

## **MAPEAMENTO DE SOLOS NO ESTADO DO PARANÁ A PARTIR DE *MACHINE LEARNING* NO *GOOGLE EARTH ENGINE***

Bruno Aparecido da Silva <sup>1</sup>  
Altair Bennert <sup>2</sup>  
Carolina Glaeser Benincá <sup>3</sup>  
Jeferson Luiz dos Santos <sup>4</sup>  
Ericson Hideki Hayakawa <sup>5</sup>  
Marcia Regina Calegari <sup>6</sup>

### **INTRODUÇÃO**

A obtenção de informações sobre a distribuição dos solos e seus atributos (GAGKAS; LILLY; BAGGALEY, 2021) se tornou uma das prioridades científicas deste século, considerando os desafios globais, como mudanças climáticas, perda de biodiversidade e saúde do solo (ARROUAYS; DAWSON, 2022). Mesmo com o aperfeiçoamento de técnicas e métodos de mapeamento de solos (p. ex., HATEFFARD et al., 2024; MCBRATNEY et al. 2000; NGU et al., 2024; ROSIN et al., 2023; SIQUEIRA et al., 2023; ŽÍŽALA et al., 2022), a ausência de informações sobre a distribuição espacial dos solos persiste em várias regiões. No caso do Brasil, cuja economia está fortemente ligada à agricultura, a falta dessas informações em escala 1:50.000 impõe dificuldades no desenvolvimento de temas prioritários ao solo (DEMATTE et al., 2022).

Nos últimos anos, com o propósito de superar a escassez dessas informações, o uso de algoritmos de *machine learning* (ML) tem se consolidado como ferramenta utilizada para obter informações espaciais dos solos (HATEFFARD et al., 2024; NGU et al., 2024). Nesse contexto, a plataforma Google Earth Engine (GEE), que é uma

---

<sup>1</sup> Doutor em Geografia – Bipark Educação e Universidade Estadual do Oeste do Paraná [brunoborchertesilva@gmail.com](mailto:brunoborchertesilva@gmail.com);

<sup>2</sup> Doutorando pelo Curso de Geografia da Universidade Estadual do Oeste do Paraná, [bennert.geo@gmail.com](mailto:bennert.geo@gmail.com);

<sup>3</sup> Doutoranda pelo Curso de Geografia da Universidade Estadual do Oeste do Paraná, [carolgbeninca@gmail.com](mailto:carolgbeninca@gmail.com);

<sup>4</sup> Mestrando do Curso de Geografia da Universidade Estadual do Oeste do Paraná, [jefer\\_santos@outlook.com](mailto:jefer_santos@outlook.com);

<sup>5</sup> Professor pelo Curso de Geografia da Universidade estadual do Oeste do Paraná, [ericson.geo@gmail.com](mailto:ericson.geo@gmail.com);

<sup>6</sup> Professora pelo Curso de Geografia da Universidade estadual do Oeste do Paraná, [marciareg\\_calegari@hotmail.com](mailto:marciareg_calegari@hotmail.com).

ferramenta singular para o mapeamento e monitoramento de tipos de cobertura e uso da terra, também pode assumir um papel promissor no mapeamento pedológico em escala 1:50.000 (POPPIEL et al., 2021; SOTHE et al., 2022). Por se tratar de um ambiente de computação em nuvem, o GEE facilita o acesso a recursos de computação com alto desempenho temporal no processamento de grande conjunto de dados geoespaciais, com custos e equipamentos mínimos (GORELICK et al., 2017; PADARIAN; MINASNY; MCBRATNEY, 2015; VELASTEGUI-MONTOYA et al., 2023; WANG et al., 2022).

