

MapBiomass Solos: Revelando a dinâmica espaço-temporal dos estoques de carbono no solo e suas relações com as mudanças no clima e na cobertura e uso da terra no Brasil

Bárbara Costa da Silva¹, Wallace Vieira da Silva², Erli Pinto dos Santos³,
Bruna Zambrano dos Santos⁴, Luciana da Luz Silva⁵, Aline Mari Huf dos Reis⁶,
Alessandro Samuel-Rosa⁷, Taciara Zborowski Horst⁷

¹ Universidade de Brasília, Brasília, DF - barbaracosta2@yahoo.com.br; ² Instituto de Pesquisa Ambiental da Amazônia, Brasília, DF - wallace.silva@ipam.org.br; ³ Universidade Federal de Viçosa, Viçosa, MG - erlipinto@gmail.com; ⁴ Universidade Federal do Paraná, Curitiba, PR - agrimensurazambrano@gmail.com; ⁵ Universidade Estadual de Feira de Santana, Feira de Santana, BA - lucisilva.luz@gmail.com; ⁶ Universidade Federal de Goiás, Goiânia, GO - huf.aline@gmail.com; ⁷ Universidade Tecnológica Federal do Paraná, Santa Helena/Dois Vizinhos, PR - alessandrorosa@utfpr.edu.br; taciara@utfpr.edu.br

RESUMO

Apoiado no conjunto de dados de solo disponíveis no Repositório de Dados do Solo Brasileiro (FEBR) e no âmbito da Rede MapBiomass, o Grupo de Trabalho em Solos (GT Solos) elaborou a primeira série temporal dos estoques de carbono orgânico nos solos brasileiros. Os mapas anuais produzidos representam os estoques de carbono orgânico do solo (COS) nos primeiros 30 cm do solo, no período de 1985-2021. Esta é uma iniciativa inédita que visa a compreensão dos efeitos da mudança da cobertura e uso da terra e do clima nas dinâmicas espaço-temporais dos estoques de COS em todo o Brasil.

Palavras-chave — Modelagem espaço-temporal, séries temporais, sensoriamento remoto, aprendizado de máquina.

ABSTRACT

Based on the set of soil data available at the Data Repository of the Brazilian Soil (FEBR) and within the scope of the MapBiomass Network, the Soil Working Group (GT Solos) produced the first time series of organic carbon stocks in Brazilian soils. The annual maps produced represent soil organic carbon stocks (SOC) in the first 30 cm of the soil, in the period of 1985-2021. This is an unprecedented initiative that aims to understand the effects of land cover and use and climate change on the space-time dynamics of SOC stocks in Brazil.

Keywords — Spatiotemporal modeling, time series, remote sensing, machine learning.

1. INTRODUÇÃO

O Brasil é um dos países cuja economia depende consideravelmente do uso da terra. Ainda assim, o país não possui informações detalhadas sobre o solo e a variação de suas propriedades no espaço e no tempo. Os esforços para mapear os estoques de carbono orgânico do solo (COS) têm

ignorado a dinâmica espacial e temporal. Iniciativas de mapeamento de propriedades dos solos no Brasil, produziram mapas sem referência temporal dos estoques de carbono orgânico do solo, atributo este que é sensível às mudanças na cobertura e uso da terra e no clima.

O Grupo de Trabalho em Solos (GT Solos), criado no âmbito da Rede MapBiomass em 2021, visa revelar a dinâmica espaço-temporal das propriedades do solo e suas relações com as mudanças no uso da terra e no clima no Brasil. Para isso, o GT Solos tem trabalhado na primeira série histórica brasileira de mapas de estoque de carbono orgânico (kg m^{-2}) nos primeiros 30 cm do solo. Utilizando dados de campo, satelitários e multi-fontes, a série está sendo produzida usando *machine learning* no *Google Earth Engine* (GEE). Ela compreende os anos de 1985 a 2021, com resolução temporal de um ano e espacial de 30 m.

