CÁSSIO MARQUES MOQUEDACE DOS SANTOS

MODELING AND MAPPING OF SOIL ORGANIC CARBON STOCK IN THE STATE OF RONDÔNIA

Dissertação apresentada a Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Solos e Nutrição de Plantas, para obtenção do título de *Magister Scientiae*.

Orientador: Renildes Lúcio Ferreira Fontes

Coorientadores: Irene Maria Cardoso
Lucas de Carvalho Gomes

VIÇOSA – MINAS GERAIS
2021

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T
S237m
2021

Santos, Cássio Marques Moquedace, 1991-
Modeling and mapping of soil organic carbon stock in the state of Rondônia / Cássio Marques Moquedace Santos. – Viçosa, MG, 2021.
1 dissertação eletrônica (119 f.): il. (algumas color.).

Texto em inglês.
Inclui apêndices.
Orientador: Renildes Lucio Ferreira Fontes.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Inclui bibliografia.
DOI: https://doi.org/10.47328/ufvbbt.2021.115
Modo de acesso: World Wide Web.

1. Solos - Trópicos. 2. Mudanças climáticas. 3. Amazônia. 4. Aprendizado do computador. 5. Húmus. 6. Mapeamento do solo. I. Universidade Federal de Viçosa. Departamento de Solos. Programa de Pós-Graduação em Solos e Nutrição de Plantas. II. Título.

CDD 22. ed. 631.498111

Bibliotecário(a) responsável: Alice Regina Pinto CRB6 2523

CÁSSIO MARQUES MOQUEDACE DOS SANTOS

MODELING AND MAPPING OF SOIL ORGANIC CARBON STOCK IN THE STATE OF RONDÔNIA

Dissertação apresentada a Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Solos e Nutrição de Plantas, para obtenção do título de *Magister Scientiae*.

APROVADA: 23 de julho de 2021.

Assentimento:

Cássio Marques Moquedace dos Santos
Autor

Renildes Lúcio Ferreira Fontes
Orientador

AGRADECIMENTOS

Talvez essa parte seja a mais difícil dessa dissertação, pois independente do fato da autoria ser individual, esta dissertação é fruto de muito esforço coletivo, inclusive indireto, e a vida não é submetida a análise de regressão, assim como não é o valor de p que mede a significância das pessoas nesta trajetória. Dito isso, agradeço...

A minha mãe Maria Suely dos Santos Marques, que sempre me motivou, acreditou quando não acreditei e me proporcionou a possibilidade do concluir o nível superior e o mestrado, mesmo não tendo tal oportunidade no passado. Desconheço pessoa mais sábia, sem você não teria sido possível, MUITO OBRIGADO.

A minha companheira Sherellyn Daphnee Alves Moretti, por estar todo tempo ao meu lado incondicionalmente. Nos momentos mais difíceis, que de longe foram raros, me fez acreditar que superaria esta difícil e gratificante etapa, obrigado.

A minha tia Solange Mendes Moquedace dos Santos, exemplo de resistência na educação brasileira, mesmo diante de tempos tão sombrios. Foi minha inspiração e apoio em diversos momentos nesta longa caminhada.

Ao meu orientador professor Renildes Lúcio Ferreira Fontes, pela liberdade e confiança na conclusão deste trabalho.

Aos meus coorientadores Irene Maria Cardoso e Lucas Gomes de Carvalho, exemplos profissionais e pessoais. Além das valiosas contribuições para este trabalho, me ensinaram a importância do conhecimento compartilhado e de praticar ciência com humildade, obrigado.

Ao professor Elpídio Inácio Fernandes Filho, por todos os conselhos e ideias, pela disponibilidade e apoio em todos os momentos desta jornada, que muitas vezes o fez renunciar a horários de descanso. Obrigado por compartilhar seu vasto conhecimento.

Ao professor e quem posso chamar de amigo Emanuel Fernando Maia de Souza, que sempre soube me conceder as palavras certas e me incentivou a ingressar na carreira acadêmica, obrigado.

Aos companheiros e irmãos na amizade que fizeram parte da minha formação e que vão continuar presentes em minha vida com certeza, em especial ao Arthur da Cruz Silva, Rafael Gomes Siqueira e Rugana Imbana.

A todos os funcionários e docentes da Universidade Federal de Viçosa e Pós-Graduação em Solos e Nutrição de Plantas, peças importantes na conclusão deste trabalho.

Pesquisa desenvolvida com o auxílio do Cluster da UFV (Universidade Federal de Viçosa).

