). In: KER, J. C.; CURI, N.; SCHAEFER, C. E. G. R.; VIDAL-TORRADO, P. (Ed.). **Pedologia: Fundamentos**. 1. ed. Viçosa - MG: Sociedade Brasileira de Ciência do Solo, 2012b. p. 207–294. 2012.

KER, J. C. Latossolos Do Brasil: Uma Revisão. **Geonomos**, v. 5, n. 1, p. 17–40, 1997. doi: 10.18285/geonomos.v5i1.187.

KER, J. C.; MOTTA, P. E.; OLIVEIRA, V. Á. de. Levantamentos pedológicos e evolução do conhecimento dos solos no Brasil. In: CURI, N.; KER, J.; NOVAIS, R.; VIDAL-TORRADO, P.; SCHAEFER, C. (Ed.). **Pedologia: Solos dos biomas brasileiros**. 1 ed. ed. Viçosa: Sociedade Brasileira de Ciência do Solo, 2017. p. 1–45. 2017.

KHALEDIAN, Y.; MILLER, B. A. Selecting appropriate machine learning methods for digital soil mapping. **Applied Mathematical Modelling**, v. 81, p. 401–418, 2020. doi: 10.1016/j.apm.2019.12.016.

KIDD, D.; SEARLE, R.; GRUNDY, M.; MCBRATNEY, A.; ROBINSON, N.; O'BRIEN, L.;

ZUND, P.; ARROUAYS, D.; THOMAS, M.; PADARIAN, J.; JONES, E.; BENNETT, J. M. L.; MINASNY, B.; HOLMES, K.; MALONE, B. P.; LIDDICOAT, C.; MEIER, E. A.; STOCKMANN, U.; WILSON, P.; WILFORD, J.; PAYNE, J.; RINGROSE-VOASE, A.; SLATER, B.; ODGERS, N.; GRAY, J.; VAN GOOL, D.; ANDREWS, K.; HARMS, B.; STOWER, L.; TRIANTAFILIS, J. Operationalising digital soil mapping – Lessons from Australia. **Geoderma Regional**, v. 23, p. e00335, 2020. doi: 10.1016/j.geodrs.2020.e00335.

KRAMM, T.; HOFFMEISTER, D. Assessing the influence of environmental factors and datasets on soil type prediction with two machine learning algorithms in a heterogeneous area in the Rur catchment, Germany. **Geoderma Regional**, v. 22, p. e00316, 2020. doi: 10.1016/j.geodrs.2020.e00316.

LAGACHERIE, P.; MCBRATNEY, A. B. Spacial soil information systems and spatial inference systems: perspectives for digital soil mapping. In: LAGACHERIE, P.; MCBRATNEY, A. B.; VOLTZ, M. (Ed.). **Developments in Soil Science**. 1. ed. New York: Elsevier B.V., 2007. 2007.

LAHJOUJ, A.; EL HMAIDI, A.; BOUHAFSA, K.; BOUFALA, M. Mapping specific groundwater vulnerability to nitrate using random forest: case of Sais basin, Morocco. **Modeling Earth Systems and Environment**, v. 6, n. 3, p. 1451–1466, 2020. doi: 10.1007/s40808-020-00761-6.

LEENAARS, J. G. B.; ELIAS, E.; WÖSTEN, J. H. M.; RUIPEREZ-GONZÁLEZ, M.; KEMPEN, B. Mapping the major soil-landscape resources of the Ethiopian Highlands using random forest. **Geoderma**, v. 361, n. October 2019, p. 114067, 2020. doi: 10.1016/j.geoderma.2019.114067.

LEGROS, J.-P. **Mapping of the soil**. 1st. ed. New Hampshire: Publishers, Science, 2006.

LI, Y.; ZHU, A.; SHI, Z.; LIU, J.; DU, F. Geoderma Supplemental sampling for digital soil mapping based on prediction uncertainty from both the feature domain and the spatial domain. **Geoderma**, v. 284, p. 73–84, 2016. doi: 10.1016/j.geoderma.2016.08.013.

MA, Y.; MINASNY, B.; MALONE, B. P.; MCBRATNEY, A. B. Pedology and digital soil mapping (DSM). **European Journal of Soil Science**, v. 70, n. 2, p. 216–235, 2019. doi: 10.1111/ejss.12790.

MACHADO, D. F. T. **Accuracy and uncertainty of digital soil mapping approaches to extract and transfer soil information from reference area**. 2017. Universidade Federal de Lavras, 2017.

MACHADO, D. F. T.; DE MENEZES, M. D.; SILVA, S. H. G.; CURI, N. Transferability, accuracy, and uncertainty assessment of different knowledge-based approaches for soil types mapping. **Catena**, v. 182, n. September 2018, p. 104134, 2019a. doi: 10.1016/j.catena.2019.104134.

- MACHADO, D. F. T.; SILVA, S. H. G.; CURI, N.; DE MENEZES, M. D. Soil type spatial prediction from random forest: Different training datasets, transferability, accuracy and uncertainty assessment. **Scientia Agricola**, v. 76, n. 3, p. 243–254, 2019b. doi: 10.1590/1678-992x-2017-0300.
- MANCINI, M.; WEINDORF, D. C.; SILVA, S. H. G.; CHAKRABORTY, S.; TEIXEIRA, A. F. dos S.; GUILHERME, L. R. G.; CURI, N. Parent material distribution mapping from tropical soils data via machine learning and portable X-ray fluorescence (pXRF) spectrometry in Brazil. **Geoderma**, v. 354, n. July, p. 113885, 2019. doi: 10.1016/j.geoderma.2019.113885.
- MCBRATNEY, A. . B.; MENDONÇA SANTOS, M. . L.; MINASNY, B. On digital soil mapping. **Geoderma**, v. 117, n. 1–2, p. 3–52, 1 nov. 2003. doi: 10.1016/S0016-7061(03)00223-4.
- MCBRATNEY, A. B.; ODEH, I. O. A.; BISHOP, T. F. A.; DUNBAR, M. S.; SHATAR, T. M. An overview of pedometric techniques for use in soil survey. **Geoderma**, v. 97, n. 3–4, p. 293–327, 2000. doi: 10.1016/S0016-7061(00)00043-4.
- MCBRATNEY, A.; FIELD, D. J.; KOCH, A. The dimensions of soil security. **Geoderma**, v. 213, p. 203–213, 2014. doi: 10.1016/j.geoderma.2013.08.013.
- MELLO, D. C. de; DEMATTÊ, J. A. M.; MELLO, F. A. de O.; POPPIEL, R. R.; SILVERO, N. E.; SAFANELLI, J. L.; SOUZA, A. B. e; RAIMO, L. A. Di; RIZZO, R.; RESENDE, M. E. B.; GONÇALVES, S. E. Applied gamma-ray spectrometry for evaluating tropical soil processes and attributes. **Geoderma**, v. 381, n. September 2020, p. 114736, 2021. doi: 10.1016/j.geoderma.2020.114736.
- MENDONÇA-SANTOS, M. de L.; SANTOS, H. G. dos. Mapeamento Digital de Classes e Atributos de Solos métodos, paradigmas e novas técnicas. **Embrapa Solos**, p. 17, 2003.
- MENDONÇA-SANTOS, M. L.; SANTOS, H. The state of the art of Brazilian soil mapping and prospects for digital soil mapping. In: LAGACHERIE, P.; MCBRATNEY, A. B.; VOLTZ, M. (Ed.). **Digital Soil Mapping: An introductory perspective**. New York, NY: Elsevier, 2007. 2007.
- MENEZES, M. D. de. **Levantamento de solos em sub-bacias hidrográficas como apoio para avaliação do uso da terra e da potencialidade de recarga de aquíferos**. 2007. UFLA, 2007.
- MENEZES, M. D. de; BISPO, F. H. A.; FARIA, W. M.; GONÇALVES, M. G. M.; CURI, N.; GUILHERME, L. R. G. Modeling arsenic content in Brazilian soils: What is relevant? **Science of the Total Environment**, v. 712, p. 136511, 2020. doi: 10.1016/j.scitotenv.2020.136511.
- MENEZES, M. D. de; MESSIAS, C. G.; SILVA, S. H. G.; CURI, N. Relações entre ambientes de solos e fitofisionomias. In: CARVALHO, D. A. DE (Ed.). **Cerrados do Sul e**

Sudoeste de Minas Gerais: Flora e Ambiente. Lavras: Ed. UFLA, 2017. p. 123–147. 2017.

MENEZES, M. D. de; SILVA, S. H. G.; OWENS, P. R.; CURI, N. Digital soil mapping approach based on fuzzy logic and field expert knowledge. **Ciência e Agrotecnologia**, v. 37, n. 4, p. 287–298, 2013. doi: 10.1590/s1413-70542013000400001.

MILLARD, K.; RICHARDSON, M. On the importance of training data sample selection in Random Forest image classification: A case study in peatland ecosystem mapping. **Remote Sensing**, v. 7, n. 7, p. 8489–8515, 2015. doi: 10.3390/rs70708489.

MINASNY, B.; MCBRATNEY, A. B. Digital soil mapping: A brief history and some lessons. **Geoderma**, v. 264, p. 301–311, 2016. doi: 10.1016/j.geoderma.2015.07.017.

MOONJUN, R.; SHRESTHA, D. P.; JETTEN, V. G.; VAN RUITENBEEK, F. J. A. Application of airborne gamma-ray imagery to assist soil survey: A case study from Thailand. **Geoderma**, v. 289, p. 196–212, 2017. doi: 10.1016/j.geoderma.2016.10.035.

MOTTA, P. E. F. da; BARUQUI, A. M.; SANTOS, H. G. DOS. **Levantamento de Reconhecimento de Média Intensidade dos Solos da Região do Alto Paranaíba, Minas Gerais.** 1º ed. Rio de Janeiro: Embrapa Solos, 2004.

OLIVEIRA, V. Á. de; JACOMINE, P. K. T.; COUTO, E. G. Solos do bioma Cerrado. In: CURI, N.; KER, J. C.; NOVAIS, R. F.; VIDAL-TORRADO, P.; SCHAEFER, C. E. G. R. (Ed.). **Pedologia: Solos dos Biomas Brasileiros.** 1º ed. Viçosa, MG: Sociedade Brasileira de Ciência do Solo, 2017. p. 178–226. 2017.

PELEGRINO, M. H. P.; SILVA, S. H. G.; MENEZES, M. D. de; SILVA, E. da; OWENS, P. R.; CURI, N. Mapping soils in two watersheds using legacy data and extrapolation for similar surrounding areas. **Ciência e Agrotecnologia**, v. 40, n. 5, p. 534–546, 2016. doi: 10.1590/1413-70542016405011416.

R CORE TEAM. **R: A Language and environment for statistical computing** Viena, Áustria R Foundation for Estatistical Computing, , 2020. .

R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing.** Viena, Áustria, 2021. .

READ, C. F.; DUNCAN, D. H.; HO, C. Y. C.; WHITE, M.; VESK, P. A. Useful surrogates of soil texture for plant ecologists from airborne gamma-ray detection. **Ecology and Evolution**, v. 8, n. 4, p. 1974–1983, 2018. doi: 10.1002/ece3.3417.

REINHARDT, N.; HERRMANN, L. Gamma-ray spectrometry as versatile tool in soil science: A critical review. **Journal of Plant Nutrition and Soil Science**, v. 182, n. 1, p. 9–27, 2019. doi: 10.1002/jpln.201700447.

RESENDE, M.; CURI, N.; REZENDE, S. A.; CORREA, G. **Pedologia: base para a**

distinção de ambientes. 6. ed. Lavras: Editora UFLA, 2014.

RESENDE, M.; CURI, N.; REZENDE, S. B.; SILVA, S. H. G. **Da rocha ao solo: enfoque ambiental.** 1. ed. Lavras: UFLA, Ed., 2019.

RIBEIRO, J. F.; WALTER, B. M. T. As principais fitofisionomias do bioma Cerrado. In: **Cerrado: Ecologia e Flora.** vol. 1 ed. Brasília: Embrapa Informação Tecnológica, 2008. p. 151–212. 2008.

RIKHVANOV, L. P. Using Radioactive Elements and the Th/U Ratio in Study of the Geochemical Typification of Granitoids and Their Intrusive Character. **Russian Geology and Geophysics**, v. 60, n. 9, p. 1018–1025, 1 set. 2019. doi: 10.15372/RGG2019067.

RODRIGUES, T. E. **Caracterização e gênese de solos Brunos do Maciço Alcalino de Poços de Caldas(MG).** 1984. Universidade de São Paulo - USP, Escola Superior Luíz de Queiróz, 1984.

ROUZE, G. S.; MORGAN, C. L. S.; MCBRATNEY, A. B. Understanding the utility of aerial gamma radiometrics for mapping soil properties through proximal gamma surveys. **Geoderma**, v. 289, p. 185–195, 2017. doi: 10.1016/j.geoderma.2016.12.004.

SANTOS, H. G. dos; ALMEIDA, J. A.; OLIVEIRA, J. B. de; LUMBRERAS, J. F.; ANJOS, L. H. C. dos; COELHO, M. R.; JACOMINE, P. K. T.; CUNHA, T. J. F.; OLIVEIRA, V. Á. de; EMBRAPA. **Sistema brasileiro de classificação de solos.** [s.l: s.n.]

SANTOS, H. G. dos; LUIZ DIAMANTE AGLIO, M.; DE OLIVEIRA DART, R.; DE LOURDES BREFFIN, M. M.; SILVA DE SOUZA, J. Distribuição Espacial dos Níveis de Levantamento de Solos no Brasil. **XXXIV Congresso Brasileiro de Ciencia do Solo**, n. 2, p. 1–4, 2013.

SCHAETZL, R.; ANDERSON, S. **Soils Genesis and Geomorphology.** 1. ed. New York: Cambridge University, 2005.

SHANNON, C. E. Cominunication Theory in the presence of Noise. **Proceedings of the IRE**, v. 37, n. 1, p. 10–21, 1949.

SHARIFIFAR, A.; SARMADIAN, F.; MALONE, B. P.; MINASNY, B. Addressing the issue of digital mapping of soil classes with imbalanced class observations. **Geoderma**, v. 350, n. October 2018, p. 84–92, 2019. doi: 10.1016/j.geoderma.2019.05.016.

SHAW, C. F. What Characteristics Distinguish Pedalfers from Pedocals? **Soil Science Society of America Journal**, v. B16, n. 2001, p. 5, 2010. doi: 10.2136/sssaj1935.036159950b1620010003x.

SILVA, B. P. C.; SILVA, M. L. N.; AVALOS, F. A. P.; DE MENEZES, M. D.; CURI, N. Digital soil mapping including additional point sampling in Posses ecosystem services pilot

watershed, southeastern Brazil. **Scientific Reports**, v. 9, n. 1, p. 1–12, 2019. doi: 10.1038/s41598-019-50376-w.

SILVA, E. da. **Mapeamento de solos e uso de algoritmos de aprendizagem em Lavras (MG)**. 2018. Universidade Federal de Lavras, 2018.

SILVA, S. H. G.; MENEZES, M. D. de; OWENS, P. R.; CURI, N. 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**, v. 267, p. 65–77, 2016. doi: 10.1016/j.geoderma.2015.12.025.

SILVA, V. A.; CURI, N.; MARQUES, J. J. G.; DE CARVALHO, L. M. T.; DOS SANTOS, W. J. R. Mapa de solos, conhecimento de campo, inventário florestal e zoneamento ecológico-econômico como base para a aptidão agrícola das terras em Minas Gerais elaborada em SIG. **Ciencia e Agrotecnologia**, v. 37, n. 6, p. 538–549, 2013. doi: 10.1590/S1413-70542013000600007.

SKORUPA, A. L.; GUILHERME, L. R. G.; CURI, N.; SILVA, C. P. de C.; SCOLFORO, J. R. S.; MELO MARQUES, J. J. G. de S. e M. Propriedades de Solos sob Vegetação Nativa em Minas Gerais: Distribuição por Fitofisionomia. **Revista Brasileira de Ciencia do Solo**, v. 36, n. 3, p. 11–22, 2012.

SMETHURST, M. A. A software tool to rotate spatial data on the surface of the sphere (Earth) for Geosoft's Oasis Montaj. **NGU Report**, 2005.

STAHR, K.; CLEMENS, G.; SCHULER, U.; ERBE, P.; HAERING, V.; CONG, N. D.; BOCK, M.; TUAN, T. D.; HAGEL, H.; VINH, B. L.; RANGUBPIT, W.; SURINKUM, A.; WILLER, J.; INGWERSEN, J.; ZAREI, M.; HERRMANN, L. Beyond the horizons: Challenges and prospects for soil science and soil care in Southeast Asia. In: FRÖHLICH, H. L.; SCHREINEMACHERS, P.; STAHR, K.; CLEMENS, G. (Ed.). **Southeast Asia: Innovations and Policies for Mountainous Areas**. 1. ed. Berlin: Springer, 2013. p. 31–107. 2013.

SURVEY UNITED STATES GEOLOGICAL. **Digital Elevation - Shuttle Radar Topography Mission (SRTM)**.

TAYLOR, K. E.; STOUFFER, R. J.; MEEHL, G. A. An overview of CMIP5 and the experiment design. **Bulletin of the American Meteorological Society**, v. 93, n. 4, p. 485–498, 2012. doi: 10.1175/BAMS-D-11-00094.1.

TEN CATEN, A.; DALMOLIN, R. S. D.; MENDONÇA-SANTOS, M. de L.; GIASSON, E. Mapeamento digital de classes de solos: características da abordagem brasileira. **Ciência Rural**, v. 42, n. 11, p. 1989–1997, 2012. doi: 10.1590/s0103-84782012001100013.

TENG, H. T.; VISCARRA ROSSEL, R. A.; SHI, Z.; BEHRENS, T. Updating a national soil classification with spectroscopic predictions and digital soil mapping. **Catena**, v. 164, n.

October 2017, p. 125–134, 2018. doi: 10.1016/j.catena.2018.01.015.

ULBRICH, H. H. G. J.; GOMES, C. B. Alkaline rocks from continental Brazil. **Earth Science Reviews**, v. 17, n. 1–2, p. 135–154, 1981. doi: 10.1016/0012-8252(81)90009-X.

VASQUES, G. de M.; DART, R. de O.; BACA, J. F. M.; CEDDIA, M. B.; MENDONÇA-; SANTOS, M. de L. **Estoque de Carbono do Brasil (0-30cm)**. 1. ed. Rio de Janeiro: Embrapa Solos, 2017.

VAYSSE, K.; LAGACHERIE, P. Evaluating Digital Soil Mapping approaches for mapping GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France). **Geoderma Regional**, v. 4, p. 20–30, 2015. doi: 10.1016/j.geodrs.2014.11.003.

WADOUX, A. M. J. C.; SAMUEL-ROSA, A.; POGGIO, L.; MULDER, V. L. A note on knowledge discovery and machine learning in digital soil mapping. **European Journal of Soil Science**, v. 71, n. 2, p. 133–136, 2020. doi: 10.1111/ejss.12909.

WILFORD, J. A weathering intensity index for the Australian continent using airborne gamma-ray spectrometry and digital terrain analysis. **Geoderma**, v. 183–184, p. 124–142, 2012. doi: 10.1016/j.geoderma.2010.12.022.

WILFORD, J.; MINTY, B. Chapter 16 The Use of Airborne Gamma-ray Imagery for Mapping Soils and Understanding Landscape Processes. **Developments in Soil Science**, v. 31, n. C, 2006. doi: 10.1016/S0166-2481(06)31016-1.