Essa iniciativa só está sendo possível graças ao compartilhamento de dados históricos de solos e seu provisionamento pelo Repositório de Dados do Solo Brasileiro (FEBR, www.pedometria.org/febr/). O objetivo do GT Solos é produzir atualizações anuais da série histórica dos mapas de COS, acompanhados de medidas de incerteza local e global. Essas informações permitirão revelar os efeitos das mudanças climáticas e de uso e cobertura da terra no estoque de COS. A estrutura necessária para produzir essa série de mapas e suas atualizações precisa ser sólida, dada a escassez de dados de campo e sua distribuição espaço-temporal heterogênea.

Neste resumo, apresentamos o histórico do trabalho de mapeamento dos estoques de COS do Brasil realizado pelo GT Solos, destacando os desafios e perspectivas, tendo em vista a continuidade desta iniciativa em um cenário de dispersão e pequena disponibilidade de dados de campo.

2. MATERIAL E MÉTODOS

A série histórica foi obtida por algoritmos de *machine learning*, que computam a correlação entre o estoque de COS medido em campo e covariáveis ambientais.

2.1. Dados de campo e covariáveis ambientais

Os dados de campo usados para treinar o modelo preditivo são oriundos do FEBR [3]. O FEBR tem a maior coleção de dados abertos para cálculo dos estoques de COS, com 247 conjuntos de dados e mais de 20 mil amostras coletadas desde a década de 1960. Esses dados são padronizados e harmonizados usando funções de pedotransferência. No trabalho de mapeamento espaço-temporal realizado pelo GT Solos, são consideradas apenas as amostras com coordenadas temporal (data de coleta) e geográficas.

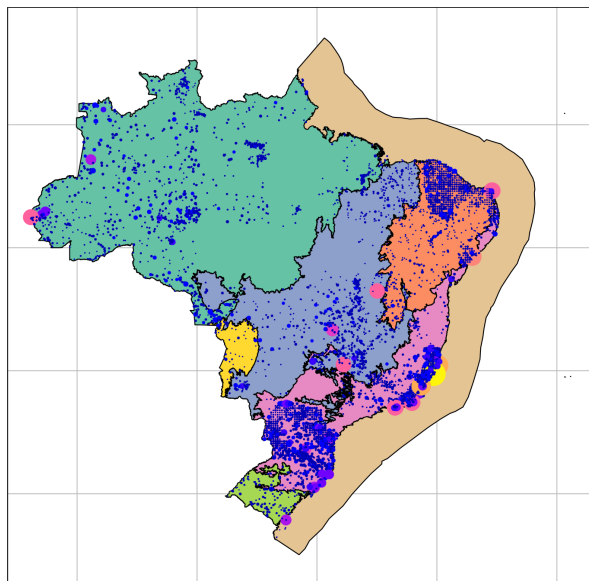


Figura 1. Distribuição espacial das amostras de campo oriundas do FEBR

As covariáveis utilizadas na predição representam fatores e processos de formação do solo (clima, organismos e relevo) e foram obtidas em bancos de dados espaciais livres e abertos. Cada covariável foi selecionada para a predição com base na plausibilidade de sua relação com a dinâmica espaço-temporal dos estoques de COS. Sua inserção no modelo preditivo se deu de modo gradual, com o intuito de avaliar sua influência nos mapas resultantes.

As covariáveis foram separadas em dinâmicas (que mudam no tempo) e estáticas. Dentre as dinâmicas, foram considerados dados climáticos [4] e bioclimáticos [5], índice de vegetação por diferença normalizada (NDVI) e uso e cobertura da terra [6]. Nas estáticas, foram incluídos dados geomorfológicos [7], [8], classificação climática [9], biomas [10] e fitofisionomias [11]. Também foram usados mapas de propriedades dos solos produzidos por iniciativas globais, como teor de carbono, areia, silte e argila, fragmentos grossos e probabilidade de ocorrência de classes de solo [12]. As covariáveis multicategóricas foram convertidas em covariáveis binárias para facilitar sua interpretação. Além disso, foram computadas a idade do uso e cobertura da terra e identificadas as áreas naturais, antrópicas e estáveis.

