

DESEMPENHO DOS DADOS SOILGRIDS NA REPRESENTAÇÃO DE VARIÁVEIS DOS SOLOS NA REGIÃO NORTE DE MINAS GERAIS

Luiza Rigoni Leão ¹
João Paulo Sena Souza ²

INTRODUÇÃO

O solo é reconhecido como importante agente ecossistêmico, tanto para a produção agrícola quanto para questões de regulação climática. Ao longo dos últimos anos a demanda por informações sobre o solo cresceu em todo o mundo, exigindo o aperfeiçoamento de técnicas de mapeamento (CATEN, et al., 2012). O Mapeamento Digital de Solos (MDS) vem se tornando uma importante ferramenta para a criação e expansão de modelos de informação espacial do solo. Este funciona a partir de modelos numéricos que aferem e predizem variações temporais dos tipos e propriedades dos solos com base em amostras destes com localização conhecida e suas relações com variáveis ambientais (LAGACHERIE, 2008). O MDS se diferencia do Mapeamento Convencional de Solos (MCS) por seu caráter quantitativo e baixo custo de implementação já o MCS é qualitativo e demanda do conhecimento especializado do mapeador (KEMPEN et al., 2011).

Existem diversos esforços para o mapeamento do solo em escala global. Um exemplo é a base de dados do SoilGrids, um sistema digital global que usa de métodos de machine learning na predição espacial de variáveis do solo (HENGL et al., 2017). Para elaborar os modelos SoilGrids, os autores utilizaram 240 mil perfis observados ao redor do mundo, e mais de 400 variáveis preditoras associadas ao clima, cobertura do solo, hidrologia e morfologia do terreno (HENGL et al., 2017). A partir deste banco de dados, podem ser gerados mapas com resolução espacial de 250 metros com uma grande diversidade de variáveis pedológicas, todos disponíveis para download gratuito em formato raster (POGGIO et al., 2021).

¹ Mestranda em geografia da Universidade Estadual de Montes Claros - UNIMONTES, luizarigoni@gmail.com;

² Professor orientador: Doutor, Universidade Estadual de Montes Claros - UNIMONTES, joão.souza@unimontes.com.

Apesar da grande utilidade destes modelos, existem incertezas inerentes aos métodos de predição espacial usados pelos autores. Essas incertezas globais tendem a aumentar quando os dados são recortados para a escala regional, principalmente em áreas de transição entre biomas. É o caso da Mesorregião do Norte de Minas Gerais, onde ocorre transições entre Caatinga, Cerrado e Mata Atlântica (IBGE, 2002; GUIMARÃES et al., 2010), além de contar com grandes áreas de Floresta Estacional Decidual (OLSON et al. 2001). Neste contexto, o objetivo deste estudo é avaliar a acurácia dos modelos disponíveis na base de dados do SoilGrids em representar atributos do solo superficial do Norte de Minas Gerais. Os resultados desse estudo devem estabelecer o quão bem esses modelos representam o solo da região, tendo em vista a complexa situação ambiental da área.

MATERIAIS E MÉTODOS

Área de estudos

A Mesorregião do Norte de Minas é composta por 89 municípios, organizados no que é a maior mesorregião em extensão territorial (128.602km²) do Estado. Destaca-se por apresentar alta complexidade de ecossistemas tendo em vista a existência de regiões de transição entre os biomas Cerrado, Caatinga e Mata Atlântica com destaque a Floresta Estacional Decidual (Mata Seca). Observa-se a transição do clima tropical semiúmido para o semiárido. Uma característica dessa região é a escassez de precipitação durante a maior parte do ano, com média de precipitação anual entre 650 mm e 1.050 mm e temperaturas médias anuais que variam entre 21 °C e 24 °C (IBGE, 2002; GUIMARAES et al., 2010). O relevo da região é composto por chapadas, serras e depressões com destaque para a Depressão do São Francisco de centro a oeste, a Serra do Espinhaço a leste e as chapadas do São Francisco a oeste. Por se tratar de uma área de transição ecológica e escassez hídrica, é uma região de alta fragilidade ambiental (PEREIRA, 2006).

Procedimentos metodológicos

Foram coletadas 50 amostras de solo ao longo de rodovias entre os municípios de Claro dos Poções-MG e Manga-MG. Do total, 39 amostras foram coletadas em áreas de vegetação preservada e 11 em pastagens. As amostras foram coletadas com um trado

holandês na profundidade 0-20 cm. As amostras foram secas ao ar, destorroadas e passadas na peneira de 2 mm de abertura para obtenção da Terra Fina Seca ao Ar (TFSA) no Laboratório de Geografia Física da Universidade Estadual de Montes Claros. Posteriormente, as amostras foram enviadas ao Laboratório de Análises de Solo da Universidade Federal de Lavras para as análises físicas e químicas de rotina. Para este estudo, as variáveis utilizadas foram teor de areia, argila, silte e pH.