YANG, L.; JIAO, Y.; FAHMY, S.; ZHU, A.-X.; HANN, S.; BURT, J. E.; QI, F. Updating Conventional Soil Maps through Digital Soil Mapping. **Soil Science Society of America Journal**, v. 75, n. 3, p. 1044, 24 maio 2011. doi: 10.2136/sssaj2010.0002.

SEGUNDA PARTE

ARTIGO: MAPEAMENTO DIGITAL DE SOLOS NO SUL DE MINAS GERAIS: PREDIÇÃO, INCERTEZA E IMPORTÂNCIA DAS COVARIÁVEIS AMBIENTAIS

RESUMO

Técnicas de mapeamento digital de solos (MDS) têm sido empregadas para atender a demanda por mapas de solos de áreas extensas, em escalas mais detalhadas, de forma rápida e menos subjetiva. A principal estratégia que tem sido aplicada é o uso de dados de solos e covariáveis ambientais para o treinamento de modelos de predição com algoritmos de aprendizado de máquina. As covariáveis ambientais representam os fatores de formação do solo e orientam a predição espacial. Entretanto, não há consenso sobre o conjunto de variáveis ambientais que controla os processos pedogenéticos em cada local. Sob a hipótese de que produtos digitais têm potencial para aprimorar o reconhecimento da variabilidade de solos e de contribuir para a geração de mapas mais detalhados, este trabalho utilizou variáveis de terreno derivadas do Modelo Digital de Elevação (MDE), dados de levantamentos aerogeofísicos (espectroscopia de raios gama), modelos climáticos e paleoclimáticos para mapear os solos de uma área de 52.982 km², a região Sul do estado de Minas Gerais. Informações sobre os solos da região que compõem o legado de décadas de pesquisas pedológicas embasaram o mapeamento que adotou uma metodologia de seleção de variáveis ambientais direcionada para predição de classes de solos. Foi gerado um mapa de solos com a identificação de 19 classes com 89% de acurácia, acompanhado das métricas da Entropia de Shannon e do nível de extrapolação dos dados aplicados na forma de mapas de incerteza. As variáveis ambientais mais importantes para a predição foram os dados aerogeofísicos (K e Th), a distância da rede de canais de drenagem derivada do MDE e o paleoclima, representado pela precipitação estimada há 22.000 anos.

Palavras-chave: Pedometria. MDS. Mapeamento de solos. Espectrometria gama. Paleoclima.

1 INTRODUÇÃO

O MDS pode ser abordado segundo uma releitura da função clássica de formação do solo concebida por Dokuchaev e formulada por Jenny (1941) em que os fatores de formação dos solos (clima, organismos, relevo, material de origem, tempo) são representados por covariáveis ambientais na forma de dados digitais, auxiliando a predição de classes ou atributos de solos. A equação dos fatores de formação dos solos foi adaptada ao MDS segundo o modelo SCORPAN (MCBRATNEY; MENDONÇA SANTOS; MINASNY, 2003), que acrescentou coordenadas geográficas ao dados, permitindo a utilização de Sistemas de Informações Geográficas no mapeamento. Essa abordagem permite predizer os padrões de distribuição dos solos na paisagem por meio de métodos matemáticos e estatísticos, quantificando o erro das predições através de métricas de acurácia.

O desenvolvimento de modelos preditivos usa como base o conhecimento pedológico (BUI, 2004) na forma de dados legados (perfis de solos) que orientam o aprendizado de máquinas e a construção de relações entre o solo e as covariáveis ambientais para o processo de aprendizagem de cada algoritmo (DUDA et al., 2017; MANCINI et al., 2019). Entretanto, como a associação espacial é altamente variável entre as paisagens, não há consenso sobre o conjunto de covariáveis ambientais que controla os processos pedogenéticos em cada local (ARROUAYS et al., 2020a) ou o modelo capaz de relacionar as covariáveis à ocorrência das classes de solo.

Dentre as questões metodológicas, um desafio observado na utilização de algoritmos de aprendizado de máquinas para MDS diz respeito à variabilidade natural da distribuição dos solos, que proporciona maior ocorrência de uma ou outra classe, dependendo da combinação entre as variáveis ambientais que ocorre em cada paisagem. Esse fenômeno, inerente ao processo de formação dos solos, gera maior frequência de amostragem de classes majoritárias e tende a induzir os algoritmos a superestimar a ocorrência das classes mais frequentes e subestimar, ou eliminar, classes minoritárias (BRUNGARD et al., 2015; MA et al., 2019; SHARIFIFAR et al., 2019). Nesses casos, a falta de amostras suficientes para um adequado aprendizado de máquina tem resultado em menor acurácia na predição (BRUNGARD et al., 2015; MA et al., 2019; SHARIFIFAR et al., 2019). Diversas estratégias já foram aplicadas para lidar com esta limitação da modelagem: utilizar níveis categóricos taxonômicos menos elevados ou agrupar em uma mesma unidade de mapeamento mais de uma unidade taxonômica, reduzindo o detalhamento do mapa (JAFARI et al., 2013; SILVA et al., 2019); aumentar o

número de amostras das classes minoritárias, realizando um balanceamento artificial da representação da paisagem (MA et al., 2019) ou; aumentar a amostragem considerando a área ao redor dos perfis de solo (PELEGRINO et al., 2016; SILVA et al., 2016; MACHADO et al., 2019b). A escolha metodológica depende dos objetivos do mapeamento e das características da área.

Quanto à escolha do modelo para relacionar solos e paisagem, tem sido frequentemente observado que modelos preditivos mais complexos como o *Random Forest*, tem apresentado maior acurácia quando comparado a modelos lineares generalizados (BRUNGARD et al., 2015; HEUNG et al., 2016; CAMERA et al., 2017; SILVA et al., 2019; MENEZES et al., 2020). No MDS, o *Random Forest* tem apresentado desempenho notável em diversos trabalhos (PELEGRINO et al., 2016; HENGL et al., 2017; MACHADO et al., 2019b; MANCINI et al., 2019; SILVA et al., 2019; KHALEDIAN; MILLER, 2020). Trata-se de um modelo não paramétrico orientado por dados. Seus recursos de *bootstrapping* e *bagging* são eficientes em revelar padrões, principalmente quando se trata de associações espaciais complexas (HEUNG et al., 2016; SILVA et al., 2019). Dentre as vantagens da modelagem de solos com uso do *Random Forest* está a sua capacidade de gerar previsões acuradas com base em um grande número de covariáveis ambientais (CHAKRABORTY et al., 2017; HENGL et al., 2017; KHALEDIAN; MILLER, 2020; MENEZES et al., 2020) e reportar a importância relativa das variáveis utilizadas na previsão. Essa característica permite novos *insights* a respeito da relação solo-paisagem, com base em covariáveis ambientais nunca antes utilizadas no mapeamentos de solos em moldes mais tradicionais (ARROUAYS et al., 2020a; KHALEDIAN; MILLER, 2020; MENEZES et al., 2020). A representação dos fatores de formação dos solos a partir de novos produtos digitais associados a variáveis já consagradas no MDS têm o potencial de contribuir para a geração de mapas de solos de maneira mais acurada e para melhorar a compreensão pedogenética da paisagem (ARROUAYS et al., 2020b; WADOUX et al., 2020).

Os dados digitais obtidos a partir da aeroespectrometria de raios gama, gerados inicialmente para exploração e mapeamento geológico mede a concentração dos radioelementos potássio (K), tório (Th) e urânio (U) na superfície da Terra, fornecendo informações sobre a distribuição espacial do material de origem dos solos (COOK et al., 1996; ELLILI et al., 2019; KIDD et al., 2020) e uma alternativa quantitativa aos mapas geológicos tradicionais como covariáveis ambientais. Diversos estudos identificaram que os processos pedogenéticos como adição (CATTLE et al., 2003), remoção (WILFORD; MINTY, 2006), translocação (STHR et al., 2013) e transformação (WILFORD, 2012) têm efeito na diluição

ou acúmulo desses radioelementos, resultando na diferenciação no sinal de raios gama e em uma preciosa informação ao mapeamento de solos (ROUZE; MORGAN; MCBRATNEY, 2017; MA et al., 2019). Existe no Brasil uma grande cobertura de levantamentos de espectrometria de raios gama aerotransportados, abrangendo 43% do território (CPRM, 2020). Esses dados, ainda não foram utilizados no mapeamento de solos (COELHO et al., 2021), e apresentam grande potencial para contribuir com o estudo dos solos tropicais.

Outros produtos digitais que estão disponíveis para todo o globo na plataforma *Worldclim – World Climate Data* (FICK; HIJMANS, 2017), são modelos paleoclimáticos que propõem estimativas para a temperatura e precipitação simulando condições pretéritas de 120.000, 22.000 e 6.000 anos (TAYLOR; STOUFFER; MEEHL, 2012). Esses dados nunca foram utilizados com a finalidade de prever a ocorrência ou mapear solos. Considerando a natureza poligenética dos solos tropicais e sendo o processo de formação do solo multivariado (WILFORD, 2012; RESENDE et al., 2014, 2019), as condições atuais e pretéritas, uma vez atuantes, são consideradas potenciais covariáveis ambientais.

Dentre as diversas opções de dados digitais para o MDS, os modelos digitais de elevação (MDE) são os produtos mais frequentemente utilizados. Trabalhados com algoritmos de cálculo do terreno, oferecem opções de covariáveis ambientais para a representação da paisagem segundo características fundamentais para a formação dos solos, tais como: altitude, declividade, índices de umidade, textura do terreno, redes de drenagem, níveis de insolação e transporte de sedimentos, dentre outros (MCBRATNEY; MENDONÇA SANTOS; MINASNY, 2003; KIDD et al., 2020; COELHO et al., 2021).

A predição dos estoques de carbono no solo para o território brasileiro está disponível para as profundidades de 0-30 cm e 0-100 cm em dois produtos digitais (VASQUES et al., 2017; GOMES et al., 2019) que ainda não foram explorados no MDS. A conservação da matéria orgânica no solo envolve questões ligadas à fertilidade, drenagem, aeração, estrutura, textura, conservação do horizonte superficial e temperatura, fornecendo referências para a diferenciação dos tipos de solo.

Algumas covariáveis condensam informações relativas a mais de um fator de formação. Por exemplo, o mapa de estoque de carbono reflete a atuação dos organismos no solo e as condições edafoclimáticas pretéritas que possibilitaram a conservação da matéria orgânica. As variáveis derivadas do MDE traduzem aspectos do relevo que refletem o tempo de exposição aos agentes do intemperismo; as variáveis paleoclimáticas representam o clima e o tempo; os dados espectrométricos estão relacionados ao material de origem dos solos, ao clima e ao

tempo. Considerando a representatividade dos dados digitais reunidos para a representação dos fatores de formação dos solos em condições tropicais, o objetivo deste trabalho foi, a partir desse conjunto de covariáveis ambientais e da modelagem preditiva com o algoritmo *Random Forest*, gerar um mapa de solos para uma área com extensão de 52.982 km². Informações sobre os solos da região que compõem o legado de décadas de pesquisa pedológica embasaram o mapeamento. As hipóteses levantadas foram: i) a aplicação de covariáveis ambientais, inéditas ou pouco exploradas no mapeamento de solos no Brasil, associada a uma metodologia de seleção de variáveis direcionada especificamente às classes de solo mapeadas, resultam em um produto mais detalhado do que o disponível; ii) a aplicação de metodologia para seleção das covariáveis mais importantes possibilita melhor interpretabilidade dos modelos de predição; iii) as estimativas espaciais de avaliação da incerteza da predição indicam as maiores incertezas onde a densidade de amostragem é menor; iv) as covariáveis relativas a modelos paleoclimáticos apresentam maior importância para a predição do que modelos climáticos atuais; v) os dados provenientes da espectrometria gama figurarão entre as principais covariáveis ambientais para a predição.

2 MATERIAL E MÉTODOS

2.1 Área de estudo

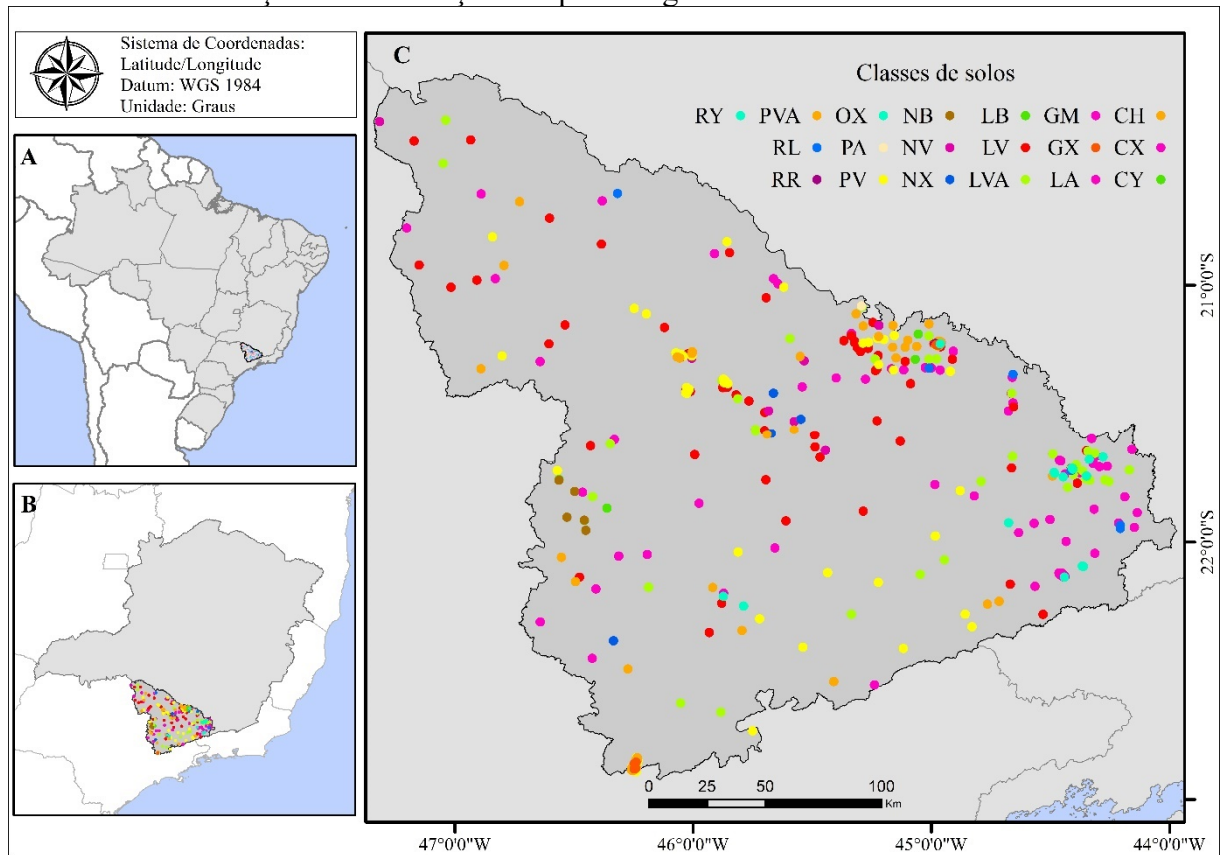
O Sul de Minas Gerais está localizado na região sudeste do Brasil., com área de 52.982 km², situado entre as coordenadas de latitude 20°10'S/22°55'S e longitude 47°18'W/47°59'7W (FIGURA 1), *datum* WGS84. A região possui grande notoriedade na produção de cafés, sendo responsável pela maior parte das exportações nacionais do produto (CONAB, 2019). Os recursos hídricos são outro destaque da região que concentra, nas partes mais altas do relevo, as nascentes responsáveis por abastecer os reservatórios de grandes usinas hidrelétricas que fornecem água e energia à região mais populosa do país (ANA, 2017). As porções de relevo movimentado expressam eventos geológicos ocorridos há mais de 500 milhões de anos, que contribuíram para a composição do quadro geológico de alta complexidade (ALKMIM, 2020), composto por rochas ígneas, metamórficas e sedimentares com diferentes mineralogias. O clima é profundamente influenciado pelas formações do relevo, apresentando as variações Cwa e Cwb, de acordo com o sistema de classificação climática Köppen & Geiger (EITEN, 1994).

A continentalidade também influencia as características climáticas, determinando uma estação seca mais marcante em direção à Oeste.

2.2 Conjunto de dados

O banco de dados de solos foi construído com pesquisa exploratória em bases físicas e digitais contendo classificações taxonômicas de perfis de solo. Foram obtidos dados legados de solos a partir das seguintes fontes: FeBR - Repositório Brasileiro Livre para Dados Abertos do Solo (BRASIL, 1983; CARVALHO et al., 1983a; MOTTA; BARUQUI; SANTOS, 2004), IDE-Sisema - Infraestrutura de dados espaciais do Sistema estadual de meio ambiente e recursos hídricos (CAIRES; FONTES; KER, 2007; FERNANDES; MARQUES, 2007), Banco de Dados de Teses e Dissertações da Universidade Federal de Lavras (MENEZES, 2007; BERTOLDO, 2008; MACHADO, 2017; SILVA, 2018; SILVA et al., 2019; GONÇALVES, 2020) e da Universidade de São Paulo (RODRIGUES, 1984). Após a coleta das informações sobre os solos, foi realizada a harmonização das informações georreferenciadas e os dados foram agrupados em um único arquivo, que serviu de base para a modelagem preditiva e validação. Ao todo, foram reunidas 559 descrições referentes a 19 tipos de solos (FIGURA 1), classificados até o segundo nível categórico segundo o Sistema Brasileiro de Classificação de Solos (SANTOS et al., 2018).

Figura 1- A) Brasil, B) Estado de Minas Gerais, C) Região Sul de Minas Gerais com a localização e classificação dos perfis legados de solos.



Legenda: CH: Cambissolo Húmico, CX: Cambissolo Háptico, CY: Cambissolo Flúvico, Gm: Gleissolo Melânico, GX: Gleissolo Háptico, LA: Latossolo Amarelo, LB: Latossolo Bruno, LV: Latossolo Vermelho, LVA: Latossolo Vermelho-Amarelo, NB: Nitossolo Bruno, NV: Nitossolo Vermelho, NX: Nitossolo Háptico, OX: Organossolo Háptico, PA: Argissolo Amarelo, PV: Argissolo Vermelho, PVA: Argissolo Vermelho-Amarelo, RL: Neossolo Litólico, RR: Neossolo Regolítico, RY: Neossolo Flúvico.

2.2.1 Estratégia para lidar com a base de dados desbalanceada

A variabilidade natural da distribuição dos solos proporciona maior ocorrência de uma ou mais classes de solos em detrimento a outras em uma dada região, dependendo da combinação entre as covariáveis ambientais e os processos pedogenéticos. Esse fenômeno inerente ao processo de formação dos solos tem influência direta na criação de modelos preditivos, induzindo os algoritmos de aprendizado de máquinas a subestimar ou não prever as classes de solos em menor proporção (BRUNGARD et al., 2015; MA et al., 2019; SHARIFIFAR et al., 2019). Para aumentar a base de treinamento dos modelos e fornecer amostras suficientes para o adequado aprendizado de máquina, a área ao redor dos pontos classificados foi trabalhada em SIG e utilizada para captar melhor a variabilidade da paisagem

(PELEGRINO et al., 2016; SILVA et al., 2016). Os pontos georreferenciados contendo a classificação dos perfis de solo foram utilizados como centro para delimitação de um círculo com raio de 30 m. A área do círculo foi convertida para o formato raster com o mesmo tamanho de pixel e alinhamento das covariáveis ambientais após processamento e harmonização. No centro de cada pixel foi extraído um ponto, que recebeu a mesma classificação de solo da coordenada original (TABELA 1). Esse procedimento foi baseado no princípio de levantamentos pedológicos tradicionais (EMBRAPA, 1995; IBGE, 2007), segundo o qual são analisados perfis de referência para a classificação dos solos de uma determinada porção da paisagem. A distância adotada seguiu o conceito de frequência de Nyquist, não ultrapassando a metade do espaçamento médio entre os pares de pontos mais próximos (SHANNON, 1949, HENGL, 2006).