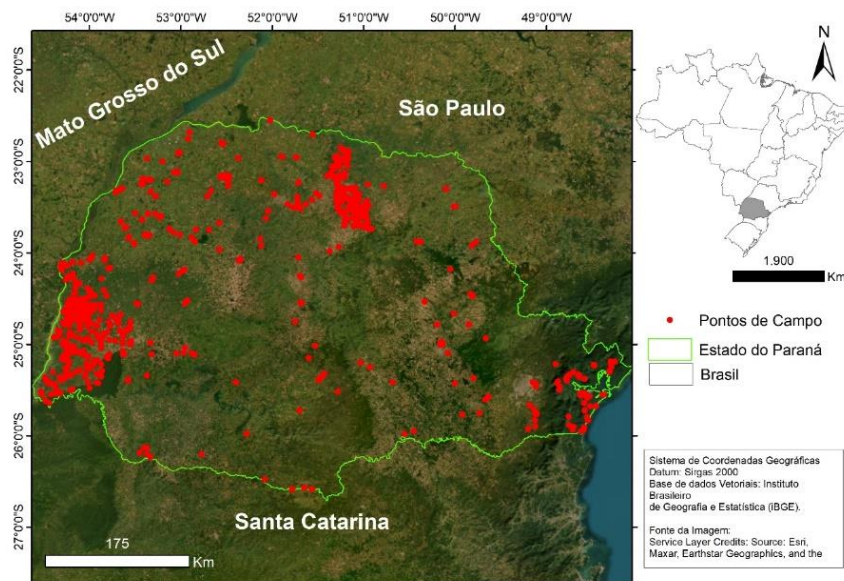
Diante dos benefícios que tais ferramentas geoespaciais podem oferecer às pesquisas sobre a distribuição dos solos na paisagem (MENDES; DEMATTÊ, 2022; SAFANELLI et al., 2021), p. ex., ganho de tempo e redução de custos de campanhas de campo, e, considerando a ausência de informações detalhadas de solos no Brasil, o objetivo deste estudo foi mapear os solos do estado do Paraná, em escala aproximada de 1:50.000, utilizando o algoritmo de *machine learning Random Forest* na plataforma do GEE.

## **MATERIAIS E MÉTODO**

### **Área do Estudo**

O estado do Paraná (199,3 mil Km<sup>2</sup>) situa-se na região sul do Brasil, cujos limites geográficos são: Estado de Santa Catarina (Sul), São Paulo (Norte), oceano Atlântico (Leste), Mato Grosso do Sul e Paraguai (Oeste) (Figura 1). As formações geológicas variam desde rochas magmáticas e metamórficas com 2,8 bi anos, rochas sedimentares paleozóicas, vulcânicas mesozóicas até sedimentos cretáceos (MINEROPAR, 2013). Esse arcabouço sustenta cinco importantes unidades morfoestruturais de paisagem: Primeiro Planalto, Segundo Planalto, Terceiro Planalto, Serra do Mar e Litoral (MAACK, 2001).

Os climas subtropical e tropical predominam no estado (NITSCHKE et al., 2019). Os principais rios são: rios Paraná, Parapanema, Ivaí, Piquiri, Tigabi e Iguaçu. Os solos mais comuns são os Latossolos, Nitossolos, Argissolos e Gleissolos (BHERING et al., 2007; EMBRAPA, 1984). As coberturas florestais das Florestas de Mata Atlântica e de Araucárias (IBGE, 2012), foram substituídas pelo uso agropecuário.