2.2. Modelo preditivo

O algoritmo *Random Forest* foi utilizado como modelo preditivo. O algoritmo, que constrói árvores de regressão, foi adotado para capturar correlações não lineares entre o estoque de COS e as covariáveis. Como hiperparâmetros foram definidos o número de árvores de regressão $n_{tree} = 1/10$ do número de amostras e o número de covariáveis selecionadas em cada divisão $m_{try} = 1/3$ do número de covariáveis. O modelo foi treinado no GEE com a função `ee.Classifier.smileRandomForest` com o modo de saída "Regression". O modelo treinado foi utilizado para prever os estoques de COS para cada ano da série de 1985 a 2021 com resolução espacial de 30 metros. Essas definições estão de acordo com as características dos produtos do MapBiomas e especificações do IPCC e FAO [1], [2].

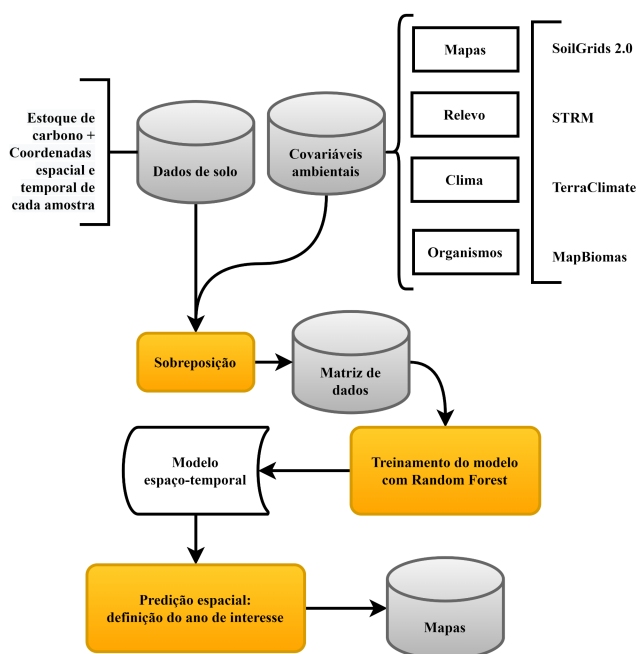


Figura 2. Fluxograma das etapas metodológicas para a modelagem espaço-temporal dos estoques de COS

3. RESULTADOS E DISCUSSÃO

A modelagem espaço-temporal dos estoques de COS incluiu uma série de exercícios com avaliação de covariáveis e agregação de dados de campo. Os resultados relevantes obtidos em cada um deles são apresentados a seguir. O estoque de COS total é apresentado para o ano de 2000, considerado o ano para o qual os produtos sem referência temporal melhor representam [12].

3.1. Covariáveis de solo e relevo

O primeiro exercício realizado foi a elaboração de um mapa sem referência temporal dos estoques de COS. O objetivo foi avaliar o padrão da distribuição espacial das predições

em comparação com os produtos disponíveis para o país até então. Para isso, o estoque de COS foi predito apenas em função de covariáveis estáticas (propriedades do solo e relevo). Como referência, foram selecionados os dados climáticos e de uso da terra do ano 2000. Nesta versão foram utilizadas 6.036 amostras de campo e 68 covariáveis. Neste mapa foi calculado um estoque de COS total de 32,99 Pg (petagrama) para os primeiros 30 cm do solo em todo o território brasileiro. O padrão espacial foi similar àquele produzido recentemente pelo PronaSolos [13].