Tabela 1- Caracterização do banco de dados legado de solos: classificação dos solos, número de amostras original e número de amostras considerando a área ao redor do perfil de solo.

Classes de solos	Legenda	Nº de amostras do legado	Nº de amostras extraídas da área ao redor do perfil
Cambissolo Húmico	CH	5	16
Cambissolo Háptico	CX	112	348
Cambissolo Flúvico	CY	3	6
Gleissolo Melânico	GM	3	10
Gleissolo Háptico	GX	8	23
Latossolo Amarelo	LA	8	30
Latossolo Bruno	LB	4	14
Latossolo Vermelho	LV	120	372
Latossolo Vermelho-Amarelo	LVA	82	255
Nitossolo Bruno	NB	5	14
Nitossolo Vermelho	NV	15	49
Nitossolo Háptico	NX	5	16
Organossolo Háptico	OX	3	7
Argissolo Amarelo	PA	17	50
Argissolo Vermelho	PV	82	259
Argissolo Vermelho-Amarelo	PVA	65	198
Neossolo Litólico	RL	7	23
Neossolo Regolítico	RR	1	4
Neossolo Flúvico	RY	14	45

2.3 Mapeamento Digital de Solos (MDS)

2.3.1 Covariáveis ambientais SCORPAN

Para representar a covariável ambiental clima, foram utilizados os dados de médias anuais de temperatura e precipitação disponibilizados na plataforma *Worldclim – World Climate Data*, referentes ao clima atual (FICK; HIJMANS, 2017) e a climas pretéritos (Paleoclima), com estimativas para 6.000, 22.000 e 120.000 anos (TAYLOR; STOUFFER; MEEHL, 2012). Justifica-se o uso de dados paleoclimáticos pelo fato da pedogênese estar frequentemente associada a diferentes circunstâncias climáticas, uma vez que longos períodos são necessários para a formação do solo, configurando a natureza poligenética dos solos (INDA JUNIOR; KÄMPF, 2005).

O clima está fortemente relacionado com outra covariável ambiental, o estoque de carbono. Climas úmidos proporcionam maior produção vegetal e temperaturas baixas contribuem para a conservação da matéria orgânica no subsolo. O estoque de carbono está relacionado a questões que envolvem a atuação dos organismos e do clima e refletem atributos do solo como fertilidade, drenagem, aeração, estrutura, textura e conservação do horizonte superficial, dentre outras. Por essas inter-relações, mapas de estoque de carbono no solo (VASQUES et al., 2017; GOMES et al., 2019) foram utilizados como covariáveis ambientais representando os fatores clima, solo e organismos.

As covariáveis relativas ao relevo foram derivadas do MDE com 30 m de resolução espacial, desenvolvido pela *Shuttle Radar Topography Mission* (SRTM) (SURVEY UNITED STATES GEOLOGICAL, 2019). As cenas referentes à área de estudos foram pré-processadas para preenchimento de falhas e para tornar o modelo hidrologicamente consistente. Deste produto, foram derivados 17 atributos de terreno (TABELA 2) utilizando o software SAGA GIS (CONRAD et al., 2015). Esses atributos são comumente utilizados em mapeamento de solos (MCBRATNEY; MENDONÇA SANTOS; MINASNY, 2003; COELHO et al., 2021) por refletirem aspectos relacionados ao clima, a distribuição da umidade, ao grau de intemperismo, a rugosidade do terreno, ao material de origem dos solos e sua granulometria, dentre outros aspectos.

Para a representação do material de origem foram utilizados dados aéreos de gamaespectrometria obtidos a partir de levantamentos aerogeofísicos coordenados pelo Serviço Geológico do Brasil (CPRM) e Companhia de Desenvolvimento Econômico de Minas Gerais (CODEMIG) entre os anos de 2001 e 2013, relativos à detecção de radionuclídeos de U, Th e

K nos primeiros 30 cm da superfície do solo. Os dados foram submetidos a processamentos padrão para levantamentos aerogeofísicos utilizando o software OASIS MONTAJ 9.7 (SMETHURST, 2005). O conjunto de dados foi interpolado utilizando curvatura mínima em grades regulares de 100 m (1/4 a 1/5 do espaçamento das linhas de vôo) gerando as grades de concentrações de K (%), Th (ppm) e U (ppm). Essas métricas foram baseadas em relatórios técnicos comumente usados em levantamentos aerogeofísicos (IAEA, 2003). As grades resultantes foram fixadas usando a ferramenta *grid knitting*. Valores negativos, identificados majoritariamente nas bordas, foram substituídos por zero e novamente interpolados. Posteriormente, o fluxo total de raios gama (dose) foi calculado com base nas adições ponderadas dos três elementos radioativos usando a fórmula $Dose = 13,078 K (\%) + 5,67 U (ppm) + 2,49 Th (ppm)$ (WILFORD, 2012). O cálculo permite a comparação quantitativa de assinaturas radiométricas, refletindo materiais de rocha e regolito (WILFORD, 2012). Outro indicador da gênese de rochas utilizado foi a razão U/Th (RIKHVANOV, 2019).

Após o processamento de todas as covariáveis ambientais (TABELA 2), os arquivos foram convertidos para o formato raster e harmonizados para uma grade comum de 30 m X 30 m utilizando o software ArcGis (ESRI, 2013).

Tabela 2- Covariáveis ambientais (continua).

Covariáveis ambientais	Resolução	Fonte
Temperatura média anual	1 km	Xu; Hutchinson (2013)
Precipitação total anual	1 km	
Temperatura média anual 6.000 anos (BC)	1 km	Taylor; Stouffer; Meehl (2012)
Temperatura média anual 6.000 anos (CC)	1 km	
Temperatura média anual 6.000 anos (CN)	1 km	
Temperatura média anual 6.000 anos (HE)	1 km	
Temperatura média anual 6.000 anos (HG)	1 km	
Temperatura média anual 6.000 anos (IP)	1 km	
Temperatura média anual 6.000 anos (ME)	1 km	
Temperatura média anual 6.000 anos (MG)	1 km	
Temperatura média anual 6.000 anos (MR)	1 km	
Temperatura média anual 22.000 anos (CC)	5 km	
Temperatura média anual 22.000 anos (ME)	5 km	
Temperatura média anual 22.000 anos (MR)	5 km	
Temperatura média anual 120.000 anos (BP)	1 km	Otto-Bliesner et al. (2008)
Precipitação total anual 6.000 anos (BC)	1 km	Taylor; Stouffer; Meehl (2012)
Precipitação total anual 6.000 anos (CC)	1 km	
Precipitação total anual 6.000 anos (CN)	1 km	

Tabela 2: Covariáveis ambientais (conclusão).

Covariáveis ambientais	Resolução	Fonte
Precipitação total anual 6.000 anos (HE)	1 km	Taylor; Stouffer; Meehl (2012)
Precipitação total anual 6.000 anos (HG)	1 km	
Precipitação total anual 6.000 anos (IP)	1 km	
Precipitação total anual 6.000 anos (ME)	1 km	
Precipitação total anual 6.000 anos (MG)	1 km	
Precipitação total anual 6.000 anos (MR)	1 km	
Precipitação total anual 22.000 anos (CC)	5 km	
Precipitação total anual 22.000 anos (ME)	5 km	
Precipitação total anual 22.000 anos (MR)	5 km	
Precipitação total anual 120.000 anos (BP)	1 km	Otto-Bliesner et al. (2008)
Matéria orgânica do solo 0-30cm (mín)	1 km	Vasques et al. (2017)
Matéria orgânica do solo 0-30cm (máx)	1 km	
Matéria orgânica do solo 0-30cm (méd)	1 km	
Matéria orgânica do solo 0-30cm (dp)	1 km	
Matéria orgânica do solo 0-1m	1 km	Gomes et al. (2019)
Modelo de elevação SRTM (MDE)	30 m	Survey United States Geological (2019)
Aspecto morfométrico do terreno	30 m	Conrad (2001)
Declividade	30 m	
Nível base da rede de drenagem	30 m	Conrad (2002)
Distância vertical da rede de drenagem	30 m	
Fator LS (comp. de rampa * declividade)	30 m	Conrad (2003)
Índice topográfico de umidade	30 m	
Índice de nivelamento do topo do vale	30 m	Conrad (2006)
Índice de nivelamento do fundo do vale	30 m	
Índice de rugosidade do terreno	30 m	Conrad (2010)
Profundidade do vale	30 m	Conrad (2012)
Abertura negativa	30 m	
Abertura positiva	30 m	
Textura do terreno	30 m	
Declividade X área de contribuição	30 m	
Posição relativa no declive	30 m	J.Boehner, O.Conrad (2008)
Gradiente de concentração de fluxo	30 m	R.Heinrich, O.Conrad (2007)
U	100 m	CPRM
Th	100 m	
K	100 m	
U/Th	100 m	
Dose	100 m	

Legenda: Mín.= mínima, máx.= máxima, méd.= média, dp = desvio padrão, comp.=comprimento.

2.3.2 Predição espacial de solos e validação

A predição espacial da ocorrência dos solos foi realizada utilizando o algoritmo *Random Forest* (BREIMAN, 2001) no software R (R DEVELOPMENT CORE TEAM, 2021). O *Random Forest* é um algoritmo do tipo árvore de classificação e regressão usado para prever uma variável alvo em locais onde ela é desconhecida, utilizando como base covariáveis previamente definidas. Para estabelecer as relações entre a variável alvo e as covariáveis, são construídas diversas árvores de classificação. Cada alvo é, então, classificado pelo conjunto de árvores (*Forest*), escolhendo a classe que recebeu o número máximo de votos. Para cada árvore, uma amostra dos dados é usada para construir a árvore e a outra parte constitui a amostra *out-of-bag* (OOB), usada para validação. A comparação entre as amostras OOB e as predições constituem o erro OOB, usado quando um conjunto de dados de validação independente não está disponível (HEUNG; BULMER; SCHMIDT, 2014). A Acurácia OOB (1- erro OOB) é semelhante à validação cruzada *k-fold*, cada árvore do modelo é construída usando dois terços dos dados e validada usando o terço restante. Esta resposta fornece valores comparáveis, desde que observado o mesmo conjunto de dados (FRIEDMAN; HASTIE; TIBSHIRANI, 2008; LAHJOUJ et al., 2020; LEENAARS et al., 2020)

A validação independente foi utilizada para a avaliação do poder preditivo da modelagem em relação a outros trabalhos de mesma natureza (KHALEDIAN; MILLER, 2020). Para isto, as amostras de cada uma das 19 classes de solo foram separadas em dois conjuntos: 70% para treinamento e 30% para validação. A partir das amostras de validação, foram calculadas métricas de acurácia comumente utilizadas em trabalhos de mapeamento digital de solos obtidos a partir da construção da matriz de confusão: acurácia global, acurácia do usuário e acurácia do produtor.

A acurácia global representa o percentual de amostras cujas classes de solo identificadas no campo correspondem às classes de solo apresentadas no mapa de predição. Os índices de acurácia do produtor e do usuário foram calculados para verificar o desempenho do modelo segundo cada classe mapeada. A acurácia do usuário (Eq. 1) demonstra a probabilidade de um local com determinada classificação no mapa corresponder à mesma classificação no campo. A acurácia do produtor (Eq. 2) diz respeito à probabilidade de um local com determinada classe de solo ser classificado corretamente no mapa (CONGALTON; GREEN, 2019).

$$A. \text{ do usuário} = \frac{X_{ii}}{\sum_{i=1}^r X_{ij}} \times 100 \quad (1) \quad A. \text{ do produtor} = \frac{X_{jj}}{\sum_{j=1}^r X_{ij}} \times 100 \quad (2)$$

Onde: A. do usuário = Acurácia do usuário, A. do produtor = Acurácia do produtor. X_{ii} e X_{jj} indicam o número de amostras classificadas corretamente. X_{ij} representa a soma de amostras incluídas na classificação de uma determinada classe no mapa (acurácia do usuário) ou a soma de amostras de validação de uma classe de solo no campo (acurácia do produtor).

O índice Kappa (Eq. 3) (COHEN, 1960) fornece o nível de acurácia para uma classificação corrigida ao acaso. É particularmente importante ao lidar com dados de classe desequilibrados (HOUNKPATIN et al., 2018) e foi calculado para cada um dos modelos selecionados.

$$\text{Índice Kappa} = \frac{P_o - P_a}{1 - P_a} \times 100 \quad (3)$$

Onde: P_o é a proporção de amostras classificadas corretamente e P_a é a probabilidade de concordância aleatória. Os resultados variam de -100 a 100 e indicam uma precisão crescente à medida que os valores se aproximam de 100.

Além das medidas de acurácia, o *Random Forest* fornece a proporção de votos que cada classe obteve para todos os locais analisados. O conjunto de respostas para cada classe de solo modelada permite a construção de mapas de probabilidade de ocorrência (BREIMAN, 2001), utilizados neste trabalho para o cálculo da entropia, conforme será apresentado a seguir.

Outra característica que melhora a interpretabilidade do algoritmo *Random Forest* é a indicação da importância relativa das covariáveis ambientais empregadas para a predição. Esta métrica é calculada a partir do aumento do erro OOB após uma variável preditora ser removida enquanto todas as outras variáveis são mantidas constantes. A diferença entre o erro para o conjunto completo e para o conjunto completo menos cada uma das variáveis é calculada para cada árvore de decisão, à medida que a floresta aleatória é construída. Quanto maior o valor da diferença, mais importante é a covariável ambiental na predição dos solos. Esta medida, chamada de decréscimo médio em acurácia (MDA), foi calculada para o modelo como um todo e para cada uma das classes de solo mapeadas.

2.3.3 Parametrização do algoritmo *Random Forest* e seleção de covariáveis ambientais

Na etapa de parametrização da *Random Forest* foi definido o número de árvores (*ntree*), o número de divisões para se chegar a uma classificação (*nt*) e o número de variáveis amostradas aleatoriamente como candidatas em cada divisão (*mtry*). Foram testados *ntree*=500, 1000, 5000; *nt* = 1, 2, 3; e *mtry* = 4, 7, 14. Os melhores valores foram selecionados segundo a acurácia global. Após o ajuste dos parâmetros, a modelagem e a seleção de covariáveis ambientais seguiu três etapas diferentes e sequenciais. Em todas as etapas foram utilizados os mesmos conjuntos de amostras de treinamento e de validação externa.

- Etapa 1 (RF1): as 55 covariáveis ambientais (Tabela 2) foram utilizadas na modelagem com o *Random Forest*. Os resultados de MDA gerados a partir de 100 repetições do algoritmo serviram de base para a seleção das covariáveis ambientais mais importantes, com $MDA > 20$, que foram empregadas na etapa posterior.

- Etapa 2 (RF2): novamente 100 modelos foram gerados a partir de 100 repetições do *Random Forest*. Nesta etapa, as covariáveis ambientais mais importantes foram selecionadas a partir do modelo de maior acurácia global. As três covariáveis ambientais com maiores valores de MDA foram selecionadas para cada uma das 19 classes de solo, aplicou-se o teste de correlação (Spearman $< 80\%$) e o conjunto de covariáveis seguiu para a terceira etapa da modelagem.

- Etapa 3 (RF3): Após 100 repetições do algoritmo *Random Forest*, o modelo de maior acurácia global foi selecionado para calcular a probabilidade espacial de ocorrência das classes de solo, para confecção do mapa de solos e para avaliação das covariáveis ambientais mais importantes para a predição.

Essas etapas objetivaram reduzir o número de covariáveis redundantes, facilitar a interpretação do modelo final e selecionar o conjunto de preditores que melhor representa a relação solo-paisagem da região de estudo. As classificações iterativas são necessárias para melhorar o diagnóstico e o desempenho da classificação (MILLARD; RICHARDSON, 2015).

2.4 Quantificação da incerteza da predição espacial

2.4.1 Avaliação da extrapolação espacial a partir da Superfície de Similaridade Ambiental Multivariada (SSAM)

Um indicativo importante para avaliar e interpretar a qualidade de mapas de predição é a representação espacial do nível de extrapolação do modelo em relação ao padrão de amostragem. Para essa análise, foi calculada a SSAM, proposta por Elith et al. (2010) para a

modelagem da distribuição de espécies e, mais recentemente, utilizado no mapeamento digital de solos (CAMERA et al., 2017; SILVA et al., 2019). A superfície SAM é produzida aplicando a Eq. 4 a cada pixel na saída do modelo de predição:

$$P(V_i) = \left\{ \begin{array}{l} \frac{(p_i - \min_i)}{(\max_i - \min_i)} \times 100; \text{ se } f_i = 0 \\ 2 \times f_i; \text{ se } 0 < f_i \leq 50 \\ 2 \times (100 - f_i); \text{ se } 50 \leq f_i < 100 \\ \frac{(\max_i - p_i)}{(\max_i - \min_i)} \times 100; \text{ se } f_i = 100 \end{array} \right\} \quad (4)$$

Onde: $P(V_i)$ = Superfície de similaridade ambiental multivariada de um ponto; p_i = valor da covariável ambiental V_i no ponto P; \min_i = valor mínimo da covariável ambiental V_i sobre o conjunto de pontos de referência, \max_i = valor máximo da variável V_i sobre o conjunto de pontos de referência; f_i = porcentagem dos pontos de referência cujo valor da covariável V_i é menor que p_i . O valor da SSAM em P é o mínimo de sua similaridade com relação a cada covariável.

O cálculo indica a semelhança de um local em relação ao conjunto de locais de referência (amostras). Os valores da SSAM são maiores onde o conjunto de covariáveis ambientais é melhor representado pelos perfis de solos. A equação permite valores negativos como resultado, indicativos dos locais onde a previsão precisou extrapolar valores para, pelo menos, uma das covariáveis ambientais preditoras. O conjunto de amostras composto exclusivamente pelo legado foi utilizado para a construção da SSAM, assumindo que o processo de criação de dados de treinamento, com a inclusão dos pontos criados ao redor da coordenada original, pode ser considerada uma primeira extrapolação. A SSAM foi confeccionada com base nas covariáveis ambientais preditoras utilizadas em cada uma das três etapas de seleção utilizando o software R (R CORE TEAM, 2020) e o pacote Dismo (HIJMANS et al., 2020).