**Figura 1** - Mapa de localização do Estado do Paraná com os pontos de dados de campo.

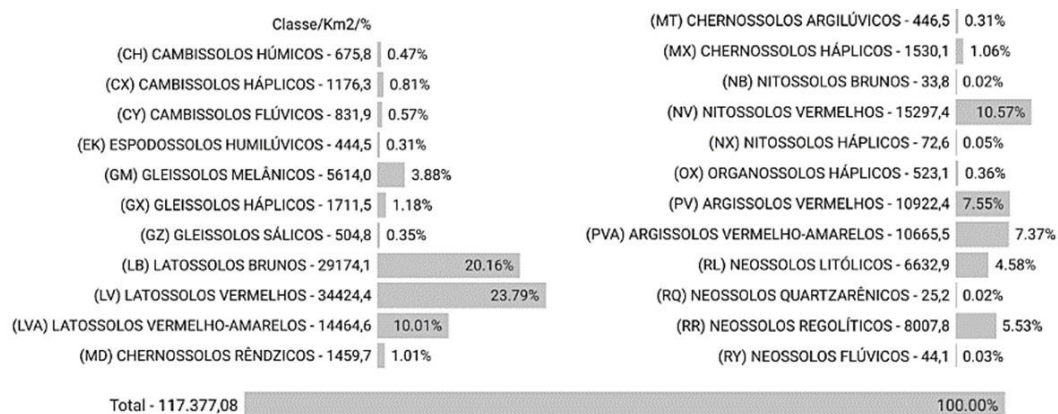
## Método

A primeira etapa do desenvolvimento da pesquisa foi a aquisição dos dados. A imagem utilizada foi a mediana dos pixels das cenas do sensor MSI (*MultiSpectral Instrument*) do satélite *Sentinel (COPERNICUS/S2\_SR)*, com correção de reflectância de superfície, entre as datas de 2022-01-03' a '2023-08-03' e percentual de cobertura de nuvens menor que 5%. No processamento dos dados, foram determinadas as variáveis do procedimento de classificação sendo: Índice de Água de Diferença Normalizada (NDWI), temperatura, geologia, Índice Topográfico de Umidade (TWI) e Posição Topográfica (IPT), Índices de Multiresolução de Topo Plano e Fundo de Vale Plano (MRRTF e MRVBF). Para realizar o treinamento do algoritmo, foram obtidas 1.278 amostras disponibilizadas no banco de dados aberto SoilData e de estudos realizados na região oeste do Paraná (SILVA; CALEGARI, 2019; SILVA; HAYAKAWA; MARTINS, 2022). Após, houve a classificação com o classificador *Random Forest (RF)* (BREIMAN, 2001, WANG et al. 2022; WHYTE et al. 2018; ZHOU et al. 2020), com 500 árvores de decisão, por meio do processo denominado de *bootstrap* (BREIMAN, 2001). Por fim, foi realizada a validação cruzada com 20% das amostras separadas utilizadas no treinamento e verificado quais variáveis foram mais significativas no procedimento de classificação.

## RESULTADOS E DISCUSSÃO

### Classes de solos identificadas

O Brasil conta com 13 ordens de solos. Por meio da metodologia deste estudo, foram identificadas e mapeadas 09 ordens de solos no estado do Paraná (Figura 2): ARGISSOLOS (13,36%), CAMBISSOLOS (1,86%), CHERNOSSOLOS (2,38%), ESPODOSSOLOS (0,31%), GLEISSOLOS (5,41%), LATOSSOLOS (53,95%), NITOSSOLOS (10,65%), ORGANOSSOLOS (0,36%) e NEOSSOLOS (10,17%). O estudo mapeou 23 subordens (Figura 3): os LATOSSOLOS VERMELHOS, LATOSSOLOS BRUNOS, e NITOSSOLOS VERMELHOS somaram 54,5% dos solos mapeados, estão no Terceiro Planalto e são originados a partir da alteração do basalto (BHERING et al., 2007; EMBRAPA, 1984). As subordens dos LATOSSOLOS VERMELHO-AMARELOS, ARGISSOLOS VERMELHOS, ARGISSOLOS VERMELHO-AMARELOS, que estão associados a material de origem sedimentar, representaram 24,92% da área estadual. Por sua vez, as subordens com menor representatividade espacial foram os NEOSSOLOS QUARTZARÊNICOS, NITOSSOLOS BRUNOS, NEOSSOLO FLÚVICO e NITOSSOLO HÁPLICO, representando menos de 0,2% da área mapeada e tem uma distribuição difusa ao longo da área mapeada, corroborando mapeamentos prévios de Bhering et al. (2007) e Embrapa (1984).



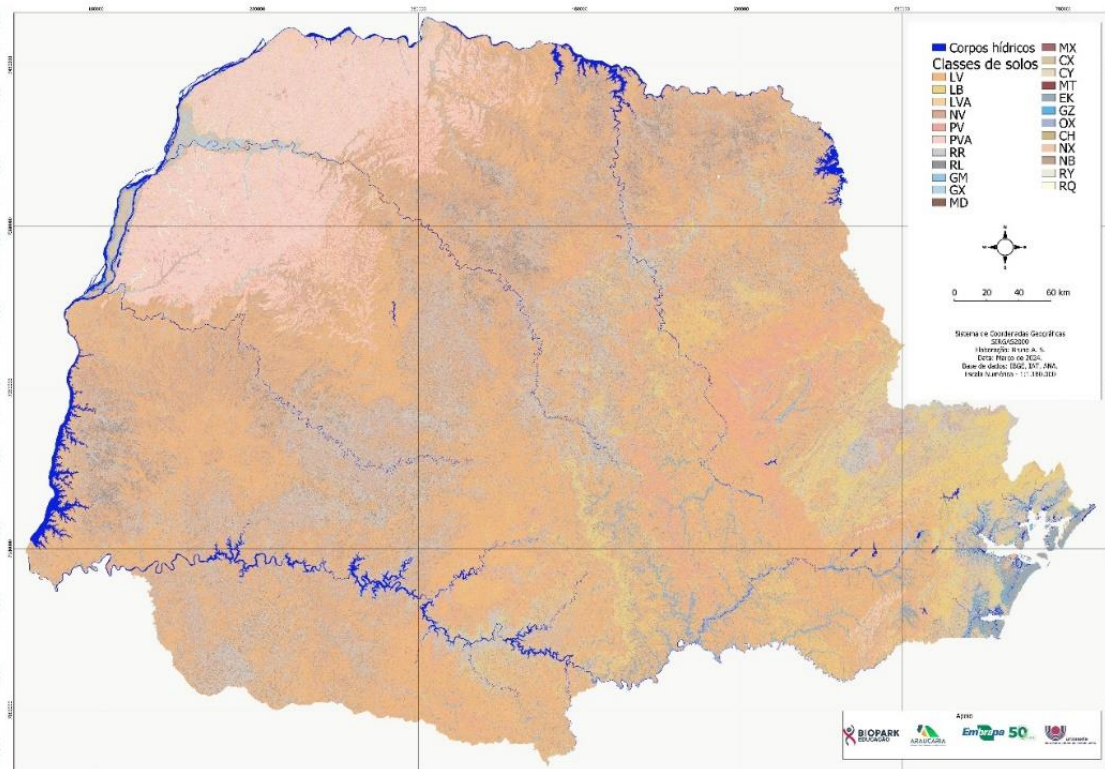
**Figura 2** – Percentual de área ocupada pelas subordens de solo mapeadas.