3.2. Efeito temporal de covariáveis bioclimáticas

O próximo exercício incluiu dados da classificação climática, variáveis bioclimáticas e probabilidade de ocorrência de classes de solo. O modelo foi treinado considerando o cenário espaço-temporal, onde o ano de coleta das amostras de campo foi incluído como uma das covariáveis predictoras. No treinamento foram utilizadas 6.036 amostras de campo e 116 covariáveis. O estoque total de COS calculado para o ano de 2000 foi de 32,98 Pg. Nos resultados, se observou artefatos como formas geométricas e pixels nulos, oriundos das covariáveis bioclimáticas. Além disso, as previsões apresentaram elevada variância no tempo, com tendência de aumento a cada cinco anos.

Como alternativa às limitações verificadas, as covariáveis bioclimáticas foram avaliadas individualmente, a fim de selecionar as mais importantes para explicar a variação dos estoques de COS. No exercício seguinte, foram selecionadas somente as covariáveis bioclimáticas temperatura média anual, isotermia, variação da temperatura anual e precipitação anual. Foram utilizadas 6.002 amostras de campo e 81 covariáveis. A predição resultou no total de 32,65 Pg de COS para o ano de 2000. Ainda assim, a variância temporal dos dados continuou alta, mesmo em áreas de cobertura e uso da terra estável durante toda a série.

Sabe-se que os estoques de COS não dependem apenas dos dados de cobertura da terra e climáticos do ano analisado, mas sofrem com a influência de anos anteriores. Por isso, o decaimento exponencial destes dados foi considerado como uma estratégia para melhor representar esses efeitos na estimativa dos estoques de COS [14]. Foram adotadas ponderações nos dados de NDVI e variáveis bioclimáticas previamente selecionadas. Neste exercício foram utilizados 5.978 dados de campo e 82 covariáveis. A predição resultou no total de 31,87 Pg de COS para o ano 2000. Como limitação foi identificado que alguns dados de treinamento não foram computados, pois os dados de propriedades dos solos apresentavam pixels vazios.

3.3. Aumento do conjunto de amostras de campo

O próximo exercício incluiu dados de cerca de 400 amostras do Inventário Florestal Nacional (IFN) [15]. Essa inclusão resultou na superestimativa dos estoques de COS em todo o território nacional a partir do ano de 2013, especialmente

em áreas de floresta natural. A comparação dos mapas de 2012 e 2014 apontou aumento do estoque de COS na ordem do quilograma, não sendo plausível do ponto de vista pedológico [1]. Destaca-se que as amostras do IFN não apresentam dados de fragmentos grossos, o que pode ter causado a superestimativa dos estoques de COS. Além disso, os dados que estavam sendo utilizados contemplavam sobretudo amostras em áreas de lavouras, com menores estoques de COS. A adição do número de amostras em áreas de floresta, desde 2013, forçou o aumento da correlação positiva entre o ano de amostragem e o estoque de COS.