2.4.2 Entropia

A quantificação do número de votos durante a modelagem com a *Random Forest* foi utilizada para a confecção dos mapas de probabilidade de ocorrência dos solos para cada classe mapeada (BREIMAN, 2001). Os resultados foram utilizados para calcular a entropia de Shannon normalizada (Eq. 5) como medida da incerteza da predição espacial (MACHADO et

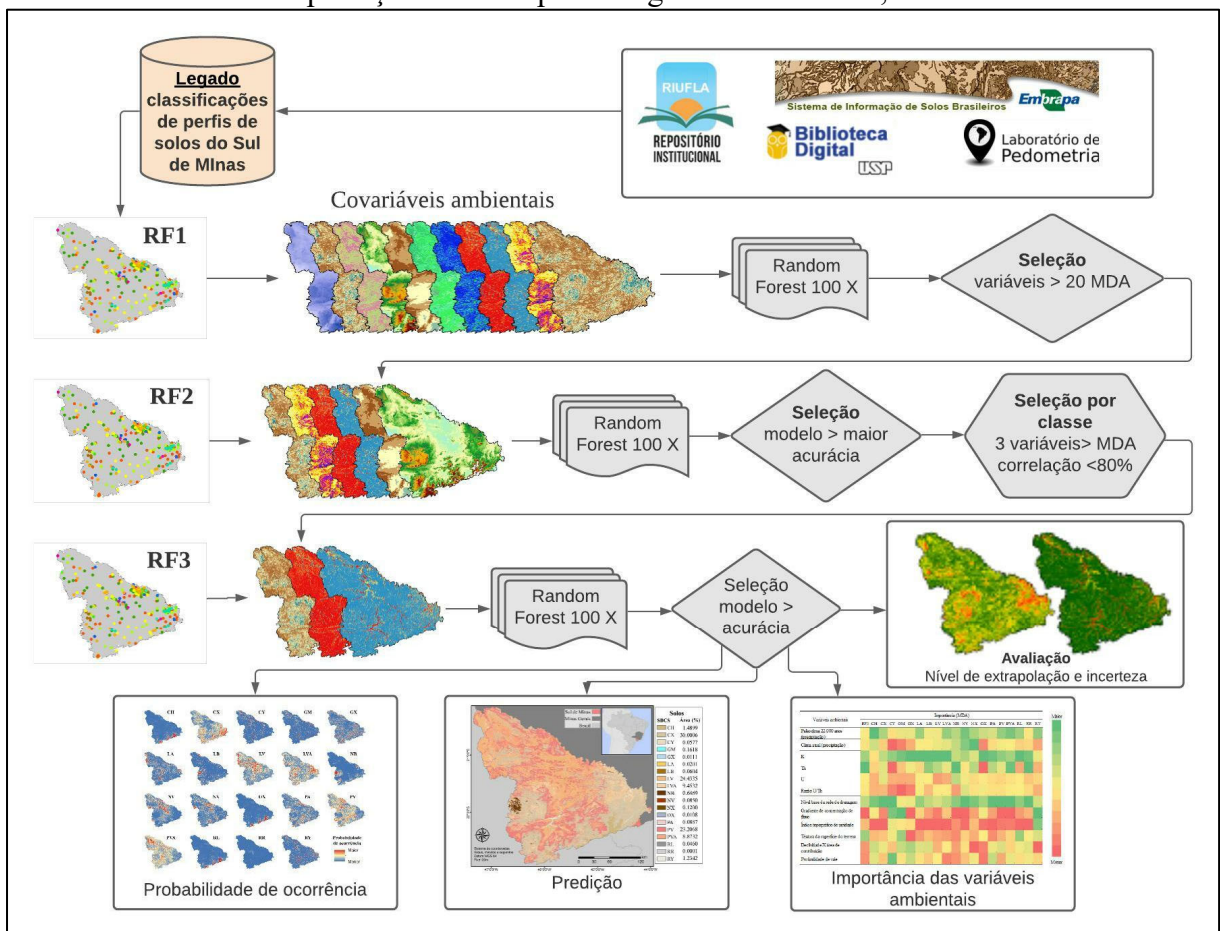
al., 2019a). Os valores de entropia variam de 0 a 1, onde 1 significa incerteza máxima, ou seja, quando o modelo apresenta muito baixa confiabilidade sobre a classe prevista e 0 quando existe confiança máxima sobre a classe predita. Para o cálculo foi utilizado o software R (R DEVELOPMENT CORE TEAM, 2021) e o pacote Raster (ETTEN et al., 2021).

$$Entropia = \frac{\sum_{k=1}^n \frac{S_k}{\sum_{k=1}^n S_k} \times \ln \frac{S_k}{\sum_{k=1}^n S_k}}{\ln(n)} \quad (5)$$

Onde: S_k é o valor da probabilidade de ocorrência de uma determinada classe de solo em um determinado local e n é o número total de classes de solo.

As etapas sequenciais da modelagem e da confecção dos produtos derivados da predição espacial são apresentados na Figura 2.

Figura 2- Fluxograma das etapas sequenciais da modelagem e da confecção dos produtos derivados da predição de solos para a região Sul de Minas, Brasil.



Legenda: RF1, RF2 e RF3: Etapas sequenciais de seleção de variáveis. MDA: métrica de avaliação da importância relativa das covariáveis ambientais.

3 RESULTADOS E DISCUSSÃO

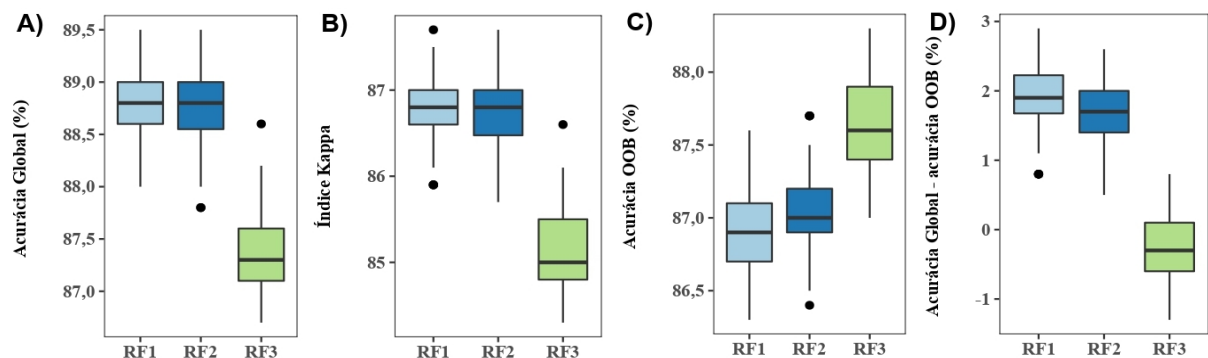
3.1 Parametrização do algoritmo *Random Forest*

Após os testes para ajuste dos parâmetros, $nt=2$, $mtry$ = igual à raiz quadrada do número de covariáveis ambientais predictoras e $n\text{tree}= 1000$ alcançaram as maiores acurácias globais e foram utilizados para as três etapas de modelagem.

3.2 Acurácia dos modelos preditivos e seleção de covariáveis ambientais

Os valores de acurácia OOB (1 - erro OOB), acurácia global e índice Kappa em cada uma das três etapas de seleção de variáveis (RF1, RF2 e RF3) foram superiores a 85% (FIGURA 3), denotando adequada estabilidade e elevada acurácia dos modelos preditivos. Os valores de acurácia global e índice Kappa nas três etapas foram superiores às médias observadas por Coelho et al. (2021) ao analisar trabalhos de mapeamento digital de solos realizados no Brasil, onde foram observadas médias de 0,56 de acurácia global e 62 de índice Kappa.

Figura 3- Métricas de acurácia após 100 repetições da modelagem com *Random Forest*.



Legenda: OOB: *Out of Bag*. Acurácia global – OOB: Subtração da acurácia global menos a acurácia OOB, calculada como (1- erro OOB). RF1, RF2 e RF3: etapas sequenciais da modelagem

No tocante à seleção das covariáveis ambientais, quando foram retiradas as covariáveis predictoras menos importantes ($MDA < 20$) na etapa RF2, foi quase negligível a diferença entre RF1 e RF2 para a acurácia global e índice Kappa. No entanto, na etapa RF3 os valores de acurácia global e índice Kappa foram menores. Ressalta-se que a seleção de covariáveis ambientais para a etapa RF3 foi baseada no MDA específico de cada classe de solo,

diferentemente de RF1 e RF2, quando o MDA utilizado para seleção foi referente à importância das covariáveis para o modelo. Essa opção teve por objetivo melhorar a acurácia de predição para as classes de solos minoritárias, mesmo que em detrimento da acurácia global.

Comparando os valores absolutos, o índice Kappa foi ligeiramente menor que a acurácia global em todas as etapas da modelagem. Esse resultado corrobora com a sensibilidade do índice para avaliar a acurácia em bases de dados desbalanceadas (HOUNKPATIN et al., 2018). No entanto, apesar da pequena diferença, os valores são altos, reforçando as métricas adequadas obtidas pelos modelos.

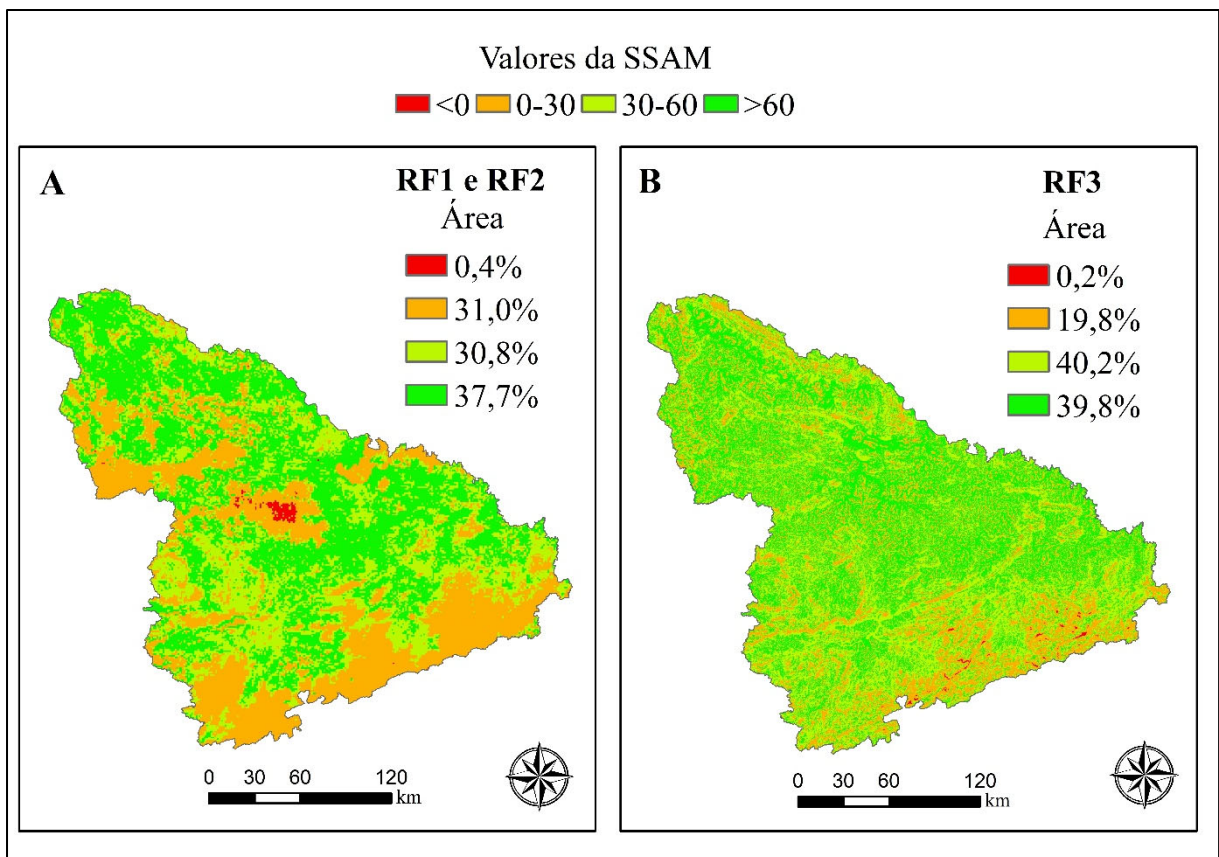
Em relação à acurácia OOB, o comportamento foi inverso ao observado para a acurácia global e índice Kappa, com valores crescentes, conforme o avanço das etapas de seleção. Alguns trabalhos ressaltam a importância do uso da acurácia OOB para avaliação de modelos, desde que analisados o mesmo conjunto de dados (FRIEDMAN; HASTIE; TIBSHIRANI, 2008; LAHJOUJ et al., 2020; LEENAARS et al., 2020). Como trata-se de uma validação semelhante à validação cruzada *k-fold*, a parametrização robusta com uso de $n_{tree}=1000$ permitiu várias iterações, o que permitiu avaliar o efeito da redução da dimensionalidade. Millard e Richardson (2015) avaliando os efeitos da seleção de variáveis para a classificação de imagens de sensoriamento remoto encontraram resultado semelhante, concluindo que classificações mais estáveis são obtidas com a redução da dimensionalidade, ou seja, com o emprego de covariáveis importantes e não correlacionadas. Além dessas vantagens, a redução do número de variáveis favorece a interpretabilidade dos modelos, contribuindo para o avanço do conhecimento pedológico em relação ao uso de novas covariáveis ambientais (ARROUAYS et al., 2020a; KHALEDIAN; MILLER, 2020; MENEZES et al., 2020; WADOUX et al., 2020). A avaliação da diferença entre a acurácia global e a acurácia OOB também é uma métrica utilizada para avaliar a modelagem preditiva, sendo que a redução do valor dessa diferença indica maior estabilidade do modelo (MILLARD; RICHARDSON, 2015), observada principalmente na etapa RF3.

3.3 Avaliação da extrapolação espacial (SSAM)

A representação do espaço pelas amostras de perfis de solos, segundo as covariáveis ambientais foi calculada segundo a SSAM para cada uma das etapas de seleção das variáveis (RF1, RF2 e RF3) (FIGURA 4A e B). Quanto maior o valor da SSAM, mais comuns são os

valores das covariáveis para o local de predição em relação à amostra e mais confiável é a predição espacial. A SSAM para RF1 e RF2 são iguais, tanto na distribuição espacial, quanto na quantificação em área (FIGURA 4A). Assim como ocorreu com as medidas de acurácia, a remoção das covariáveis ambientais consideradas menos importantes ($MDA < 20$) não afetou o nível de extrapolação dos dados.

Figura 4- Superfície de Similaridade Ambiental Multivariada (SSAM) dos modelos que apresentaram a maior acurácia global em cada uma das três etapas de seleção de variáveis.



Legenda: SSAM: Superfície de similaridade ambiental multivariada. RF1, RF2 e RF3: etapas da modelagem com *Random Forest*.

Os resultados obtidos para a SSAM indicam adequada representatividade das amostras de solos (perfis legados) em relação às covariáveis ambientais. Menos de 0,5% das áreas apresentaram valores negativos nas três etapas de seleção (RF1, RF2 e RF3). A seleção de variáveis na etapa RF3 aumentou a representatividade do espaço pelas amostras, reduziu as áreas com menor similaridade (categorias < 0 e 0-30) e aumentou áreas com maior similaridade (categorias 30-60 e > 60) em relação aos modelos RF1 e RF2.

A investigação acerca dos valores negativos na SSAM em RF1 e RF2 indicou que o mapa de carbono no solo (GOMES et al., 2019) foi a covariável responsável pela área em vermelho na SSAM. Nas etapas iniciais esta covariável foi selecionada como uma das mais importantes para a predição de solos e, especificamente na área onde a SSAM é negativa, a estimativa de carbono no solo é muito baixa ($< 0,1 \text{ kg.m}^{-2}$) e não se ajustou ao modelo de predição. Na etapa RF3, os locais que apresentaram valores negativos representam uma área menos extensa e localizam-se principalmente nas regiões altas da Serra da Mantiqueira, onde a amostragem de perfis de solo era reduzida, dada a dificuldade de acesso oferecida pelas características do local.

3.4 Entropia

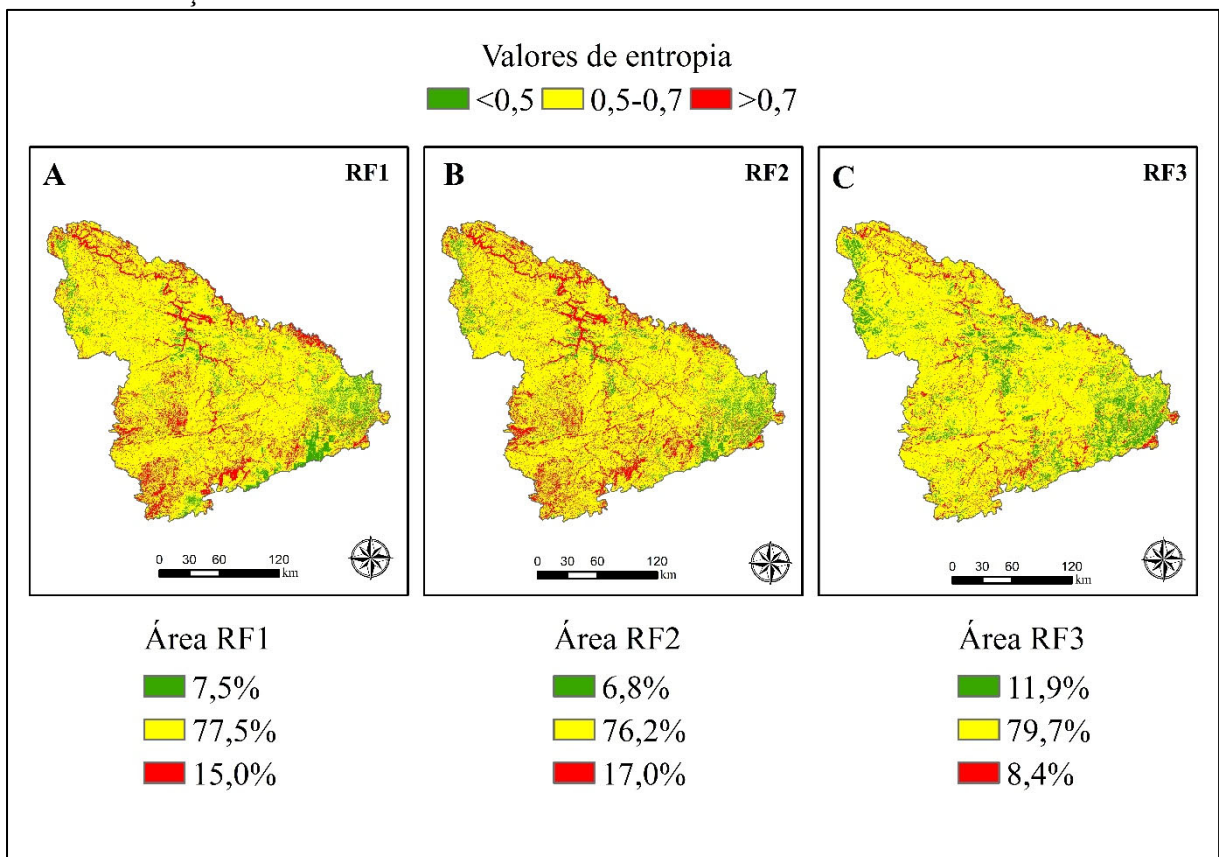
A incerteza das predições foi verificada por meio da entropia normalizada de *Shannon*, aplicada a cada píxel de predição espacial. O resultado está relacionado à incerteza durante o processo de construção das árvores de decisão no algoritmo *Random Forest*. Quanto maior a associação de uma classe de solo ao resultado da predição e menor a associação a outras classes, maior a certeza e menores os valores de entropia. Quanto maior a entropia, maior a complexidade da paisagem para a classificação do solo e maiores as chances de classificação incorreta (MACHADO et al., 2019a). Os valores variaram de 0 a 0,89, observando-se valores mais elevados para RF1 e RF2 e menores para RF3 (TABELA 3). Foram observados valores decrescentes de entropia conforme o conjunto de covariáveis ambientais foi se restringindo às mais importantes. Para melhor visualização do efeito espacial da seleção de variáveis foram construídos mapas de entropia (FIGURA 5).