### Acurácia

A análise dos resultados da validação demonstrou exatidão global de 0,70 e índice Kappa de 0,62 (Figura 4a), indicando bom desempenho do RF na espacialização dos solos (SIQUEIRA et al., 2023). Em relação a acurácia do usuário (AU) as classes LB, LVA, GX, EK obtiveram valor de 1,00, seguido de LV (0,84), NV (0,73), PVA (0,67), PV

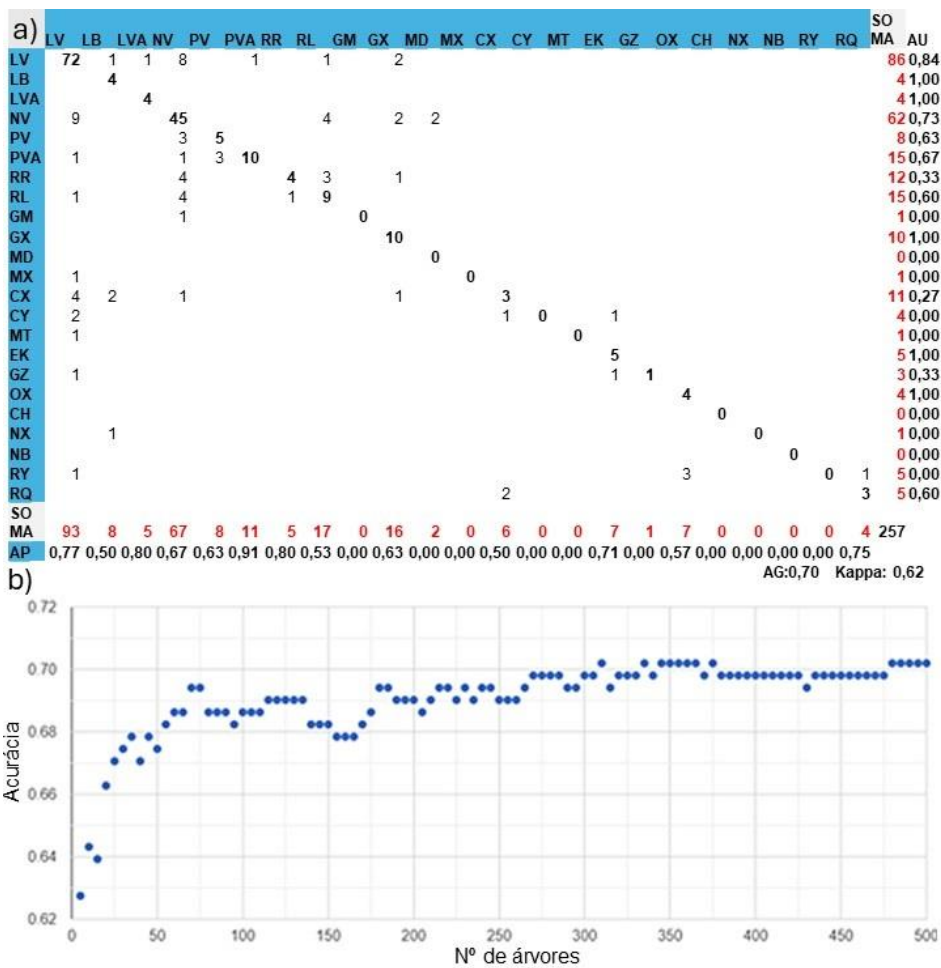


(0,63), RQ (0,60), RR (0,33), GZ (0,33) e CX (0,27). Por outro lado, as classes PVA (0,91), LVA (0,80), LV (0,77), RQ (0,75), EK (0,71), NV (0,67), PV (0,63), OX (0,57), RL (0,53), LB (0,50), CX (0,50) obtiveram os maiores valores de acurácia do produtor (AP). Quanto aos baixos valores de AU e AP, estes refletem os locais com pouca cobertura de dados, conforme constatado por Žížala et al. (2022).



