

IMAGENS SINTÉTICAS COMO SUBSÍDIO AO MAPEAMENTO DO TEOR DE ARGILA SUPERFICIAL DOS SOLOS

Waldir de Carvalho Junior¹, Adinan Marzulo Maia Martins², Ricardo de Oliveira Dart¹, Nilson Rendeiro Pereira¹, José Ronaldo Macedo¹, Silvio Barge Bhering¹, Gustavo M. Vasques¹, César da Silva Chagas¹ e Braz Calderano Filho¹

¹Embrapa Solos, Rua Jardim Botânico, 1024. Rio de Janeiro – RJ. 22460-000, {waldir.carvalho; ricardo.dart; nilson.pereira; jose.ronaldo; silvio.bhering; gustavo.vasques; cesar.chagas; braz.calderano}@embrapa.br; ² Universidade Federal do Rio de Janeiro, Avenida Athos da Silveira Ramos, 274, Cidade Universitária, Rio de Janeiro – RJ, 27941-853, adinanmaia@gmail.com

RESUMO

Imagens sintéticas podem ser geradas sobre coleções de imagens visando obter composições onde os pixels representam o solo exposto. A correlação destas imagens com dados de solos foi possível, demonstrando boa correlação e bons resultados na aplicação de aprendizado de máquina. O R2 de 0,64 mostrou ser promissor o uso desta técnica para mapeamento digital de solos e para interpretações com finalidades de manejo adequado e sustentável dos solos. Os mapas digitais médios de distribuição de argila superficial dos solos dos municípios de Inocência e Rio Brilhante no MS, foram gerados e as principais covariáveis foram as bandas 5, 6, 7 e o NDVI.

Palavras-chave — aprendizado de máquina, mapeamento digital de solos, randomForest.

ABSTRACT

Synthetic images can be generated over image collections in order to obtain compositions where the pixels represent the barest soil. The correlation of these images with soil data was possible, demonstrating good correlation and good results in the application of machine learning. The R2 of 0.64 showed to be promising the use of this technique for digital soil mapping and for interpretations with the purpose of adequate and sustainable soil management. The average digital maps of surface clay distribution of soils in the municipalities of Inocência and Rio Brilhante in MS were generated and the main covariates were bands 5, 6, 7 and the NDVI.

Key words — Machine Learning, digital soil mapping, random Forest.

1. INTRODUÇÃO

O Estado do Mato Grosso do Sul possui importante papel na produção agrícola nacional, com expectativa de produção de 21,5 milhões de toneladas de leguminosas e oleaginosas neste ano de 2022 [1]. Para manter sustentável esta produção, o

conhecimento dos solos é fundamental para um manejo adequado e manutenção dos ecossistemas.

O solo desempenha um papel fundamental nos ecossistemas, na produção de alimentos, na regulação do clima e nos ciclos da água e dos elementos [2] e [3]. Apesar de sua importância, os solos são um dos componentes ambientais mais impactados do planeta [4].

A falta de conhecimento pedológico em escala adequada levanta questões importantes sobre a utilização sustentável dos solos e exigem a aplicação de novas técnicas e ferramentas de mapeamento de solos visando suprir esta demanda [5]. Dentre as novas ferramentas podemos citar o sensoriamento remoto e as técnicas de mapeamento digital de solos (MDS) [6].

Neste contexto, o objetivo deste trabalho foi testar uma metodologia de aprendizado de máquina aplicada sobre imagens sintéticas e dados de solos, para mapear o teor de argila superficial dos solos em dois municípios do Estado do MS.

2. MATERIAL E MÉTODOS

As imagens sintéticas foram obtidas através de processamento de coleção de imagens do satélite Landsat - 8 OLI/TIRS [7] no ambiente Google Earth Engine [8]. Neste processamento foram criadas imagens sintéticas de média para os municípios de Rio Brilhante e Inocência (ambos no MS), onde os pixels são representações de solo descoberto de vegetação ou solo exposto (Figura 01). As imagens foram processadas com correção atmosférica e radiométrica para transformação do nível de cinza em reflectância.