Tabela 3- Valores de entropia dos modelos de classificação em cada uma das três etapas de seleção de covariáveis ambientais.

Etapa	Número de covariáveis	Mínimo	Máximo	Média
RF1	55	0,03	0,89	0,62
RF2	37	0,03	0,88	0,63
RF3	12	0,01	0,84	0,60

Legenda: RF1, RF2 e RF3: etapas sequenciais de seleção de variáveis

Figura 5- Mapas de entropia dos modelos de classificação em cada uma das três etapas de seleção de variáveis ambientais.



Legenda: RF1, RF2, RF3: Etapas de seleção de variáveis ambientais.

As áreas que apresentaram maiores valores de entropia ($> 0,7$) são coincidentes com as regiões próximas à rede de drenagem nos três mapas. A variabilidade dos solos nesses segmentos da paisagem é elevada (RY, CY, GX, OX), refletindo inundações periódicas, canais abandonados e fontes múltiplas de sedimentos (CURI; RESENDE; SANTANA, 1985; KÄMPF; CURI, 2012b). Modelos teóricos para a ocorrência de diferentes tipos de solos próximos à rede de drenagem podem ser entendidos de acordo com a geomorfologia fluvial. O volume de água escoado na rede de drenagem varia regularmente, conforme a estação do ano (períodos mais chuvosos e mais secos). Essa variação interfere na largura do leito de drenagem e proporciona a formação de diques separando o leito vazante do leito marginal (onde o canal de drenagem é ocupado apenas esporadicamente) (CUNHA, 2013). No dique ocorrem solos aluviais (CY e RY); no leito marginal podem ocorrer solos hidromórficos (GX, GM e OX) que se diferenciam na paisagem segundo a profundidade do leito, altura e estabilidade relativa do nível de água e deficiência de arejamento (CURI; RESENDE; SANTANA, 1985). A

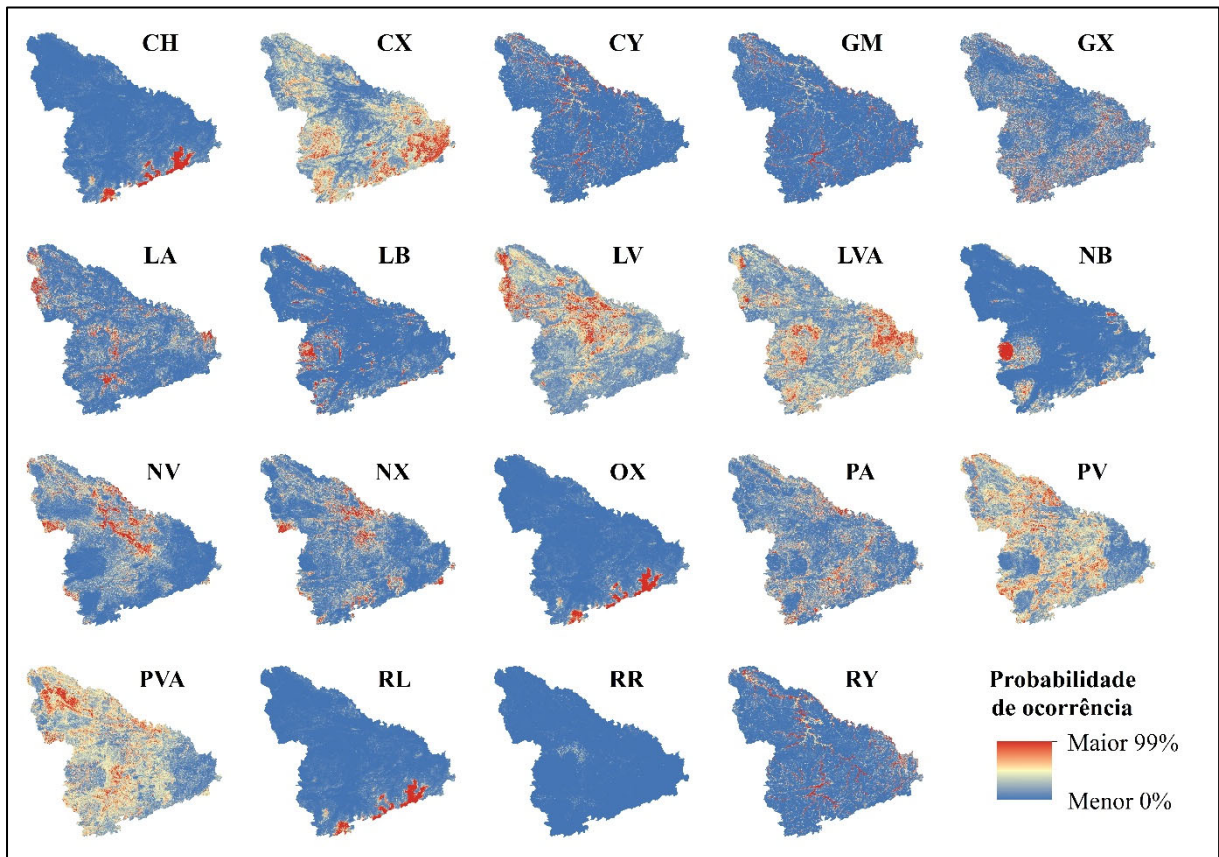
diversidade de ambientes em uma porção relativamente estreita da paisagem proporciona maior complexidade na ocorrência dos solos, aumentando a incerteza da predição nesses locais.

As áreas que apresentaram menor entropia estão concentradas em três regiões: a) no planalto da bacia do Alto Rio Grande; b) na porção central, próximo aos grandes reservatórios das usinas hidrelétricas de Furnas e; c) na porção noroeste, próximo à divisa com o estado de São Paulo. Nas duas primeiras regiões, a densidade de amostras de perfis de solo é alta em relação à última (FIGURA 1). Esse comportamento não era esperado, pois trabalhos relacionaram maiores incertezas na predição ao menor número de amostras (YANG et al., 2011; BRUNGARD et al., 2015; LI et al., 2016). Esse comportamento permite inferir que a maior ou menor certeza na classificação não está relacionada apenas à quantidade de amostras, mas também pode estar relacionada à complexidade da relação solo-paisagem e à sua representatividade por meio das covariáveis ambientais. Quando o processo de classificação resulta em altos níveis de entropia, indica que as instâncias não são bem capturadas pelas configurações solo-ambiente, ou o banco de dados não é suficiente para capturar as variações do solo. Avaliações de entropia podem ser extremamente úteis para auxiliar o planejamento amostral de levantamentos de solo, direcionando esforços à locais onde a complexidade do terreno exige maior esforço amostral.

3.5 Probabilidade de ocorrência

Os mapas de probabilidade de ocorrência das classes de solo foram gerados a partir do modelo com menor nível de extrapolação (SSAM), menor entropia da predição e maior acurácia global selecionado na última etapa da modelagem (RF3) (FIGURA 6). Embora essas não sejam probabilidades baseadas em funções de distribuição, mas em número de votos durante a construção das árvores de decisão, elas fornecem uma medida da confiança na classificação e podem ser usadas para indicar locais onde o conjunto das variáveis ambientais compõe ambientes mais favoráveis à ocorrência de determinadas classes de solo (TENG et al., 2018).

Figura 6- Mapas de probabilidade de ocorrência dos solos no Sul de Minas Gerais.



Legenda: CH: Cambissolo Húmico, CX: Cambissolo Háplico, CY: Cambissolo Flúvico, GM: Gleissolo Melânico, GX: Gleissolo Háplico, LA: Latossolo Amarelo, LB: Latossolo Bruno, LV: Latossolo Vermelho, LVA: Latossolo Vermelho-Amarelo, NB: Nitossolo Bruno, NV: Nitossolo Vermelho, NX: Nitossolo Háplico, OX: Organossolo Háplico, PA: Argissolo Amarelo, PV: Argissolo Vermelho, PVA: Argissolo Vermelho-Amarelo, RL: Neossolo Litólico, RR: Neossolo Regolítico, RY: Neossolo Flúvico.

As regiões com maior probabilidade de ocorrência de CH são coincidentes com as áreas de ocorrência de OX e RL, e concentram-se nas regiões mais altas da Serra da Mantiqueira. Esses locais são caracterizados por baixas temperaturas médias e alta pluviosidade, características que proporcionam condições para produção a conservação da matéria orgânica no solo, favorecendo a ocorrência de CH e OX. Paralelamente, o relevo movimentado proporciona a ocorrência de afloramentos rochosos e de locais onde a conformação do terreno contribui para a formação de solos com horizonte A menos espesso, o caso do RL.

Os CX são os solos com ocorrência mais dispersa ao longo de toda a paisagem. São solos minerais, com horizonte B pouco desenvolvido (SANTOS et al., 2018). A extensa área com alta probabilidade de ocorrência expressa a heterogeneidade dessa classe com relação ao seu ambiente (RESENDE et al., 2014). As características destes solos variam muito, podem se originar de diferentes materiais, formas de relevo e condições climáticas. A classe comporta solos desde fortemente a imperfeitamente drenados, de rasos a profundos, além de diferentes

cores (ARAÚJO et al., 2018). A área contínua com maior probabilidade de ocorrência está na região do planalto do Alto Rio Grande, em concordância com o estudo de Araújo et al. (2018), que relataram ocorrência de 62% de CX para a Bacia do Alto Rio Grande e estudo de Giarola et al. (1997), que avaliando a região sob a influência do reservatório de Camargos/Itutinga reportaram a ocorrência de 42% da área com Cambissolos.

As áreas de maior probabilidade de ocorrência de CY localizam-se ao longo das margens dos canais de drenagem e são coincidentes com a ocorrência de GM e RY. A associação dessas três classes de solo é comum em regiões de várzea (CURI; RESENDE; SANTANA, 1985). Os solos de caráter flúvico, CY e RY, ocorrem nas porções mais próximas ao leito menor dos rios, onde a deposição e remoção constante de sedimentos, proporcionada pela energia do movimento da água, não favorece o desenvolvimento pedológico em profundidade, resultando em solos pouco desenvolvidos. Nessa porção da paisagem, é comum a ocorrência de diques, elevados em relação ao leito de vazão e em relação ao leito maior, com melhor condição de drenagem (CUNHA, 2013). Os solos GM ocorrem após os diques, onde o represamento periódico de água quando o leito do rio se expande e a drenagem prejudicada pela menor distância vertical em relação ao lençol freático proporciona a ocorrência de solos hidromórficos, como os GM mapeados nessas porções (CURI; RESENDE; SANTANA, 1985). Ainda em relação aos solos hidromórficos, a ocorrência dos GX, com maior dispersão geográfica e não associados aos canais de drenagem foi surpreendente e reflete uma falha do modelo preditivo, pois, conforme a gênese desses solos, os GX deveriam ocorrer em regiões semelhantes aos GM.

Os Latossolos encontram-se amplamente distribuídos pela região Sul de Minas Gerais. São solos caracterizados pelo avançado estágio de intemperização, explícito pela transformação quase completa dos minerais facilmente alteráveis, ocorrendo sob diferentes condições climáticas, relevo e material de origem (KER, 1997). Os LV apresentaram maior probabilidade de ocorrência da porção central, no entorno do reservatório da usina hidrelétrica de Furnas, o que corrobora com o levantamento detalhado de solos realizado nesta região pelo CNEPA (1962), que indicou presença de LV em 37% da área. Os LVA apresentaram grande dispersão geográfica, assim como ocorre pelo território nacional (KER, 1997). As maiores probabilidades localizaram-se na região do planalto do Alto Rio Grande. Reforçando esse resultado, Araújo et al. (2018) identificaram que os LVA se estendem por aproximadamente 590 km² nessa região. Os LA apresentaram menor probabilidade de ocorrência. Ker (1997), em estudo sobre os Latossolos do Brasil, verificou que a região Sul de Minas não é típica para a ocorrência dessa

classe, que se concentra na porção mais central do país e nos tabuleiros costeiros. Os LB são normalmente originados de rochas eruptivas básicas, intermediárias e alcalinas, e circunscritos a regiões de maiores altitudes, frias e úmidas (KER, 1997). Em acordo com essas características, os LB tiveram maior probabilidade de ocorrência na região do planalto de Poços de Caldas, que forma uma estrutura circular de 30 km de diâmetro, constituída por grandes massas de depósitos alcalinos ricos em bauxita, zircônio e molibdênio formados no período Fanerozóico (ULBRICH; GOMES, 1981).

Nesta mesma região do Planalto de Poços de Caldas, além dos LB, são predominantes os NB, conforme reportado por Carvalho et al. (1983). Ainda na ordem dos Nitossolos, os NV e NX têm maior probabilidade de ocorrência na porção central, principalmente na região do entorno do reservatório da usina hidrelétrica de Furnas. O levantamento de solos realizado pelo CNEPA (1962) mapeou manchas contínuas de LV nesta região.

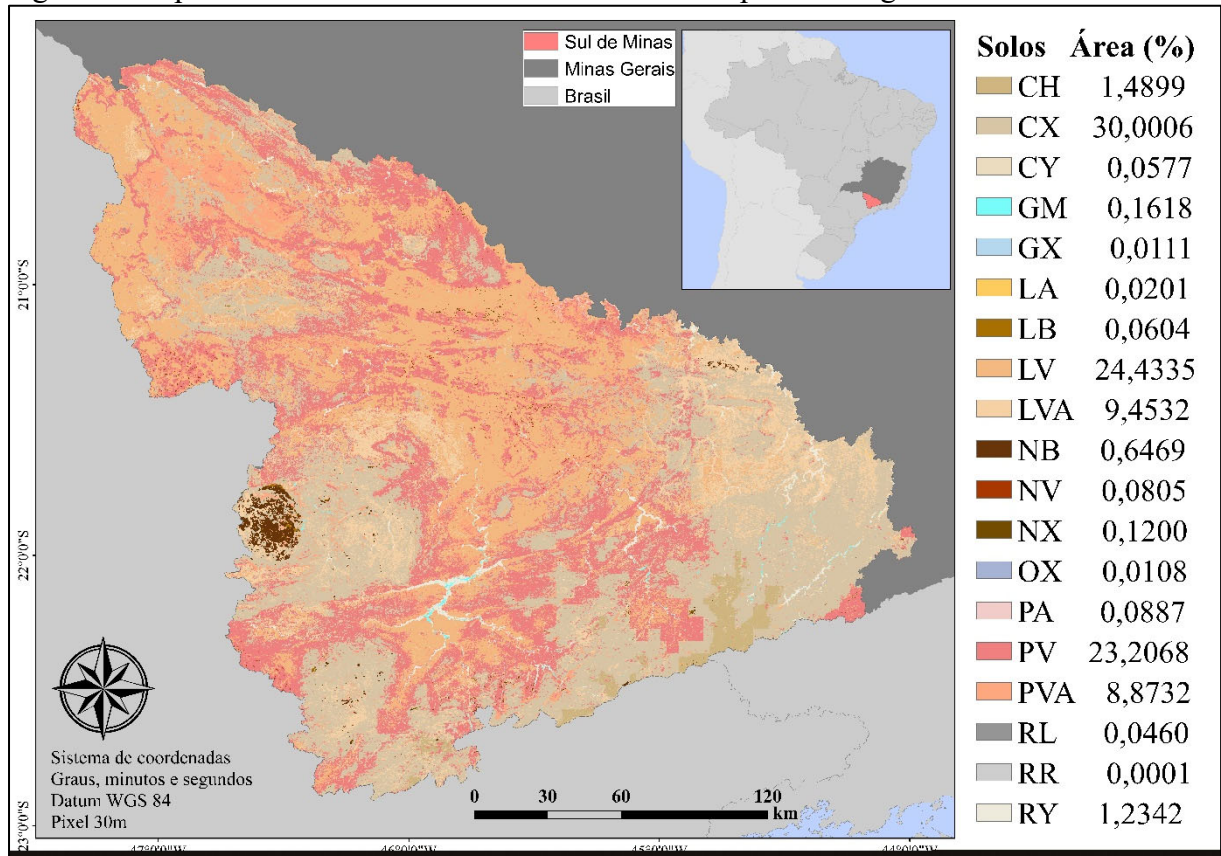
A literatura reporta ocorrência expressiva de Argissolos no Sul de Minas (SILVA et al., 2013; RESENDE et al., 2014), o que corrobora com o resultado da predição. Os PV apresentaram probabilidade de ocorrência pela maior parte do território, com exceção da região da bacia do Alto Rio Grande e do planalto de Poços de Caldas. Os PVA apresentaram menores probabilidades de ocorrer apenas na região do Alto Rio Grande, e maior probabilidade de ocorrer na região norte. Os PA apresentaram baixas probabilidades de ocorrência em toda a área do estudo.

3.6 Predição espacial de solos

O mapa de solos (FIGURA 7) indica predominância das classes CX, LV, PV, LVA e PVA. Juntas, essas classes foram preditas para mais de 95% da área total do Sul de Minas. Dentre as classes majoritárias, a de maior expressão, recobrando mais de 30% da área do Sul de Minas é a dos CX. Esses solos estão associados a diferentes tipos de rochas e podem ocupar praticamente todas as situações de topografia, mas com certa constância nas formas mais acidentadas (OLIVEIRA; JACOMINE; COUTO, 2017). O mapa de solos de Minas Gerais (CETEC et al., 2010) reporta CX em caráter de dominância como a terceira classe de maior ocorrência, com 15,7% da área do Sul de Minas, muito difundidos na condição de componente secundário como inclusões em quase todas as unidades de mapeamento. A aplicação de uma metodologia baseada no MDS, em que a resolução espacial das variáveis ambientais

determinou a área mínima mapeável, permitiu mapear os CX que, em mapeamentos em menor escala, ocorriam apenas como inclusões.

Figura 7- Mapa de solos do sul de Minas Gerais obtido a partir do algoritmo *Random Forest*



Legenda: SBCS: Sistema Brasileiro de Classificação de Solos, CH: Cambissolo Húmico, CX: Cambissolo Háplico, CY: Cambissolo Flúvico, GM: Gleissolo Melânico, GX: Gleissolo Háplico, LA: Latossolo Amarelo, LB: Latossolo Bruno, LV: Latossolo Vermelho, LVA: Latossolo Vermelho-Amarelo, NB: Nitossolo Bruno, NV: Nitossolo Vermelho, NX: Nitossolo Háplico, OX: Organossolo Háplico, PA: Argissolo Amarelo, PV: Argissolo Vermelho, PVA: Argissolo Vermelho-Amarelo, RL: Neossolo Litólico, RR: Neossolo Regolítico, RY: Neossolo Flúvico.

A acurácia do produtor e do usuário foram calculadas a partir da matriz de confusão (TABELA 4). A acurácia do produtor reporta os erros de omissão, ou seja, as áreas que deveriam ter sido mapeadas segundo determinada classe, mas não foram. A acurácia do usuário reporta os erros de inclusão, ou seja, as áreas que não correspondem à classe mapeada, mas que estão incluídas no mapa. Os erros de omissão e inclusão foram estimados em termos de área segundo a Tabela 5.

Tabela 4- Matriz de confusão do modelo selecionado para a predição .