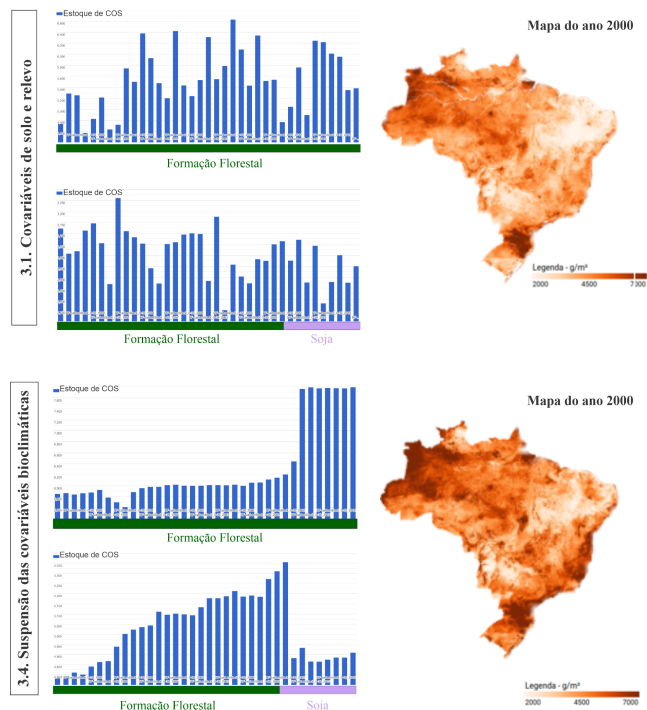


Figura 3. Os gráficos apresentam a variação temporal dos estoques de COS observada em um pixel de área estável e um pixel com conversão para soja.

Para tentar reduzir o efeito causado pela adição de dados do IFN, foi removida do modelo a covariável “ano” do modelo. A partir deste exercício, o modelo passou a ser treinado para a predição espacial e extrapolado para o tempo (modelagem do espaço para o tempo). As covariáveis dinâmicas foram amostradas com base no ano de amostragem de campo, e utilizadas para predição espacial, conforme o ano de interesse. Neste exercício foram utilizados dados de 6.461 amostras de campo e 82 covariáveis. As estimativas do modelo computaram um estoque de COS total de 37,27 Pg para o país em 2000. Ainda assim, não foi observado melhora na estimativa temporal dos estoques de COS.

Considerando que a superestimativa dos estoques estava concentrada em florestas, foram incluídos no modelo mais duas formas de representação da vegetação. Primeiro foram incluídos os dados binários dos biomas brasileiros e

fitofisionomias [10], [11]. O objetivo deste exercício foi treinar o modelo para considerar o tipo de formação natural e superar as limitações identificadas com a adição dos dados do IFN. Foram utilizadas 6.867 amostras e 84 covariáveis. O estoque de COS predito para o ano 2000 foi de 37,95 Pg. Depois disso, foram incluídos os dados de áreas naturais e antrópicas e dados de áreas estáveis e não estáveis. O objetivo foi melhor representar o efeito dos sistemas de uso da terra nos estoques de COS no modelo. Neste exercício foram utilizados 6.965 dados de campo e 87 covariáveis. O estoque de COS calculado para o ano 2000 foi de 38,20 Pg. Os desafios com a superestimativa do estoque de COS a partir de 2013 persistiram nesses exercícios.

3.4. Suspensão das covariáveis bioclimáticas

Para melhor avaliar a dinamicidade das mudanças de uso e cobertura da terra e do clima nos estoques de COS, o grupo decidiu avaliar individualmente esses dois vetores de mudança. Neste exercício, foram removidas as covariáveis climáticas. A variação temporal foi realizada utilizando as covariáveis dinâmicas de uso e cobertura da terra e seus subprodutos (áreas naturais e antrópicas e áreas estáveis e não estáveis). Utilizando 7.094 dados de campo e 83 covariáveis, o estoque de COS para o ano 2000 foi de 38,50 Pg. O padrão temporal das estimativas melhorou consideravelmente em relação às versões anteriores, com leve aumento nos estoques de COS nos últimos anos da série. A ausência de covariáveis climáticas reduziu a variação sazonal dos estoques e possibilitou a análise dos efeitos da mudança no uso e cobertura da terra nos estoques de COS ao longo da série.

4. CONCLUSÕES

A principal meta do GT Solos é estabelecer o arcabouço para atualização anual da série de mapas de estoque de COS. O grupo está empenhado em analisar os padrões temporais para avaliar incrementalmente os dados inseridos no modelo preditivo. Os resultados mostram dificuldades em avaliar os efeitos de médio e longo prazo do clima e uso e cobertura da terra. Além disso, há efeitos negativos do desbalanço amostral no espaço e no tempo. Por isso, está sendo considerada a adoção de modelos separados para áreas estáveis e não estáveis para isolar os fatores dinâmicos e compensar o excesso de amostras em áreas florestais. Já para avaliar os efeitos das covariáveis dinâmicas climáticas, são necessárias mais amostras de campo. Covariáveis importantes para a dinâmica do estoque de COS, como as práticas de manejo, serão produzidas.