Estas imagens foram correlacionadas com dados de solos obtidos pelo Zoneamento Agroecológico do Estado do Mato Grosso do Sul (ZAEMS).

Nos trabalhos de campo foram descritos, classificados, coletados e analisados os solos em 98 pontos (Figura 01), e as principais classes de solos encontradas em Inocência e Rio Brilhante estão apresentadas na Tabela 01, abaixo.

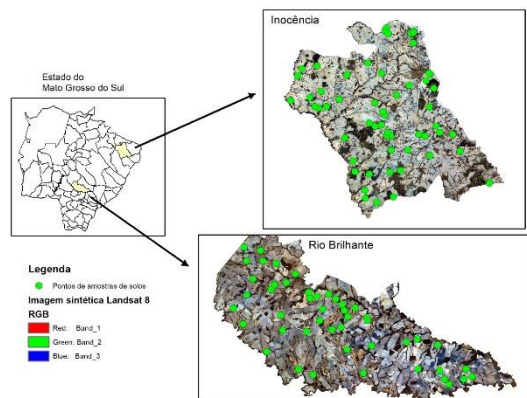


FIGURA 1: Área de Estudo com uma composição colorida das imagens sintéticas e os pontos amostrais de solos.

Classe de solo	Textura superficial	Inocência	Rio Brilhante
LVd	Média ou arenosa	15	--
LVd	Argilosa, média ou arenosa	--	04
LVdf	Argilosa ou muito argilosa	--	35
LAd	Argilosa	01	02
CXbd	Média	01	--
FXe	Arenosa	01	--
PVd	Arenosa	04	--
GMbd e GXbd	Arenosa, Média e argilosa	01	03
OOs	Média	01	--
RQo	Arenosa	30	--

Tabela 01: Classe de solos, com textura superficial e número de pontos coletados em cada município.

Onde: LVd – Latossolo Vermelho distrófico; LVdf – Latossolo Vermelho Háplico Tb distrófico; LAd – Latossolo Amarelo distrófico; CXbd – Cambissolo Háplico Tb distrófico; FXe – Plintossolo Háplico eutrófico; PVd – Argissolo Vermelho distrófico; GMbd e GXbd – Gleissolo Melânico ou Háplico Tb distrófico; OOs – Organossolo; RQo – Neossolo Quartazrênico órtico

Todo o processamento dos dados para modelagem foi feito no R [9]. As variáveis independentes (bandas da imagem sintética) foram pre-processadas para eliminar aquelas altamente correlacionadas (correlação de Spearman > |0.95|) e aquelas com variância próxima de zero.

A elaboração dos layouts das imagens foi feita utilizando-se o ArcGIS 10. Os teores de argila dos pontos amostrais de solos (Figura 01) foram correlacionados com os dados espectrais das imagens sintéticas, buscando por aprendizado de máquina elaborar um modelo de mapeamento digital baseado no algoritmo RandomForest [10].

Nesta etapa, os dados foram divididos aleatoriamente em Teste e Treino na proporção de 70 e 30%, respectivamente. Essa divisão foi repetida 100 vezes para eliminar qualquer viés na observação dos resultados. Foram usados os pacotes Caret [11] e Raster [12] para processar toda a modelagem. Foi usado o coeficiente de determinação (R2) médio, o

RMSE e o erro médio absoluto (MAE) para verificar a performance da modelagem.

Os mapas de argila foram gerados a cada repetição, e ao final do processo foi calculado o mapa médio de teor superficial de argila dos solos.

3. RESULTADOS

No município de Inocência, os solos são em sua maioria de textura superficial média ou arenosa (Figura 2), composto de Latossolos Vermelhos e Vermelhos Amarelos, além dos Neossolos Quartazrênicos (Tabela 01).

Já o município de Rio Brilhante possui em geral solos com textura superficial argilosa ou muito argilosa (Figura 2), composto principalmente por Latossolos Vermelhos Distrofêricos argilosos ou muito argilosos e Latossolos Vermelhos e Amarelos argilosos (Tabela 01).