	CH	CX	CY	GM	GX	LA	LB	LV	LVA	NB	NV	NX	OX	PA	PV	PVA	RL	RR	RY	AP
CH	4																			1,00
CX		99					3	4	4				2			1	6			3 0,81
CY			1																	1,00
GM				1																1,00
GX					6															1,00
LA						7			2											0,78
LB							1													1,00
LV		1						104	1		2			2	2	1				0,92
LVA		2			1			4	60							2				0,87
NB										5										1,00
NV											16						1			0,94
NX												5								1,00
OX													1							1,00
PA														13						1,00
PV	1								3					1	73	3				0,90
PVA		1						2						1		56				0,93
RL																	3			1,00
RR																		1		1,00
RY		2							2											11 0,73
AU	0,80	0,94	1,00	1,00	0,86	1,00	0,25	0,91	0,83	1,00	0,89	1,00	0,33	0,76	0,97	0,88	0,33	1,00	0,79	0,89

Legenda: CH: Cambissolo Húmico, CX: Cambissolo Háplico, CY: Cambissolo Flúvico, GM: Gleissolo Melânico, GX: Gleissolo Háplico, LA: Latossolo Amarelo, LB: Latossolo Bruno, LV: Latossolo Vermelho, LVA: Latossolo Vermelho-Amarelo, NB: Nitossolo Bruno, NV: Nitossolo Vermelho, NX: Nitossolo Háplico, OX: Organossolo Háplico, PA: Argissolo Amarelo, PV: Argissolo Vermelho, PVA: Argissolo Vermelho-Amarelo, RL: Neossolo Litólico, RR: Neossolo Regolítico, RY: Neossolo Flúvico, AP: Acurácia do produtor, AU: Acurácia do usuário.

Tabela 5- Área predita, erros de omissão e de inclusão para cada classe de solo mapeada.

Solos	Área predita (ha)	Erro de omissão (%)	Erro de omissão (ha)	Erro de inclusão (%)	Erro de inclusão (ha)
CX	1.589.470,2	5,7	9.082.686,9	18,9	29.965.421,8
PV	1.229.527,6	2,7	3.278.740,3	9,9	12.143.482,7
LV	1.294.516,4	8,8	11.355.406,6	8,0	10.310.307,2
LVA	500.844,2	16,7	8.347.402,5	13,0	6.532.749,8
PVA	470.115,9	12,5	5.876.448,8	6,7	3.134.106,0
RY	65.387,8	21,4	1.401.166,9	26,7	1.743.674,4
NV	4.504,2	11,1	50.047,0	5,9	26.495,5
LA	1.067,3	0,0	0,0	22,2	23.718,0
LB	3.200,4	75,0	240.030,0	0,0	0,0
RL	2.437,5	66,7	162.498,0	0,0	0,0
OX	570,1	66,7	38.010,0	0,0	0,0
PA	4.698,5	23,5	110.553,9	0,0	0,0
CH	78.939,2	20,0	1.578.783,6	0,0	0,0
GX	585,8	14,3	8.368,7	0,0	0,0
NB	34.271,6	0,0	0,0	0,0	0,0
GM	8.570,7	0,0	0,0	0,0	0,0
NX	6.358,3	0,0	0,0	0,0	0,0
CY	3.059,1	0,0	0,0	0,0	0,0
RR	5,5	0,0	0,0	0,0	0,0

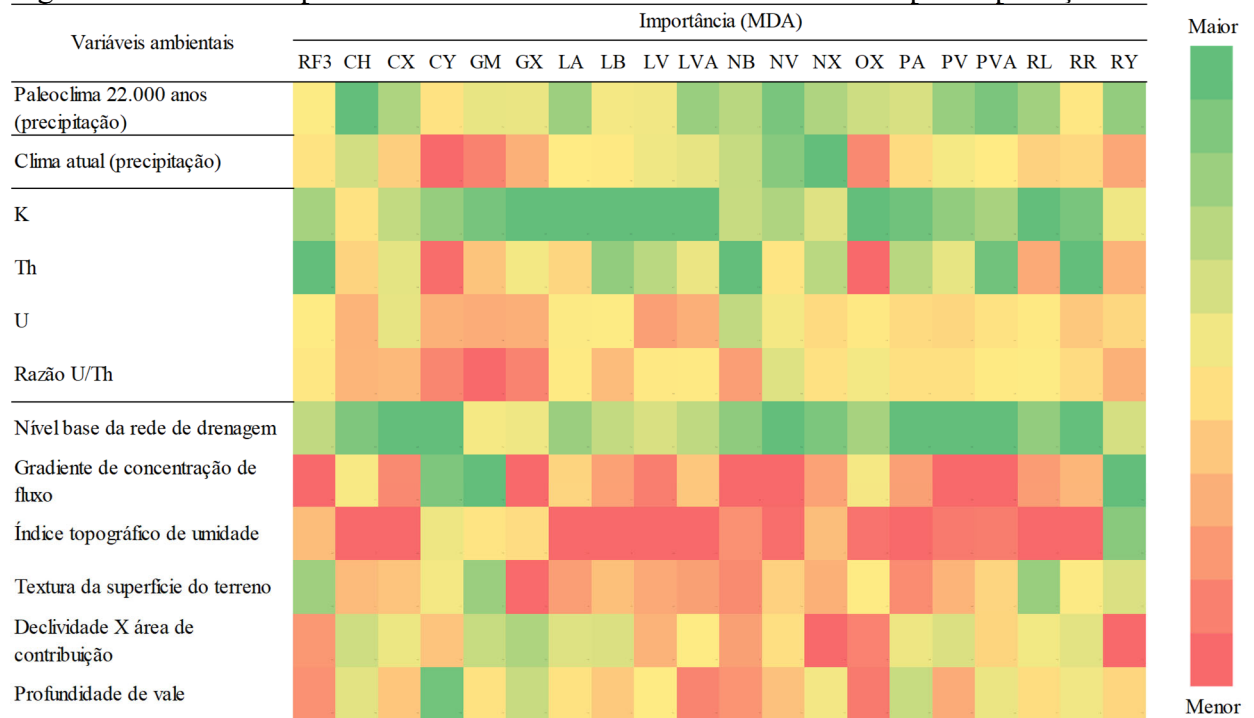
Legenda: CH: Cambissolo Húmico, CX: Cambissolo Háptico, CY: Cambissolo Flúvico, GM: Gleissolo Melânico, GX: Gleissolo Háptico, LA: Latossolo Amarelo, LB: Latossolo Bruno, LV: Latossolo Vermelho, LVA: Latossolo Vermelho-Amarelo, NB: Nitossolo Bruno, NV: Nitossolo Vermelho, NX: Nitossolo Háptico, OX: Organossolo Háptico, PA: Argissolo Amarelo, PV: Argissolo Vermelho, PVA: Argissolo Vermelho-Amarelo, RL: Neossolo Litólico, RR: Neossolo Regolítico, RY: Neossolo Flúvico

Os LB apresentaram maiores erros de omissão (75%), seguidos do RL e OX em 67% da área não mapeada. Os OX ocorreram em locais de drenagem livre, sob clima altimontano (RESENDE et al., 2014), porém, os OX típicos de várzeas que ocorrem próximos à rede de drenagem e a outros solos hidromórficos (CURI; RESENDE; SANTANA, 1985) não foram preditos. De maneira geral, os erros de inclusão (%) foram menores que os erros de omissão. O maior erro de inclusão foi observado para RY (27%). Novamente, a elevada complexidade da relação solo-paisagem nas regiões próximas à rede de drenagem ajudam a explicar esse resultado.

3.7 Importância das covariáveis ambientais

Após três etapas de seleção, foram identificadas as 12 covariáveis ambientais mais importantes para a predição espacial de solos. Os resultados do melhor modelo foram representados segundo a ordem de importância (MDA) para o modelo (RF3) e para cada uma das classes de solo (FIGURA 8). Considerando a importância para o modelo, as covariáveis mais importantes foram teores de Th e K. Considerando a importância relativa para as classes, as covariáveis ambientais K, nível base da rede de drenagem e paleoclima foram as mais importantes para a predição.

Figura 8- Escala de importância das covariáveis ambientais utilizadas para a predição.



Legenda: MDA: métrica fornecida pela *Random Forest* para indicar a importância relativa das covariáveis ambientais. A *Random Forest* fornece valores de MDA para o modelo como um todo (RF3) e para cada uma das classes preditas. CH: Cambissolo Húmico, CX: Cambissolo Háplico, CY: Cambissolo Flúvico, GM: Gleissolo Melânico, GX: Gleissolo Háplico, LA: Latossolo Amarelo, LB: Latossolo Bruno, LV: Latossolo Vermelho, LVA: Latossolo Vermelho-Amarelo, NB: Nitossolo Bruno, NV: Nitossolo Vermelho, NX: Nitossolo Háplico, OX: Organossolo Háplico, PA: Argissolo Amarelo, PV: Argissolo Vermelho, PVA: Argissolo Vermelho-Amarelo, RL: Neossolo Litólico, RR: Neossolo Regolítico, RY: Neossolo Flúvico.

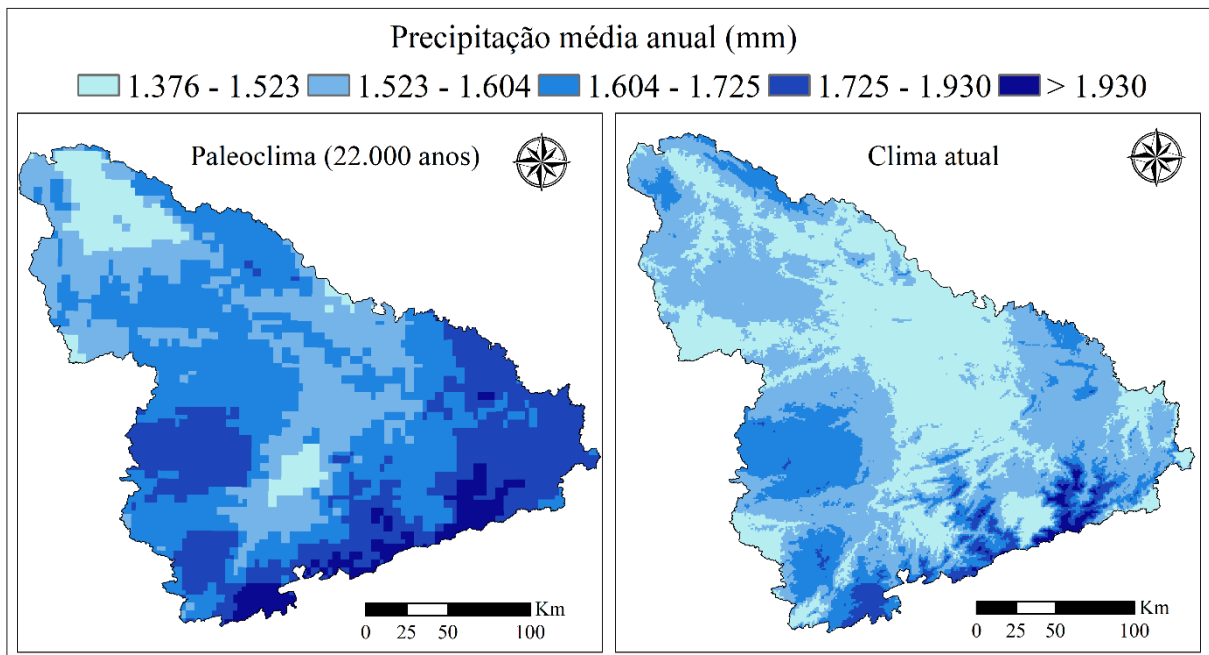
O K é o principal constituinte da maioria das rochas e o elemento de alteração predominante na maioria dos depósitos minerais (McBRATNEY; MENDONÇA SANTOS; MINASNY, 2003). Durante o intemperismo das rochas, as concentrações de K diminuem à medida que as concentrações de Th aumentam. O K é um elemento móvel, que tende a se redistribuir na paisagem conforme ocorrem os processos erosivos, enquanto o Th, como elemento imóvel, tende a se concentrar conforme outros elementos são removidos durante o intemperismo e a erosão (MA et al., 2019; MELLO et al., 2021). Juntamente com outras variáveis ambientais, os dados de espectrometria gama fornecem informações importantes para a compreensão de diversos processos de formação da paisagem (WILFORD, 2012; MOONJUN et al., 2017; ROUZE; MORGAN; McBRATNEY, 2017; MA et al., 2019; REINHARDT; HERRMANN, 2019; MELLO et al., 2021), como a redistribuição de sedimentos (GATEHOUSE; WILLIAMS; PILLANS, 2001), identificação de aspectos hídricos (CATTLE et al., 2003; BEAMISH, 2013; READ et al., 2018) e caracterização da textura do solo

(REINHARDT; HERRMANN, 2019). A versatilidade das interpretações a partir dos resultados obtidos com a aeroespectrometria gama justificam a posição de destaque das variáveis K e Th para a predição.

O nível base da rede de drenagem foi a covariável representativa do fator de formação do solo relevo que apresentou maior importância relativa para a predição da maioria das classes. Esta covariável reflete características dos solos como profundidade, drenagem e cor, aspectos diretamente relacionados aos critérios de classificação de solos, principalmente ao nível de Ordem e Subordem do SiBCS (SANTOS et al., 2018), que guarda estreita correspondência entre os conceitos de cada classe e a ocorrência dos solos na paisagem (RESENDE et al., 2014). Para as classes de solos típicas de ambientes de várzea, RY, CY e GM, as covariáveis gradiente de concentração de fluxo, índice topográfico de umidade e profundidade de vale se destacaram quanto ao nível de importância. Essas variáveis derivadas do DEM que refletem aspectos da hidrologia do terreno também figuram em posição de destaque em diversos trabalhos de mapeamento de solos (SILVA et al., 2016, 2019; HENGL et al., 2017; KRAMM; HOFFMEISTER, 2020).

O paleoclima apresentou maior importância relativa que o clima atual para quase todas as classes de solo, exceto para o NV. Esse resultado reflete a natureza poligenética dos solos tropicais, onde muitas paisagens não estão em equilíbrio com as condições climáticas atuais. A variável paleoclimática representa um recorte do período Quaternário, caracterizado por glaciações periódicas e períodos interglaciais. Acredita-se que no Brasil, o período interglacial corresponde a clima mais quente e chuvoso, conforme representado pela covariável do paleoclima selecionada para predição (FIGURA 9). Essas condições favorecem tanto a erosão quanto a pedogênese, o impacto das chuvas escava o leito rochoso e carrega sedimentos, retirando-os do sistema ou depositando-os em locais mais baixos, próximos à rede de drenagem. Este processo proporciona a formação de depósitos aluviais em diques e brejos. O escoamento da água provoca dissecação e adelgaçamento do perfil onde as condições de relevo proporcionem escoamento, incentivando o rejuvenescimento do solo. Em condições de relevo mais suavizado, a infiltração de água no perfil do solo e favorece o intemperismo das rochas em contato com o regolito, aprofundando o perfil e acelerando a pedogênese (RESENDE et al., 2019).

Figura 9- Precipitação média anual referente ao Paleoclima (22.000 anos pretéritos) e precipitação média anual atual na região Sul de Minas Gerais, Brasil.



Segundo o enfoque pedogeomorfológico, as paisagens e os solos que as compõem resultam do efeito cumulativo dos ambientes. Assim, de certa forma, só há um ambiente original, todos os outros herdam algo dos ambientes posteriores, são relictos (RESENDE et al., 2014). Considerando a natureza poligenética dos solos brasileiros, avançar nas questões climáticas de períodos remotos auxiliará nas predições das características atuais da distribuição dos solos nas paisagens.

4 CONCLUSÕES

A aplicação do MDS utilizando covariáveis ambientais inéditas aliadas a outras já consagradas no mapeamento de solos no Brasil possibilitou a confecção de um mapa de solos com adequada acurácia e métricas de avaliação de incerteza. Classes de solo, com menor extensão territorial que figuravam como inclusões em levantamentos de solos anteriores foram mapeadas, agregando informação sobre os solos da região.

Os mapas de avaliação do nível de extrapolação do modelo indicaram a boa representatividade das covariáveis ambientais selecionadas segundo as amostras derivadas do legado, indicando o grande potencial de uso das informações disponíveis para avançar no mapeamento de solos no Brasil. Iniciativas de confecção de bancos de dados digitais devem ser estimuladas e priorizadas visando ao resgate do conhecimento sobre os solos e a diminuição de custos com novos levantamentos.

A avaliação da incerteza indicou que, ao contrário da hipótese inicial, a incerteza nem sempre foi diretamente relacionada à densidade amostral. Locais com baixa densidade de amostras apresentaram elevado nível de certeza da predição.

O protocolo desenvolvido em etapas sequenciais possibilitou selecionar um conjunto de covariáveis ambientais representativo dos principais fatores de formação dos solos, o que contribuiu para a interpretabilidade do modelo. O paleoclima apresentou destaque na predição, com maior nível de importância que a covariável do clima atual. Os resultados apresentados pelas covariáveis K e Th, derivadas dos dados de aeroespectrometria gama demonstraram o grande potencial desses dados para o mapeamento de solos. A natureza poligenética dos solos tropicais demanda avanços no desenvolvimento e aplicação de novas covariáveis ambientais para decifrar a complexidade da distribuição dos solos nessas paisagens.

REFERÊNCIAS

- 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**, v. 214–215, n. 2014, p. 101–113, 2014. doi: 10.1016/j.geoderma.2013.09.023.
- ALKMIM, F. F. **História Geológica de Minas Gerais**. . Acesso em: 26 jul. 2021.
- ALMEIDA, F. F. M. de. Origem e evolução da plataforma brasileira. In: **Boletim 241**. Rio de Janeiro: DNPM/Divisão de Geologia e Mineralogia, 2000. p. 36. 2000.
- AMARAL, F. C. S. do; SANTOS, H. G. dos; ÁGILO, M. L. D.; DUARTE, M. N.; OLIVEIRA, R. P. de; CARVALHO JÚNIOR, W. **Mapeamento de Solos e Aptidão Agrícola das terras do estado de Minas Gerais**. Rio de Janeiro: Embrapa, 2004.
- ANA, A. N. de Á. **Plano Integrado de Recursos Hídricos da Bacia Hidrográfica do Rio Grande**, 2017. .
- ARAÚJO, A. R.; OLIVEIRA, J. M.; PEREIRA, P.; CURI, N.; MARQUES, J. J. G. M. **Geomorfologia , solos e aptidão agrícola das terras da bacia do Alto Rio Grande , Minas Gerais**. Lavras: Ed: UFLA, 2018.
- ARROUAYS, D.; MCBRATNEY, A.; BOUMA, J.; LIBOHOVA, Z.; RICHER-DEFORGES, A. C.; MORGAN, C. L. S.; ROUDIER, P.; POGGIO, L.; MULDER, V. L. Impressions of digital soil maps: The good, the not so good, and making them ever better. **Geoderma Regional**, v. 20, p. e00255, 2020a. doi: 10.1016/j.geodrs.2020.e00255.
- ARROUAYS, D.; POGGIO, L.; SALAZAR GUERRERO, O. A.; MULDER, V. L. Digital soil mapping and GlobalSoilMap. Main advances and ways forward. **Geoderma Regional**, v. 21, p. e00265, 2020b. doi: 10.1016/j.geodrs.2020.e00265.
- BARBOSA, O.; BRAUN, O. P. G.; DYER, R. C.; CUNHA, C. A. B. R. Geologia da região do Triângulo Mineiro. **Boletim DNPM/DFPM**, 136, p. 140, 1970.
- BEAMISH, D. Gamma ray attenuation in the soils of Northern Ireland, with special reference to peat. **Journal of Environmental Radioactivity**, v. 115, p. 13–27, 2013. doi: 10.1016/j.jenvrad.2012.05.031.
- BERTOLDO, M. A. **CARACTERIZAÇÃO EDAFOAMBIENTAL DA CAFEICULTURA NA REGIÃO DE TRÊS PONTAS, MINAS GERAIS**. 2008. Universidade Federal de Lavras - UFLA, 2008.
- BRASIL, M. das M. e E. **Projeto Radam Brasil. Levantamento de Recursos Naturais. Volume 32. Folhas SF 23/24 Rio de Janeiro/Vitória**. Rio de Janeiro: MME, 1983.