Dos dados primários gerados foram selecionados aqueles que possuem modelos de predição feitos pelo projeto SoilGrids250m, sendo estes: areia, argila, silte e pH. Os dados do SoilGrids estão disponíveis em soilgrids.org. Apenas os modelos correspondentes às camadas superficiais do solo do SoilGrids foram usados na avaliação de desempenho. Os valores dos modelos SoilGrids foram extraídos para as coordenadas geográficas dos dados amostrados e organizados em uma planilha única pela unidade de medida decagrama por quilograma (dag/kg) para a granulometria.

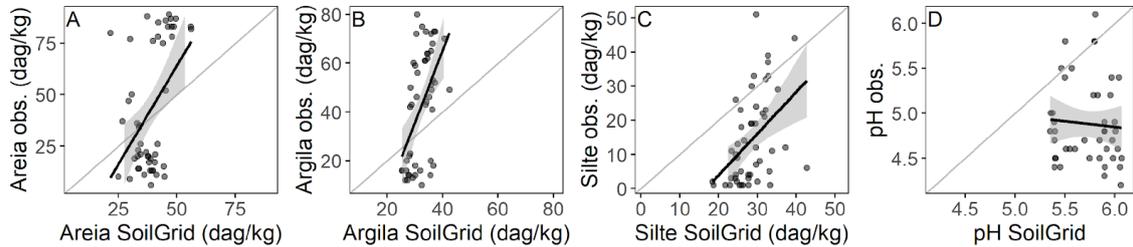
Os valores dos modelos do SoilGrids foram extraídos para cada ponto amostrado no Norte de Minas Gerais para formar uma matriz de regressão. O desempenho dos modelos foi medido por meio de análise da regressão linear para confrontar os dados observados com os dados preditos pelo modelo. A partir da regressão linear podemos recolher dados de desempenho, como o coeficiente de determinação (R^2) e a Raiz do Erro Quadrático Médio (RMSE – sigla em inglês), que indica uma média do erro do dado observado em relação ao dado extraído do modelo. Foi realizada a subtração dos valores do SoilGrids com os dados observados para demonstrar o padrão espacial dos erros de predição, indicando áreas onde os valores são subestimados ou superestimados. Todos os procedimentos estatísticos foram realizados no software R (R Core Team, 2024).

RESULTADOS E DISCUSSÃO

A variabilidade dos dados dos modelos é relevantemente menor do que os encontrados nos dados primários, o que nos sugere que os modelos não estão capturando características relevantes para a predição na área de estudo.

(Fig.1)

Fig.1 Gráficos de regressão linear. Y referente aos dados primários analisados e X referente aos dados correspondentes no SoilGrids. Área em cinza referente ao intervalo de confiança de 95%. A) Regressão linear da areia. B) Regressão linear da argila. C) Regressão linear do Silte. D) Regressão linear do pH.



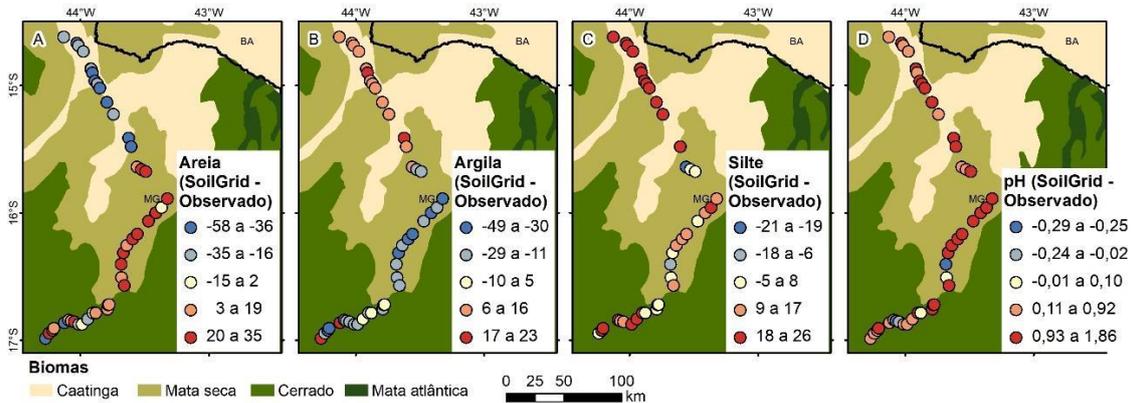
Os R^2 dos modelos lineares entre as variáveis de textura do solo estão baixos, mas explicam estatisticamente correlação entre os dados confrontados. Para os dados de Areia, o valor observado máximo foi de 89 dag/kg e mínimo 6 dag/kg, com uma média de 42,88 dag/kg. Já no modelo o máximo encontrado foi de 56,10 dag/kg e o mínimo de 21,60 dag/kg, com uma média de 38,94 dag/kg. O desempenho do SoilGrids em representar os dados reais foi $R^2 = 0,21$ e $RMSE = 28,11$ dag/kg.