**Figura 3 -** Subordens de solos mapeadas no estado do Paraná.

A acurácia foi afetada principalmente devido ao número de árvores de decisão utilizadas no *Random Forest* (500). Conforme ilustrado na Figura 4b, de 0 a 250 árvores de decisão a acurácia apresentou um gradativo aumento de 0,62 até 0,70, a partir deste ponto mantém-se um padrão que varia de 0,69 a 0,70 até 500 árvores de decisão. Este estudo obteve acurácia final de 0,70, e isso é em função do número de árvores de decisão adotadas que melhorou os resultados de classificação (BESSINGER et al., 2022). Por fim, verificou-se que o limite de árvores de decisão para o mapeamento das classes de solo é de 250, acima disso não houve melhora da precisão da classificação.



**Figura 4.** Matrix de erro (a) e variação da acurácia em relação às árvores de decisão do classificador *Random Forest* (b).

## CONSIDERAÇÕES FINAIS

Nosso estudo obteve informações inéditas sobre a distribuição espacial dos solos em grande escala para todo o estado do Paraná, identificando 23 subordens de solo. As subordens dos LATOSSOLOS VERMELHOS e NEOSSOLOS QUARTZARÊNICOS, respectivamente, tiveram maior e menor representatividade espacialmente. Os resultados da validação cruzada demonstraram-se satisfatórios em consideração ao número de pontos amostrados em campo, com acurácia de 0,70. No entanto, a eficiência deste tipo de metodologia esbarra na disponibilidade de amostras para treinamento do algoritmo, considerando que resultados mais expressivos de acurácia ocorreriam com o aumento do número de pontos em locais não amostrados, principalmente nas regiões central-sul e leste do estado.

**Palavras-chave:** *Random forest; soildata*, índices espectrais, índices geomorfométricos, pedometria.

**AGRADECIMENTOS:** A Fundação Araucária do Estado do Paraná - Termo de Convênio n. 215/2023, ao Biopark Educação. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001”. Ao CNPq (Processo 313757/2021-6).

## REFERÊNCIAS

ARROUAYS, D.; DAWSON, L. **Soil priorities around the world - an introduction. Geoderma Regional.** Elsevier B.V., 1 set. 2022.

BHERING, S. B. et al. **Mapa de Solos do Estado do Paraná.** Rio de Janeiro: Embrapa Solos, 2007.

BREIMAN, L. Random Forests. **Machine Learning**, v. 45, p. 5–32, 2001.

BESSINGER, M.; VOGEL, M.L.; SKOWNO, A.; CONRAD, F. Landsat-8 based coastal ecosystem mapping in South Africa using random forest classification in Google Earth Engine. *South African Journal of Botany*. 150, 2022.

DEMATTÊ, J. A. M. et al. The Brazilian soil priorities. **Geoderma Regional**, v. 29, 1 jun. 2022.

EMBRAPA. **Levantamento de Reconhecimento dos Solos do Estado do Paraná: Tomo I e II.** Curitiba: EMBRAPA - SNLCS/SUDESUL/IAPAR, 1984.

GAGKAS, Z.; LILLY, A.; BAGGALEY, N. J. Digital soil maps can perform as well as large-scale conventional soil maps for the prediction of catchment baseflows. **Geoderma**, v. 400, n. May, 2021.

GORELICK et al. Google Earth Engine: Planetary – scale geospatial analysis for everyone. **Remote Sensing of Environment**, 202, p.18-27, 2017. Doi.org/10.1016/j.rse.2017.06.031.

HATEFFARD, F.; STEINBUCH, L.; HEUVELINK, G. B. M. Evaluating the extrapolation potential of random forest digital soil mapping. **Geoderma**, v. 441, p. 116740, jan. 2024.