BREIMAN, L. Random Forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001. doi: 10.1023/A:1010933404324.

BREVIK, E. C.; CALZOLARI, C.; MILLER, B. A.; PEREIRA, P.; KABALA, C.; BAUMGARTEN, A.; JORDÁN, A. Soil mapping, classification, and pedologic modeling: History and future directions. **Geoderma**, v. 264, p. 256–274, 2016. doi: 10.1016/j.geoderma.2015.05.017.

BRITO NEVES, B. B.; CAMPOS-NETO, M. C.; FUCK, R. A. From Rodinia to Western Gondwana: An approach to the Brasiliano-Pan-Africa. v. 22, n. 3, p. 155–166, 1999.

BRUNGARD, C. W.; BOETTINGER, J. L.; DUNIWAY, M. C.; WILLS, S. A.; EDWARDS, T. C. Machine learning for predicting soil classes in three semi-arid landscapes. **Geoderma**, v. 239–240, p. 68–83, 1 fev. 2015. doi: 10.1016/j.geoderma.2014.09.019.

BUI, E. N. Soil survey as a knowledge system. **Geoderma**, v. 120, n. 1–2, p. 17–26, 2004. doi: 10.1016/j.geoderma.2003.07.006.

CAIRES, S.; FONTES, M. P. F.; KER, J. C. Dados para o mapa de solos de Minas Gerais. In: UFV (Ed.). **Infraestrutura de Dados Espaciais do Sistema Estadual de Meio Ambiente e Recursos Hídricos**. Belo Horizonte: SISEMA, 2007. 2007.

CAMERA, C.; ZOMENI, Z.; NOLLER, J. S.; ZISSIMOS, A. M.; CHRISTOFOROU, I. C.; BRUGGEMAN, A. Geoderma A high resolution map of soil types and physical properties for Cyprus : A digital soil mapping optimization. **Geoderma**, v. 285, p. 35–49, 2017. doi: 10.1016/j.geoderma.2016.09.019.

CAMPOS-NETO, M. C. **Orogenic Systems from Southwestern Gondwana: An approach to Brasiliano-PanAfrican Cycle and Orogenic Collage in Southeastern Brazil**. Rio de Janeiro: 31º International Geological Congress, 2000.

CARVALHO, C. C.; NUNES, F. C.; ANTUNES, A. M. H. Histórico do levantamento de solos no Brasil: Da industrialização brasileira à era da informação. **Revista brasileira de cartografia**, v. 65, p. 997–1013, 2013.

CARVALHO, A. P. DE; LARACH, J. O. I.; CAMARGO, M. N.; OLIVEIRA, J. B. de; SANTOS, H. G. dos; WITTERN, E. P.; CONCEIÇÃO, M. da; TAVARES, N. P.; LOUZADA, E. S.; RAMOS, D. P.; PADRO, H.; SOUZA, J. L. R. de; MONIZ, A. C.; ALMEIDA, C. L. F.; DURIEZ, M. A. de; JOHAS, R. A. L.; MELO, M. E. C. C. de; ARAUJO, W. S. de; BLOISE, R. M.; MOREIRA, G. N. C.; PAULA, J. L. de; LIMA, T. da C.; RODRIGUES, E. M.; ANTONELLO, L. L.; LIMA, P. C.; PINTO, J. da S. **Conjunto de dados dos Anais da II Reunião de Classificação, Correlação de Solo e Interpretação de Aptidão Agrícola**. 1º ed. Rio de Janeiro: Embrapa Solos, 1983a.

CARVALHO, A. P. de; LARACH, J. O. I.; CAMARGO, M. N.; OLIVEIRA, J. B. de; SANTOS, H. G. dos; WITTERN, E. P. M. K. P.; CONCEIÇÃO, M. da; TAVARES, N. P.;

LOUZADA, E. S.; RAMOS, D. P.; PADRO, H.; SOUZA, J. L. R. de; MONIZ, A. C.; ALMEIDA, C. L. F. de; DURIEZ, M. A. de M.; JOHAS, R. A. L.; MELO, M. E. C. C. de M.; ARAUJO, W. S. de; BLOISE, R. M.; MOREIRA, G. N. C.; PAULA, J. L. de; LIMA, T. da C.; RODRIGUES, E. M.; ANTONELLO, L. L.; LIMA, P. C.; PINTO, J. da S. **Conjunto de dados dos Anais da II Reunião de Classificação, Correlação de Solo e Interpretação de Aptidão Agrícola**. Rio de Janeiro: Embrapa Solos, 1983b.

CATTLE, S. R.; MEAKIN, S. N.; RUSZKOWSKI, P.; CAMERON, R. G. Using radiometric data to identify aeolian dust additions to topsoil of the Hillston district, western NSW. **Soil Research**, v. 41, n. 8, p. 1439–1456, 2003. doi: 10.1071/SR03026.

CETEC; FEAM; UFV; UFLA. **Mapa de solos do Estado de Minas Gerais**. 1º ed. Belo Horizonte: Fundação Estadual do Meio Ambiente - FEAM, 2010.

CHAKRABORTY, S.; MAN, T.; PAULETTE, L.; DEB, S.; LI, B.; WEINDORF, D. C.; FRAZIER, M. Rapid assessment of smelter/mining soil contamination via portable X-ray fluorescence spectrometry and indicator kriging. **Geoderma**, v. 306, n. April, p. 108–119, 2017. doi: 10.1016/j.geoderma.2017.07.003.

CNEPA, C. N. de P. A. **Levantamento de reconhecimento dos solos da região sob influência do reservatório de Furnas**. 1º ed. Rio de Janeiro: Ministério da Agricultura, 1962.

COELHO, F. F.; GIASSON, E.; CAMPOS, A. R.; TIECHER, T.; COSTA, J. J. F.; COBLINSKI, J. A. Digital soil class mapping in Brazil: A systematic review. **Scientia Agricola**, v. 78, n. 5, p. 1–11, 2021. doi: 10.1590/1678-992x-2019-0227.

COELHO, R. M.; ROSSI, M.; MATTOS, I. F. de A. Solos da Mata Atlântica. In: **Pedologia: Solos dos Biomas Brasileiros**. 1º ed. Viçosa, MG: Sociedade Brasileira de Ciência do Solo, 2017. p. 261–302. 2017.

COHEN, J. A coefficient of agreement for nominal scales. **Educational and Psychological Measurement**, v. 20, n. 1, p. 37- 46 ST- A coefficient of agreement for nominal, 1960.

CONAB. Acompanhamento da safra brasileira de café - Primeiro Levantamento. **Companhia Nacional de Abastecimento**, v. 6, 2019.

CONGALTON, R. G.; GREEN, K. **Assessing the Accuracy of Remotely Sensed Data Principles and Practices**. 3. ed. London: CRC Press, 2019. v. 53

CONRAD, O.; BECHTEL, B.; BOCK, M.; DIETRICH, H.; FISCHER, E.; GERLITZ, L.; WEHBERG, J.; WICHMANN, V.; BÖHNER, J. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. **Geoscientific Model Development**, v. 8, n. 7, p. 1991–2007, 2015. doi: 10.5194/gmd-8-1991-2015.

COOK, S. E.; CORNER, R. J.; GROVES, P. R.; GREALISH, G. J. Use of airborne gamma

radiometric data for soil mapping. **Australian Journal of Soil Research**, v. 34, n. 1, p. 183–194, 1996. doi: 10.1071/SR9960183.

CUNHA, S. B. da. Geomorfologia fluvial. In: CUNHA, S. B. C.; GUERRA, A. J. T. (Ed.). **Geomorfologia: Exercícios, técnicas e aplicações**. Rio de Janeiro: Bertrand Brasil, 2013. p. 348. 2013.

CURI, N.; RESENDE, M.; SANTANA, D. P. Solos de várzeas de Minas Gerais. **Informe Agropecuário**, v. 13, n. 152, p. 3–10, 1985.

DE ALMEIDA, F. F. M.; HASUI, Y.; DE BRITO NEVES, B. B.; FUCK, R. A. Brazilian structural provinces: An introduction. **Earth Science Reviews**, v. 17, n. 1–2, p. 1–29, 1981. doi: 10.1016/0012-8252(81)90003-9.

DIAS, P. H. A.; NOCE, C. M.; PEDROSA-SOARES, A. C.; SEER, H. J.; DUSSIN, I. A.; VALERIANO, C. de M.; KUCHENBECKER, M. O Grupo Ibiá (Faixa Brasília Meridional): Evidências Isotópicas Sm-Nd E U-Pb De Bacia Colisional Tipo Flysch. **Geonomos**, v. 19, p. 90–99, 2013. doi: 10.18285/geonomos.v19i2.45.

DUDA, B. M.; WEINDORF, D. C.; CHAKRABORTY, S.; LI, B.; MAN, T.; PAULETTE, L.; DEB, S. Soil characterization across catenas via advanced proximal sensors. **Geoderma**, v. 298, p. 78–91, 2017. doi: 10.1016/j.geoderma.2017.03.017.

EITEN, G. Vegetação do Cerrado. In: PINTO, M. N. (Ed.). **Cerrado: caracterização, ocupação e perspectivas**. Brasília: UnB/SEMATEC, 1994. p. 17–73. 1994.

ELITH, J.; KEARNEY, M.; PHILLIPS, S. The art of modelling range-shifting species. **Methods in Ecology and Evolution**, v. 1, n. 4, p. 330–342, 2010. doi: 10.1111/j.2041-210x.2010.00036.x.

ELLILI, Y.; MALONE, B. P.; MICHOT, D.; MINASNY, B.; VINCENT, S.; WALTER, C.; LEMERCIER, B. Comparing three approaches of spatial disaggregation of legacy soil maps based on DSMART algorithm. **Soil Discussions**, n. June, p. 1–39, 2019.

EMBRAPA. **Procedimentos normativos de levantamentos pedológicos**. 1 Ed. ed. Brasília: Embrapa, 1995.

EMBRAPA. Programa Nacional de Solos do Brasil (Pronasolos). In: KER, J. C.; CURI, N.; SCHAEFER, C. E. G. R.; VIDAL-TORRADO, P. (Ed.). **Série Documentos**. 1º ed. Rio de Janeiro: Embrapa Solos, 2016. 183 p. 93. 2016.

ESRI. **ArcGIS Desktop: Release 10.2Redlands CA**, 2013. .

ETTEN, J. Van; SUMNER, M.; CHENG, J.; BASTON, D.; BEVAN, A.; BIVAND, R.; Busetto, L.; CANTY, M.; FASOLI, B.; FORREST, D.; GRAY, J.; GREENBERG, J. A.; HIEMSTRA, P.; KARNEY, C.; MATTIUZZI, M. Package ‘ raster ’ R topics documented :

2021.

FAO, F. and A. O. of U. N. **Our Soils Under Threat** International Year of Soils fao.org, , 2015. .

FELFILI, J. M.; DA SILVA, M. C. A comparative study of cerrado (sensu stricto) vegetation in Central Brazil. **Journal of Tropical Ecology**, v. 9, n. 3, p. 277–289, 1993. doi: 10.1017/S0266467400007306.

FERNANDES, M. M.; MARQUES, F. S. M. e. Dados para o mapa de solos de Minas Gerais. In: CETEC (Ed.). **Infraestrutura de Dados Espaciais do Sistema Estadual de Meio Ambiente e Recursos Hídricos**. Belo Horizonte: SISEMA, 2007. 2007.

FICK, S. E.; HIJMANS, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. **International Journal of Climatology**, v. 37, n. 12, p. 4302–4315, 2017. doi: 10.1002/joc.5086.

FLACH, C. W.; CORRÊA, E. A. Levantamento De Solos No Brasil: Métodos, Práticas E Dificuldades. **Geographia Meridionalis**, v. 3, n. 3, p. 420, 2018. doi: 10.15210/gm.v3i3.12237.

FRIEDMAN, J.; HASTIE, T.; TIBSHIRANI, R. *The Elements of Statistical Learning Preface to the Second Edition*. 2008.

GAN-LIN, Z.; FENG, L. I. U.; XIAO-DONG, S.; ZHANG, G. lin; LIU, F.; SONG, X. dong. Recent progress and future prospect of digital soil mapping: A review. **Journal of Integrative Agriculture**, v. 16, n. 12, p. 2871–2885, 2017. doi: 10.1016/S2095-3119(17)61762-3.

GATEHOUSE, R. D.; WILLIAMS, I. S.; PILLANS, B. J. Fingerprint windblow dust in south-eastern Australian soils by uranium-lead dating of detrital zircon. **Australian Journal of Soil Research**, v. 39, n. 1, p. 7–12, 2001. doi: 10.1071/SR99078.

GIAROLA, N. F. B.; CURI, N.; SIQUEIRA, J. O.; CHAGAS, C. da S.; FERREIRA, M. M. **Solos da região sob influência do reservatório da hidrelétrica de Itutinga/Camargos (MG): perspectiva ambiental**. Lavras, MG: CEMIG, 1997.

GOMES, L. C.; FARIA, R. M.; DE SOUZA, E.; VELOSO, G. V.; SCHAEFER, C. E. G. R.; FILHO, E. I. F. Modelling and mapping soil organic carbon stocks in Brazil. **Geoderma**, v. 340, n. December 2017, p. 337–350, 2019. doi: 10.1016/j.geoderma.2019.01.007.

GONÇALVES, M. G. M. **Soil Survey As Support To Precision Coffee Crop and Winter Wines Development in Southeast Lavras - Mg**. 2020. Universidade Federal de Lavras, 2020.

GRIMM, R.; BEHRENS, T.; MÄRKER, M.; ELSENBEEER, H. Soil organic carbon

concentrations and stocks on Barro Colorado Island - Digital soil mapping using Random Forests analysis. **Geoderma**, 2008. doi: 10.1016/j.geoderma.2008.05.008.

HARIDASAN, M. Aluminium accumulation by some cerrado native species of central Brazil. **Plant and Soil**, v. 65, n. 2, p. 265–273, 1982. doi: 10.1007/BF02374657.

HARTEMINK, A. E. On global soil science and regional solutions. **Geoderma Regional**, v. 5, p. 1–3, 2015. doi: 10.1016/j.geodrs.2015.02.001.

HARTEMINK, A. E.; MCBRATNEY, A. A soil science renaissance. **Geoderma**, v. 148, n. 2, p. 123–129, 2008. doi: 10.1016/j.geoderma.2008.10.006.

HEILBRON, M.; RIBEIRO, A.; VALERIANO, C. N.; PACIULLO, F. V.; ALMEIDA, J. C. H.; TROUW, R. J. A.; TUPINAMBÁ, M.; SILVA, L. G. E. The Ribeira Belt. **Springer**, p. 326, 2017. doi: 10.1007/978-3-319-01715-0_15.

HENGL, T. Finding the right pixel size. **Computers and Geosciences**, v. 32, n. 9, p. 1283–1298, 2006. doi: 10.1016/j.cageo.2005.11.008.

HENGL, T.; DE JESUS, J. M.; HEUVELINK, G. B. M.; GONZALEZ, M. R.; KILIBARDA, M.; BLAGOTIĆ, A.; SHANGGUAN, W.; WRIGHT, M. N.; GENG, X.; BAUER-MARSCHALLINGER, B.; GUEVARA, M. A.; VARGAS, R.; MACMILLAN, R. A.; BATJES, N. H.; LEENAARS, J. G. B.; RIBEIRO, E.; WHEELER, I.; MANTEL, S.; KEMPEN, B. SoilGrids250m: Global gridded soil information based on machine learning. **PLoS ONE**, v. 12, n. 2, p. 1–40, 2017. doi: 10.1371/journal.pone.0169748.

HEUNG, B.; BULMER, C. E.; SCHMIDT, M. G. Predictive soil parent material mapping at a regional-scale: A Random Forest approach. **Geoderma**, v. 214–215, p. 141–154, 2014. doi: 10.1016/j.geoderma.2013.09.016.

HEUNG, B.; HO, H. C.; ZHANG, J.; KNUDBY, A.; BULMER, C. E.; SCHMIDT, M. G. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. **Geoderma**, v. 265, p. 62–77, 2016. doi: 10.1016/j.geoderma.2015.11.014.

HIJMANS, A. R. J.; PHILLIPS, S.; LEATHWICK, J.; ELITH, J.; HIJMANS, M. R. J. Package ‘dismo’. 2020.

HOUNKPATIN, K. O. L.; SCHMIDT, K.; STUMPF, F.; FORKUOR, G.; BEHRENS, T.; SCHOLTEN, T.; AMELUNG, W.; WELP, G. Predicting reference soil groups using legacy data: A data pruning and Random Forest approach for tropical environment (Dano catchment, Burkina Faso). **Scientific Reports**, v. 8, n. 1, p. 1–16, 2018. doi: 10.1038/s41598-018-28244-w.

HUDSON, H. D. Division s-5-soil genesis, morphology & classification. v. 841, p. 836–841, 1992.

IAEA. **Guidelines for Radioelement Mapping Using Gamma-ray Spectrometry Data**. International Atomic Energy Agency, Austria, 2003.

IBGE. Manual técnico de pedologia. In: **Manuais Técnicos em Geociências**. 2. ed. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística - IBGE, 2007. p. 316. 2007.

INDA JUNIOR, A. V.; KÄMPF, N. Variabilidade de goethita e hematita via dissolução redutiva em solos de região tropical e subtropical. **Revista Brasileira de Ciência do Solo**, v. 29, n. 6, p. 851–866, 2005. doi: 10.1590/s0100-06832005000600003.

JAFARI, A.; AYOUBI, S.; KHADEMI, H.; FINKE, P. A.; TOOMANIAN, N. Selection of a taxonomic level for soil mapping using diversity and map purity indices: A case study from an Iranian arid region. **Geomorphology**, v. 201, p. 86–97, 2013. doi: 10.1016/j.geomorph.2013.06.010.

JAFARI, A.; KHADEMI, H.; FINKE, P. A.; VAN DE WAUW, J.; AYOUBI, S. Spatial prediction of soil great groups by boosted regression trees using a limited point dataset in an arid region, southeastern Iran. **Geoderma**, 2014. doi: 10.1016/j.geoderma.2014.04.029.

JENNY, H. **Factors of soil formation: a system of quantitative pedology**. 1. ed. New York: Soil Science, 1941.

JOHNSON, D. L.; WATSON-STEGNER, D. Evolution model of pedogenesis. **Soil Science**, v. 143, n. 5, p. 349–366, 1987. doi: 10.1097/00010694-198705000-00005.

KÄMPF, N.; CURI, N. Conceito de solo e sua evolução histórica. In: KER, J. C.; CURI, N.; SCHAEFER, C. E. G. R.; VIDAL-TORRADO, P. (Ed.). **Pedologia: Fundamentos**. 1. ed. Viçosa: Sociedade Brasileira de Ciência do Solo, 2012a. p. 343. 2012.

KÄMPF, N.; CURI, N. Formação e evolução do solo (Pedogênese). In: KER, J. C.; CURI, N.; SCHAEFER, C. E. G. R.; VIDAL-TORRADO, P. (Ed.). **Pedologia: Fundamentos**. 1. ed. Viçosa - MG: Sociedade Brasileira de Ciência do Solo, 2012b. p. 207–294. 2012.

KER, J. C. Latossolos Do Brasil: Uma Revisão. **Geonomos**, v. 5, n. 1, p. 17–40, 1997. doi: 10.18285/geonomos.v5i1.187.