5. REFERÊNCIAS

[1]“2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories — IPCC”. [Online]. Disponível em: <https://www.ipcc.ch/report/2019-refinement-to-the-2006-ipcc-guidelines-for-national-greenhouse-gas-inventories/>.

[2] FAO and ITPS, Global Soil Organic Carbon Map (GSOCmap) Version 1.5: Technical report. Rome, Italy: FAO, 2020 [Online]. Disponível em: <https://www.fao.org/documents/card/en/c/ca7597en>.

[3]A. Samuel-Rosa, R. S. D. Dalmolin, J. M. Moura-Bueno, W. G. Teixeira, e J. M. F. Alba, “Open legacy soil survey data in Brazil: geospatial data quality and how to improve it”, *Sci. Agric.*, vol. 77, no 1, p. e20170430, 2020, doi: 10.1590/1678-992x-2017-0430.

[4]“TerraClimate”, Climatology Lab. [Online]. Disponível em: <https://www.climatologylab.org/terraclimate.html>.

[5]“WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas - Fick - 2017 - International Journal of Climatology - Wiley Online Library”. [Online]. Disponível em: <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.5086>.

[6]“Mapbiomas Brasil | Coleção 7”. [Online]. Disponível em: <https://brasil.mapbiomas.org/>

[7]G. Amatulli, D. McInerney, T. Sethi, P. Strobl, e S. Domisch, “Geomorpho90m, empirical evaluation and accuracy assessment of global high-resolution geomorphometric layers”, *Sci. Data*, vol. 7, no 1, p. 162, maio 2020, doi: 10.1038/s41597-020-0479-6.

[8]OpenTopography, “NASADEM Global 1 arc-second Digital Elevation Model”, 2021, doi: 10.5069/G93T9FD9. [Online]. Disponível em: <https://opentopography.org/meta/OT.032021.4326.2>.

[9]C. A. Alvares, J. L. Stape, P. C. Sentelhas, G. de Moraes, J. Leonardo, e G. Sparovek, “Köppen 's climate classification map for Brazil”, *Meteorol. Z.*, vol. 22, no 6, p. 711–728, 2013.

[10]“IBGE Biomas”. [Online]. Disponível em: <https://www.ibge.gov.br/geociencias/cartas-e-mapas/informacoes-ambientais/15842-biomas.html?=&t=acesso-ao-produto>.

[11]“IBGE Vegetação”. [Online]. Disponível em: <https://www.ibge.gov.br/geociencias/informacoes-ambientais/vegetacao.html>.

[12]L. Poggio et al., “SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty”, *SOIL*, vol. 7, no 1, p. 217–240, jun. 2021, doi: 10.5194/soil-7-217-2021.

[13]“Mapa de estoque de carbono orgânico do solo (COS) a 0-30 cm do Brasil. - Portal Embrapa”. [Online]. Disponível em: <https://www.embrapa.br/busca-de-publicacoes/-/publicacao/1085197/mapa-de-estoque-de-carbono-organico-do-solo-cos-a-0-30-cm-do-brasil>.

[14] G. B. M. Heuvelink et al., “Machine learning in space and time for modeling soil organic carbon change”, *Eur. J. Soil Sci.*, vol. 72, no 4, p. 1607–1623, 2021, doi: 10.1111/ejss.12998.

[15] “Inventário Florestal Nacional”, Ministério da Agricultura, Pecuária e Abastecimento. [Online]. Disponível em: <https://www.gov.br/agricultura/pt-br/assuntos/servico-florestal-brasil/ileiro/ifn-inventario-florestal-nacional/>