IBGE. Manual Técnico da Vegetação Brasileira: Sistema fitogeográfico; Inventário das formações florestais e campestres; Técnicas e manejo de coleções botânicas; Procedimentos para mapeamentos. **Instituto Brasileiro de Geografia e Estatística - IBGE**, p. 271, 2012.

IBGE. **Instituto Brasileiro de Geografia e Estatística.** Acesso em: 19 nov. 2023.

MAACK, R. Breves Notícias Sobre a Geologia dos Estados do Paraná e Santa Catarina. **Brazilian Archives of Biology and Technology**, v. Jubilee Vo, p. 169–288, 2001.

MENDES, W. DE S.; DEMATTÊ, J. A. M. Digital soil mapping outputs on soil classification and sugarcane production in Brazil. **Journal of South American Earth Sciences**, v. 116, 1 jun. 2022.

MINEROPAR. **O Grupo Serra Geral no estado do Paraná: mapeamento geológico na escala de 1:250.000 (I e II).** 1ª ed. Curitiba: Mineropar, 2013. v. 1 e 2

NGU, N. H. et al. Active learning-based random forest algorithm used for soil texture classification mapping in Central Vietnam. **Catena**, v. 234, 1 jan. 2024.

NITSCHKE, P. R. et al. **Atlas Climático do Estado do Paraná**. Londrina: IAPAR, 2019.

PADARIAN, J.; MINASNY, B.; MCBRATNEY, A.B. Using Google's cloud-based platform for digital soil mapping. **Computers & Geosciences**, 83, p.80-88, 2015.

POPPIEL, R. R. et al. High resolution middle eastern soil attributes mapping via open data and cloud computing. **Geoderma**, v. 385, n. August 2020, p. 1–15, 2021.

ROSIN, N. A. et al. Mapping Brazilian soil mineralogy using proximal and remote sensing data. **Geoderma**, v. 432, 1 abr. 2023.

SAFANELLI, J. L. et al. Leveraging the application of Earth observation data for mapping cropland soils in Brazil. **Geoderma**, v. 396, n. February, p. 1–13, 2021.

SILVA, B. A.; CALEGARI, M. R. Atualização do Mapa de Solos na Bacia do Paraná 3: Estudo de Caso das Sub-bacias Hidrográficas dos Rios São Francisco Verdadeiro e São Francisco Falso. **Perspectiva Geográfica**, v. 14, n. 20, p. 48–58, 2019.

SILVA, B. A.; HAYAKAWA, E. H.; MARTINS, V. M. Espacialização das classes solos a partir da utilização de atributos geomorfométricos na Bacia do Paraná 3, Brasil. **Revista Brasileira de Geografia Física**, v. 14, n. 7, 2022.

SIQUEIRA, R. G. et al. Machine learning applied for Antarctic soil mapping: Spatial prediction of soil texture for Maritime Antarctica and Northern Antarctic Peninsula. **Geoderma**, v. 432, 1 abr. 2023.

SOIL DATA. <https://soildata.mapbiomas.org/>. Acessado em Novembro de 2023.

SOTHE, C. et al. Large scale mapping of soil organic carbon concentration with 3D machine learning and satellite observations. **Geoderma**, v. 405, p. 1–14, 2022.

VELASTEGUI-MONTOYA, A.; BURBANO, N.M.; MERO, P.C.; TORRES, H.R.; SADECK, L.; ADAMI, M. Google Earth Engine: A Global Analysis and Future Trends. **Remote Sens.** 2023, 15(14), 3675; <https://doi.org/10.3390/rs15143675>

WANG, S. et al. Using soil library hyperspectral reflectance and machine learning to predict soil organic carbon: Assessing potential of airborne and spaceborne optical soil sensing. **Remote Sensing of Environment**: 271. 2022. doi: 10.1016/j.rse.2022.112914.

WHYTE, A.; FERENTINOS, K. P.; PETROPOULOS, G. P. A new synergistic approach for monitoring wetlands using Sentinels -1 and 2 data with object-based machine learning algorithms. **Environmental Modelling and Software** 104: 2018. 40-54. doi: 10.1016/j.envsoft.2018.01.023.

ŽÍŽALA, D. et al. High-resolution agriculture soil property maps from digital soil mapping methods, Czech Republic. **Catena**, v. 212, 1 maio 2022.

ZHOU, B.; OKIN, G.S.; ZHANG, J. Leveraging Google Earth Engine (GEE) and machine learning algorithms to incorporate in situ measurement from different times for rangelands monitoring. **Remote Sensing of Environment**, 236. 2020. doi:10.1016/j.rse.2019.111521.