KER, J. C.; MOTTA, P. E.; OLIVEIRA, V. Á. de. Levantamentos pedológicos e evolução do conhecimento dos solos no Brasil. In: CURI, N.; KER, J.; NOVAIS, R.; VIDAL-TORRADO, P.; SCHAEFER, C. (Ed.). **Pedologia: Solos dos biomas brasileiros**. 1 ed. ed. Viçosa: Sociedade Brasileira de Ciência do Solo, 2017. p. 1–45. 2017.

KHALEDIAN, Y.; MILLER, B. A. Selecting appropriate machine learning methods for digital soil mapping. **Applied Mathematical Modelling**, v. 81, p. 401–418, 2020. doi: 10.1016/j.apm.2019.12.016.

KIDD, D.; SEARLE, R.; GRUNDY, M.; MCBRATNEY, A.; ROBINSON, N.; O'BRIEN, L.;

ZUND, P.; ARROUAYS, D.; THOMAS, M.; PADARIAN, J.; JONES, E.; BENNETT, J. M. L.; MINASNY, B.; HOLMES, K.; MALONE, B. P.; LIDDICOAT, C.; MEIER, E. A.; STOCKMANN, U.; WILSON, P.; WILFORD, J.; PAYNE, J.; RINGROSE-VOASE, A.; SLATER, B.; ODGERS, N.; GRAY, J.; VAN GOOL, D.; ANDREWS, K.; HARMS, B.; STOWER, L.; TRIANTAFILIS, J. Operationalising digital soil mapping – Lessons from Australia. **Geoderma Regional**, v. 23, p. e00335, 2020. doi: 10.1016/j.geodrs.2020.e00335.

KRAMM, T.; HOFFMEISTER, D. Assessing the influence of environmental factors and datasets on soil type prediction with two machine learning algorithms in a heterogeneous area in the Rur catchment, Germany. **Geoderma Regional**, v. 22, p. e00316, 2020. doi: 10.1016/j.geodrs.2020.e00316.

LAGACHERIE, P.; MCBRATNEY, A. B. Spacial soil information systems and spatial inference systems: perspectives for digital soil mapping. In: LAGACHERIE, P.; MCBRATNEY, A. B.; VOLTZ, M. (Ed.). **Developments in Soil Science**. 1. ed. New York: Elsevier B.V., 2007. 2007.

LAHJOUJ, A.; EL HMAIDI, A.; BOUHAFSA, K.; BOUFALA, M. Mapping specific groundwater vulnerability to nitrate using random forest: case of Sais basin, Morocco. **Modeling Earth Systems and Environment**, v. 6, n. 3, p. 1451–1466, 2020. doi: 10.1007/s40808-020-00761-6.

LEENAARS, J. G. B.; ELIAS, E.; WÖSTEN, J. H. M.; RUIPEREZ-GONZÁLEZ, M.; KEMPEN, B. Mapping the major soil-landscape resources of the Ethiopian Highlands using random forest. **Geoderma**, v. 361, n. October 2019, p. 114067, 2020. doi: 10.1016/j.geoderma.2019.114067.

LEGROS, J.-P. **Mapping of the soil**. 1st. ed. New Hampshire: Publishers, Science, 2006.

LI, Y.; ZHU, A.; SHI, Z.; LIU, J.; DU, F. Geoderma Supplemental sampling for digital soil mapping based on prediction uncertainty from both the feature domain and the spatial domain. **Geoderma**, v. 284, p. 73–84, 2016. doi: 10.1016/j.geoderma.2016.08.013.

MA, Y.; MINASNY, B.; MALONE, B. P.; MCBRATNEY, A. B. Pedology and digital soil mapping (DSM). **European Journal of Soil Science**, v. 70, n. 2, p. 216–235, 2019. doi: 10.1111/ejss.12790.

MACHADO, D. F. T. **Accuracy and uncertainty of digital soil mapping approaches to extract and transfer soil information from reference area**. 2017. Universidade Federal de Lavras, 2017.

MACHADO, D. F. T.; DE MENEZES, M. D.; SILVA, S. H. G.; CURI, N. Transferability, accuracy, and uncertainty assessment of different knowledge-based approaches for soil types mapping. **Catena**, v. 182, n. September 2018, p. 104134, 2019a. doi: 10.1016/j.catena.2019.104134.

- MACHADO, D. F. T.; SILVA, S. H. G.; CURI, N.; DE MENEZES, M. D. Soil type spatial prediction from random forest: Different training datasets, transferability, accuracy and uncertainty assessment. **Scientia Agricola**, v. 76, n. 3, p. 243–254, 2019b. doi: 10.1590/1678-992x-2017-0300.
- MANCINI, M.; WEINDORF, D. C.; SILVA, S. H. G.; CHAKRABORTY, S.; TEIXEIRA, A. F. dos S.; GUILHERME, L. R. G.; CURI, N. Parent material distribution mapping from tropical soils data via machine learning and portable X-ray fluorescence (pXRF) spectrometry in Brazil. **Geoderma**, v. 354, n. July, p. 113885, 2019. doi: 10.1016/j.geoderma.2019.113885.
- MCBRATNEY, A. . B.; MENDONÇA SANTOS, M. . L.; MINASNY, B. On digital soil mapping. **Geoderma**, v. 117, n. 1–2, p. 3–52, 1 nov. 2003. doi: 10.1016/S0016-7061(03)00223-4.
- MCBRATNEY, A. B.; ODEH, I. O. A.; BISHOP, T. F. A.; DUNBAR, M. S.; SHATAR, T. M. An overview of pedometric techniques for use in soil survey. **Geoderma**, v. 97, n. 3–4, p. 293–327, 2000. doi: 10.1016/S0016-7061(00)00043-4.
- MCBRATNEY, A.; FIELD, D. J.; KOCH, A. The dimensions of soil security. **Geoderma**, v. 213, p. 203–213, 2014. doi: 10.1016/j.geoderma.2013.08.013.
- MELLO, D. C. de; DEMATTÊ, J. A. M.; MELLO, F. A. de O.; POPPIEL, R. R.; SILVERO, N. E.; SAFANELLI, J. L.; SOUZA, A. B. e; RAIMO, L. A. Di; RIZZO, R.; RESENDE, M. E. B.; GONÇALVES, S. E. Applied gamma-ray spectrometry for evaluating tropical soil processes and attributes. **Geoderma**, v. 381, n. September 2020, p. 114736, 2021. doi: 10.1016/j.geoderma.2020.114736.
- MENDONÇA-SANTOS, M. de L.; SANTOS, H. G. dos. Mapeamento Digital de Classes e Atributos de Solos métodos, paradigmas e novas técnicas. **Embrapa Solos**, p. 17, 2003.
- MENDONÇA-SANTOS, M. L.; SANTOS, H. The state of the art of Brazilian soil mapping and prospects for digital soil mapping. In: LAGACHERIE, P.; MCBRATNEY, A. B.; VOLTZ, M. (Ed.). **Digital Soil Mapping: An introductory perspective**. New York, NY: Elsevier, 2007. 2007.
- MENEZES, M. D. de. **Levantamento de solos em sub-bacias hidrográficas como apoio para avaliação do uso da terra e da potencialidade de recarga de aquíferos**. 2007. UFLA, 2007.
- MENEZES, M. D. de; BISPO, F. H. A.; FARIA, W. M.; GONÇALVES, M. G. M.; CURI, N.; GUILHERME, L. R. G. Modeling arsenic content in Brazilian soils: What is relevant? **Science of the Total Environment**, v. 712, p. 136511, 2020. doi: 10.1016/j.scitotenv.2020.136511.
- MENEZES, M. D. de; MESSIAS, C. G.; SILVA, S. H. G.; CURI, N. Relações entre ambientes de solos e fitofisionomias. In: CARVALHO, D. A. DE (Ed.). **Cerrados do Sul e**

Sudoeste de Minas Gerais: Flora e Ambiente. Lavras: Ed. UFLA, 2017. p. 123–147. 2017.

MENEZES, M. D. de; SILVA, S. H. G.; OWENS, P. R.; CURI, N. Digital soil mapping approach based on fuzzy logic and field expert knowledge. **Ciência e Agrotecnologia**, v. 37, n. 4, p. 287–298, 2013. doi: 10.1590/s1413-70542013000400001.

MILLARD, K.; RICHARDSON, M. On the importance of training data sample selection in Random Forest image classification: A case study in peatland ecosystem mapping. **Remote Sensing**, v. 7, n. 7, p. 8489–8515, 2015. doi: 10.3390/rs70708489.

MINASNY, B.; MCBRATNEY, A. B. Digital soil mapping: A brief history and some lessons. **Geoderma**, v. 264, p. 301–311, 2016. doi: 10.1016/j.geoderma.2015.07.017.

MOONJUN, R.; SHRESTHA, D. P.; JETTEN, V. G.; VAN RUITENBEEK, F. J. A. Application of airborne gamma-ray imagery to assist soil survey: A case study from Thailand. **Geoderma**, v. 289, p. 196–212, 2017. doi: 10.1016/j.geoderma.2016.10.035.

MOTTA, P. E. F. da; BARUQUI, A. M.; SANTOS, H. G. DOS. **Levantamento de Reconhecimento de Média Intensidade dos Solos da Região do Alto Paranaíba, Minas Gerais.** 1º ed. Rio de Janeiro: Embrapa Solos, 2004.

OLIVEIRA, V. Á. de; JACOMINE, P. K. T.; COUTO, E. G. Solos do bioma Cerrado. In: CURI, N.; KER, J. C.; NOVAIS, R. F.; VIDAL-TORRADO, P.; SCHAEFER, C. E. G. R. (Ed.). **Pedologia: Solos dos Biomas Brasileiros.** 1º ed. Viçosa, MG: Sociedade Brasileira de Ciência do Solo, 2017. p. 178–226. 2017.

PELEGRINO, M. H. P.; SILVA, S. H. G.; MENEZES, M. D. de; SILVA, E. da; OWENS, P. R.; CURI, N. Mapping soils in two watersheds using legacy data and extrapolation for similar surrounding areas. **Ciência e Agrotecnologia**, v. 40, n. 5, p. 534–546, 2016. doi: 10.1590/1413-70542016405011416.

R CORE TEAM. **R: A Language and environment for statistical computing** Viena, Áustria R Foundation for Estatistical Computing, , 2020. .

R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing.** Viena, Áustria, 2021. .

READ, C. F.; DUNCAN, D. H.; HO, C. Y. C.; WHITE, M.; VESK, P. A. Useful surrogates of soil texture for plant ecologists from airborne gamma-ray detection. **Ecology and Evolution**, v. 8, n. 4, p. 1974–1983, 2018. doi: 10.1002/ece3.3417.

REINHARDT, N.; HERRMANN, L. Gamma-ray spectrometry as versatile tool in soil science: A critical review. **Journal of Plant Nutrition and Soil Science**, v. 182, n. 1, p. 9–27, 2019. doi: 10.1002/jpln.201700447.

RESENDE, M.; CURI, N.; REZENDE, S. A.; CORREA, G. **Pedologia: base para a**

distinção de ambientes. 6. ed. Lavras: Editora UFLA, 2014.

RESENDE, M.; CURI, N.; REZENDE, S. B.; SILVA, S. H. G. **Da rocha ao solo: enfoque ambiental.** 1. ed. Lavras: UFLA, Ed., 2019.

RIBEIRO, J. F.; WALTER, B. M. T. As principais fitofisionomias do bioma Cerrado. In: **Cerrado: Ecologia e Flora.** vol. 1 ed. Brasília: Embrapa Informação Tecnológica, 2008. p. 151–212. 2008.

RIKHVANOV, L. P. Using Radioactive Elements and the Th/U Ratio in Study of the Geochemical Typification of Granitoids and Their Intrusive Character. **Russian Geology and Geophysics**, v. 60, n. 9, p. 1018–1025, 1 set. 2019. doi: 10.15372/RGG2019067.

RODRIGUES, T. E. **Caracterização e gênese de solos Brunos do Maciço Alcalino de Poços de Caldas(MG).** 1984. Universidade de São Paulo - USP, Escola Superior Luíz de Queiróz, 1984.

ROUZE, G. S.; MORGAN, C. L. S.; MCBRATNEY, A. B. Understanding the utility of aerial gamma radiometrics for mapping soil properties through proximal gamma surveys. **Geoderma**, v. 289, p. 185–195, 2017. doi: 10.1016/j.geoderma.2016.12.004.

SANTOS, H. G. dos; ALMEIDA, J. A.; OLIVEIRA, J. B. de; LUMBRERAS, J. F.; ANJOS, L. H. C. dos; COELHO, M. R.; JACOMINE, P. K. T.; CUNHA, T. J. F.; OLIVEIRA, V. Á. de; EMBRAPA. **Sistema brasileiro de classificação de solos.** [s.l: s.n.]

SANTOS, H. G. dos; LUIZ DIAMANTE AGLIO, M.; DE OLIVEIRA DART, R.; DE LOURDES BREFFIN, M. M.; SILVA DE SOUZA, J. Distribuição Espacial dos Níveis de Levantamento de Solos no Brasil. **XXXIV Congresso Brasileiro de Ciencia do Solo**, n. 2, p. 1–4, 2013.

SCHAETZL, R.; ANDERSON, S. **Soils Genesis and Geomorphology.** 1. ed. New York: Cambridge University, 2005.

SHANNON, C. E. Cominunication Theory in the presence of Noise. **Proceedings of the IRE**, v. 37, n. 1, p. 10–21, 1949.

SHARIFIFAR, A.; SARMADIAN, F.; MALONE, B. P.; MINASNY, B. Addressing the issue of digital mapping of soil classes with imbalanced class observations. **Geoderma**, v. 350, n. October 2018, p. 84–92, 2019. doi: 10.1016/j.geoderma.2019.05.016.

SHAW, C. F. What Characteristics Distinguish Pedalfers from Pedocals? **Soil Science Society of America Journal**, v. B16, n. 2001, p. 5, 2010. doi: 10.2136/sssaj1935.036159950b1620010003x.

SILVA, B. P. C.; SILVA, M. L. N.; AVALOS, F. A. P.; DE MENEZES, M. D.; CURI, N. Digital soil mapping including additional point sampling in Posses ecosystem services pilot

watershed, southeastern Brazil. **Scientific Reports**, v. 9, n. 1, p. 1–12, 2019. doi: 10.1038/s41598-019-50376-w.

SILVA, E. da. **Mapeamento de solos e uso de algoritmos de aprendizagem em Lavras (MG)**. 2018. Universidade Federal de Lavras, 2018.

SILVA, S. H. G.; MENEZES, M. D. de; OWENS, P. R.; CURI, N. 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**, v. 267, p. 65–77, 2016. doi: 10.1016/j.geoderma.2015.12.025.

SILVA, V. A.; CURI, N.; MARQUES, J. J. G.; DE CARVALHO, L. M. T.; DOS SANTOS, W. J. R. Mapa de solos, conhecimento de campo, inventário florestal e zoneamento ecológico-econômico como base para a aptidão agrícola das terras em Minas Gerais elaborada em SIG. **Ciencia e Agrotecnologia**, v. 37, n. 6, p. 538–549, 2013. doi: 10.1590/S1413-70542013000600007.

SKORUPA, A. L.; GUILHERME, L. R. G.; CURI, N.; SILVA, C. P. de C.; SCOLFORO, J. R. S.; MELO MARQUES, J. J. G. de S. e M. Propriedades de Solos sob Vegetação Nativa em Minas Gerais: Distribuição por Fitofisionomia. **Revista Brasileira de Ciencia do Solo**, v. 36, n. 3, p. 11–22, 2012.

SMETHURST, M. A. A software tool to rotate spatial data on the surface of the sphere (Earth) for Geosoft's Oasis Montaj. **NGU Report**, 2005.

STAHR, K.; CLEMENS, G.; SCHULER, U.; ERBE, P.; HAERING, V.; CONG, N. D.; BOCK, M.; TUAN, T. D.; HAGEL, H.; VINH, B. L.; RANGUBPIT, W.; SURINKUM, A.; WILLER, J.; INGWERSEN, J.; ZAREI, M.; HERRMANN, L. Beyond the horizons: Challenges and prospects for soil science and soil care in Southeast Asia. In: FRÖHLICH, H. L.; SCHREINEMACHERS, P.; STAHR, K.; CLEMENS, G. (Ed.). **Southeast Asia: Innovations and Policies for Mountainous Areas**. 1. ed. Berlin: Springer, 2013. p. 31–107. 2013.

SURVEY UNITED STATES GEOLOGICAL. **Digital Elevation - Shuttle Radar Topography Mission (SRTM)**.

TAYLOR, K. E.; STOUFFER, R. J.; MEEHL, G. A. An overview of CMIP5 and the experiment design. **Bulletin of the American Meteorological Society**, v. 93, n. 4, p. 485–498, 2012. doi: 10.1175/BAMS-D-11-00094.1.

TEN CATEN, A.; DALMOLIN, R. S. D.; MENDONÇA-SANTOS, M. de L.; GIASSON, E. Mapeamento digital de classes de solos: características da abordagem brasileira. **Ciência Rural**, v. 42, n. 11, p. 1989–1997, 2012. doi: 10.1590/s0103-84782012001100013.

TENG, H. T.; VISCARRA ROSSEL, R. A.; SHI, Z.; BEHRENS, T. Updating a national soil classification with spectroscopic predictions and digital soil mapping. **Catena**, v. 164, n.

October 2017, p. 125–134, 2018. doi: 10.1016/j.catena.2018.01.015.

ULBRICH, H. H. G. J.; GOMES, C. B. Alkaline rocks from continental Brazil. **Earth Science Reviews**, v. 17, n. 1–2, p. 135–154, 1981. doi: 10.1016/0012-8252(81)90009-X.

VASQUES, G. de M.; DART, R. de O.; BACA, J. F. M.; CEDDIA, M. B.; MENDONÇA-; SANTOS, M. de L. **Estoque de Carbono do Brasil (0-30cm)**. 1. ed. Rio de Janeiro: Embrapa Solos, 2017.

VAYSSE, K.; LAGACHERIE, P. Evaluating Digital Soil Mapping approaches for mapping GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France). **Geoderma Regional**, v. 4, p. 20–30, 2015. doi: 10.1016/j.geodrs.2014.11.003.

WADOUX, A. M. J. C.; SAMUEL-ROSA, A.; POGGIO, L.; MULDER, V. L. A note on knowledge discovery and machine learning in digital soil mapping. **European Journal of Soil Science**, v. 71, n. 2, p. 133–136, 2020. doi: 10.1111/ejss.12909.

WILFORD, J. A weathering intensity index for the Australian continent using airborne gamma-ray spectrometry and digital terrain analysis. **Geoderma**, v. 183–184, p. 124–142, 2012. doi: 10.1016/j.geoderma.2010.12.022.

WILFORD, J.; MINTY, B. Chapter 16 The Use of Airborne Gamma-ray Imagery for Mapping Soils and Understanding Landscape Processes. **Developments in Soil Science**, v. 31, n. C, 2006. doi: 10.1016/S0166-2481(06)31016-1.

YANG, L.; JIAO, Y.; FAHMY, S.; ZHU, A.-X.; HANN, S.; BURT, J. E.; QI, F. Updating Conventional Soil Maps through Digital Soil Mapping. **Soil Science Society of America Journal**, v. 75, n. 3, p. 1044, 24 maio 2011. doi: 10.2136/sssaj2010.0002.

XU, Tingbao; HUTCHINSON, Michael F. New developments and applications in the ANUCLIM spatial climatic and bioclimatic modelling package. **Environmental Modelling & Software**, v. 40, p. 267-279, 2013.