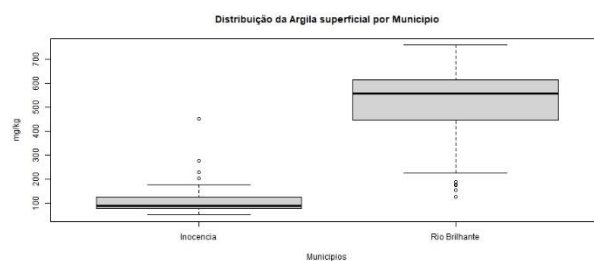


Figura 2. Boxplot da distribuição de argila superficial para os municípios estudados.

Essa distribuição diferenciada da textura superficial dos solos entre os dois municípios estudados, foi associada as diferenças de reflectância das bandas da imagem sintética, o que contribui para a correlação ambiental entre teor superficial de argila dos solos e a reflectância das bandas da imagem sintética.

No processo de aprendizado de máquina foram excluídas as covariáveis que tinham correlação maior que 0,95, além daquelas com variância zero ou próximo disso. Assim as bandas SR_B2 (alta correlação), QA_PIXEL e QA_RADSAT (variância zero) foram excluídas do processo de modelagem. Desta forma, restaram oito covariáveis, a saber SR_B1, SR_B3, SR_B4, SR_B5, SR_B6, SR_B7, ST_TRAD e NDVI.

Prosseguindo com o processo de aprendizado de máquina, foram gerados subconjuntos aleatórios de dados para treino e teste na proporção de 70/30. Os dados de treino contem 70 observações e os de treino 28. Estes dados de treino foram usados para gerar o modelo RandomForest, repetindo o processo por 100 vezes, buscando a estabilização do modelo e observação do R2, do RMSE e do MAE.

Os valores médios R2 foram de 0,64 para treino e de 0,56 para teste. O RMSE médio foi de 147 g.kg⁻¹ e 156 g.kg⁻¹ para treino e teste, respectivamente, assim como o MAE médio foi de 111 g.kg⁻¹ e 108 g.kg⁻¹. Estes valores indicam uma performance de média a alta do modelo RandomForest, porém interessante, já que foram utilizadas apenas as

covariáveis da imagem sintética, sem nenhum atributo do terreno ou covariável baseada no modelo digital de elevação.

Quanto as imagens geradas de argila (média de 100 repetições) (Figura 03), podemos notar que as áreas com vegetação permanentemente e aquelas consideradas como solo exposto tiveram predições bastante diferentes, apresentando descontinuidades que não ocorrem na natureza.

As principais covariáveis preditoras ou variáveis independentes foram, em ordem de importância, as bandas B5, B6, ST_TRAD, NDVI e B7. Na figura 04 podemos observar o comportamento médio da resposta espectral por classe de solo e notar as diferenças mais marcantes para as bandas de maior importância, ou seja, B5, B6 e NDVI.

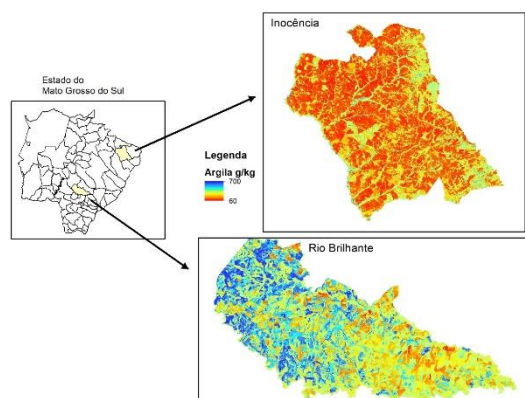


Figura 03. Mapas de distribuição de argila superficial para os municípios de Inocência e Rio Brilhante (ambos em MS).