Para a argila os dados primários apresentaram um máximo de 80 dag/kg e mínimo de 10 dag/kg, com uma média de 42.96 dag/kg. No SoilGrids, o máximo encontrado foi de 42,48 e mínima de 25,35, com uma média de 32.52. O desempenho do Soilgrid em representar os dados de Argila foi $R^2 = 0,27$ e $RMSE = 23,2$ dag/kg.

Para representar a distribuição espacial do Silte, o Soilrid também apresentou um baixo desempenho, com $R^2 = 0,20$ e $RMSE = 18,46$ dag/kg. O pior desempenho do modelo foi com dados de pH, com $R^2 = 0,007$ e $RMSE = 1,02$.

É possível notar um padrão espacial nos erros de predição do SoilGrids (Fig. 2 A). Para areia, o modelo superestimou os valores obtidos na Mata Seca e subestimou os da Caatinga. Para os teores de argila, a tendência é a inversa, com predições que superestimaram os valores observados na Caatinga e os subestimaram na Mata Seca. O Cerrado apresentou uma maior aleatoriedade dos erros de predição do SoilGrids, tanto para a areia quanto para a argila (Fig. 2 B). Teores de Silte foram superestimados pelo SoilGrid nos pontos referentes a Caatinga e ao Cerrado (Fig. 2 C). O pH foi subestimado pelo SoilGrid em apenas dois dos 50 pontos amostrados. A maior parte dos pontos foram superestimados pelo SoilGrid, com diferenças variando de 0,93 a 1,86, significativos para o pH do solo.

Fig.2 Mapas da das diferenças entre os valores do SoilGrid e os valores observados.



CONSIDERAÇÕES FINAIS

Tendo em vista os dados apresentados e discutidos podemos concluir que os modelos analisados apresentam baixa acurácia para a área de estudo. Embora os modelos tenham apresentado um coeficiente de determinação baixo com os dados primários observados, as correlações lineares foram significativas, o que indica que estes modelos podem ser usados para melhorar futuros mapas para a região, com exceção do pH. Entretanto, os resultados mostraram que modelos globais devem ser usados com cautela para a elaboração de estudos do recorte devido à grande discrepância entre valores preditos e valores reais. Aumentar a amostragem de solos em regiões com alta complexidade ecológica pode ajudar a melhorar a acurácia de modelos dessas áreas.

Palavras-chave: Mapeamento Digital de Solo; Modelos Espaciais; Validação.

AGRADECIMENTOS

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais - FAPEMIG pelos financiamentos de pesquisa no âmbito do Edital nº 001/2021 Universal (Processo APQ02912-21) e do Edital nº 004/2023 - Centros de Tecnologia e Infraestrutura para Pesquisa na UEMG e UNIMONTES (Processo APQ-03759-23); e pelo apoio a

participação coletiva em eventos de caráter técnico-científico no país com a Chamada nº 13-2023 (Processo PCE-00393-24).

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES).

REFERÊNCIAS

CATEN, A. et al. Mapeamento digital de classes de solos: características da abordagem brasileira. *Ciência Rural*, v. 42, n. 11, p. 1989–1997, 1 nov. 2012.

GUIMARAES, D.P. et al. Índices pluviométricos em Minas Gerais. Embrapa Milho e Sorgo. Sete Lagoas, 2010.

HENGL, T. et al. SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, v. 12, n. 2, p. e0169748, 16 fev. 2017.

IBGE. Mapa de clima do Brasil. IBGE. Rio de Janeiro, 2002.

KEMPEN, B. et al. Efficiency comparison of conventional and digital soil mapping for updating soil maps. *Soil Sci. Soc. Am. J.* v.76, p. 2097-2115, 2011.

LAGACHERIE, P. Digital Soil Mapping: a state of the art. In: HARTEMINK, A. E.; McBRATNEY, A. B.; MENDONÇA-SANTOS, M. L. (Ed.). *Digital Soil Mapping with Limited Data*. Wageningen: Springer, 2008. p. 3-14.

OLSON, D.M. et al. Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience*, v. 51, n. 11, 2001.

PEREIRA, A. M. Múltiplos olhares sobre a região Norte de Minas. *Revista Cerrados*, [S. l.], v. 4, n. 01, p. 23–42, 2006.

POGGIO, L. et al. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty, *SOIL*, 7, 217–240, 2021.

R CORE TEAM. R: A Language and Environment for Statistical Computing. Versão 4.3.2. R Foundation for Statistical Computing, Vienna, Austria, 2023. Software de internet. Disponível em: <https://www.R-project.org/>. Acesso em: 07 mar. 2024.