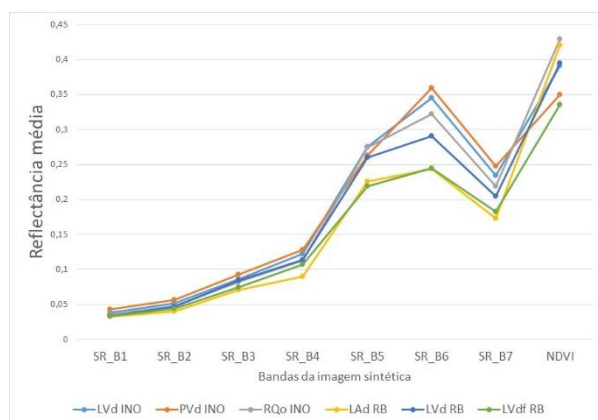


Figura 04. Resposta espectral média para as principais classes de solos dos municípios estudados.

4. DISCUSSÃO

Considerando que foram usadas como covariáveis ou variáveis independentes apenas as bandas das imagens sintéticas, verificou-se uma boa correlação entre os teores de argila do horizonte A dos solos e a reflectância dos pixels das imagens.

Durante o pré-processamento das bandas das imagens sintéticas, algumas delas foram excluídas em função de alta correlação entre elas e também devido a sua variância estar próxima de zero. Isso simplificou o modelo e tornou o processo mais rápido.

Os resultados dos índices R2, RMSE e MAE foram médios, mas mostraram a importância de se usar este tipo de imagem para mapeamento digital de atributos de solos, mormente o teor de argila superficial.

5. CONCLUSÕES

O presente estudo demonstrou a possibilidade de utilização de técnicas de aprendizado de máquina para mapeamento digital de atributos de solos, principalmente em relação a distribuição do teor de argila superficial dos solos. Estas técnicas serão usadas no projeto de ZAEMS (Zoneamento Agroecológico do Estado do Mato Grosso do Sul) para auxiliar nas interpretações para uso e manejo adequado e sustentável dos solos.

6. REFERÊNCIAS

- [1] IBGE. Levantamento Sistemático da Produção Agrícola – maio 2022. Disponível em: <https://sidra.ibge.gov.br/home/lspa/mato-grosso-do-sul>. Consultado em 13 de junho 2022.
- [2] S. Diek, F. Fornallaz, M. E. Schaepman and R. de Jong. Barest Pixel Composite for Agricultural Areas Using Landsat Time Series. *Remote Sens.* 9, 1245; doi:10.3390/rs9121245 www.mdpi.com/journal/remotesensing. 2017
- [3] J. A. M. Demattê, C. T. Fongaro, R. Rizzo, and J. L. Safanelli. Geospatial Soil Sensing System (GEOS3): A powerful data mining procedure to retrieve soil spectral reflectance from satellite images, *Remote Sensing of Environment*, Volume 212, Pages 161-175, ISSN 0034-4257, <https://doi.org/10.1016/j.rse.2018.04.047>. (<https://www.sciencedirect.com/science/article/pii/S0034425718302049>). 2018
- [4] R. Lal. Soil carbon sequestration to mitigate climate change. *Geoderma* 123, 1–22. 2004.
- [5] C. C. N. de Carvalho, F. C. Nunes, M. A. Homem Antunes, and M. C. Nolasco. Soil surveys in Brazil and perspectives in digital soil mapping. *Soil Horiz.* 56. <http://dx.doi.org/10.2136/sh14-01-0002>. 2015.
- [6] B. Stenberg, R. A. Viscarra Rossel, A. M. Mouazen, and J. Wetterlind. Visible and near infrared spectroscopy in soil science. In: Sparks, D.L. (Ed.), *Advances in Agronomy*. 2010.
- [7] USGS, United State Geological Survey. Site: Landsat 8 | U.S. Geological Survey (usgs.gov). acessado em Abril de 2022.
- [8] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, v. 202, pp. 18–27. <http://dx.doi.org/10.1016/j.geoderma.2017>.

[9] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>. 2022.

[10] L. Breiman. Random Forests. Mach. Learn. 5–32. <https://doi.org/https://doi.org/10.1023/A:1010933404324> 2001.

[11] M. Kuhn. caret: Classification and Regression Training. R package version 6.0-93. <https://CRAN.R-project.org/package=caret>. 2022.

[12] R. J. Hijmans. raster: Geographic Data Analysis and Modeling. R package version 3.5-29. <https://CRAN.R-project.org/package=raster>. 2022.