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

RESUMO

SANTOS, Cássio Marques Moquedace, M.Sc., Universidade Federal de Viçosa, julho de 2021. **Modelagem e mapeamento do estoque de carbono orgânico do solo no estado de Rondônia**. Orientador: Renildes Lúcio Ferreira Fontes. Coorientadores: Irene Maria Cardoso e Lucas de Carvalho Gomes.

As mudanças no climáticas ocasionadas pelas atividades antrópicas tem sido uma das principais pautas ambientais em todo o mundo. O solo, pode atuar como moderador dessas alterações, pois armazena dois terços de todo carbono terrestre e pequenas mudanças neste estoque podem aumentar as concentrações de gases de efeito estufa (GEE) e agravar o aquecimento global. Portanto, objetivou-se modelar e mapear o estoque de carbono orgânico do solo no estado de Rondônia, no presente e em cenários futuros até 2100. Utilizou-se dados de quase três mil perfis de solo oriundos do estudo para segunda aproximação do Zoneamento Socioeconômico Ecológico de Rondônia. Estratificou-se o estoque de SOC nas profundidades de 0-5, 5-15, 15-30, 30-60 e 60-100 cm e para a seleção de variáveis adotou-se dois métodos, um por correlação $|< 0,95|$ e outro por importância (recursive feature elimination - RFE). Foram testados sete diferentes algoritmos de machine learning e executou-se o processo de seleção e modelagem (treinamento, validação, teste e predição) 100 vezes para cada profundidade com diferentes subconjuntos de amostras. Para as previsões de mudanças no estoque do SOC entre os anos 2020 e 2100, utilizou-se os dados de mudanças do clima projetados em cenários contrastantes do CMIP6 (ssp126 e ssp585), previstos pelos modelos CanESM5 e MIROC6. Ao final elaborou-se mapas finais médios e mapas de incerteza. O algoritmo com a melhor performance e selecionado para a predição foi o Random Forest (rf). O melhor desempenho do rf é explicado pela sua capacidade em detectar relações não lineares e hierárquicas entre o estoque de SOC e os preditores ambientais. O processo de seleção de variáveis, reduziu em 50% o número de preditores por correlação ($r > |0,95|$), reduzindo o tempo de processamento. A classe de solo foi um dos preditores que mais influenciou o estoque de SOC, sobretudo nas camadas superficiais (0-5, 5-15 e 15-30 cm). Além disso o clima foi outro importante preditor, especialmente nas camadas subsuperficiais (30-60, 60-100 cm). Mais da metade dos 2.530,91 Tg de SOC presentes em um metro de profundidade estão concentrados nos primeiros 30 cm de solo. O maior estoque de SOC estão na região sul do estado, nas baixas altitudes e locais com dinâmica de alagamento sazonal e os menores foram encontrados na região central do estado, sobretudo em superfície. Os menores valores de SOC estão associados ao elevado intemperismo dos Latossolos e a profunda antropização da área. A maior densidade de SOC estão nos Gleissolos

e Organossolos, consequência de sua gênese e limitação de decomposição pela hidromorfia. As regiões de maiores altitudes sofrerão elevada redução do estoque de SOC nos cenários futuros, especialmente no cenário ssp585. No entanto nas áreas onde o estoque de SOC é limitado, as projeções indicam aumento. Nos dois modelos climáticos (CanESM5 e MIROC6) as diferenças de impacto no SOC serão nas áreas com mais carbono, sendo o CanESM5 com as maiores variações no clima e no SOC demonstrando o quanto o SOC de Rondônia pode ser dependente do clima. A dispersão dos pontos observados e preditos do modelo ajustado comparado a outros estudos a nível nacional e internacional, mostra a importância em modelar em menores extensões para alcançar maior precisão. Embora exista forte relação dos preditores climáticos com o SOC, recomenda-se cautela na interpretação dos resultados, sobretudo nas áreas onde o SOC tende a aumentar nas projeções futuras. Rondônia encontra-se em região com baixa amplitude climática, portanto, o modelo ajustado realizou predições em cenários fora do seu alcance de treinamento.

Palavras-chave: aprendizado de máquina. Amazônia. mudanças climáticas. random forest. solo tropical. pedometria

ABSTRACT

SANTOS, Cássio Marques Moquedace, M.Sc., Universidade Federal de Viçosa, July 2021. **Modeling and mapping of soil organic carbon stock in the state of Rondônia**. Advisor: Renildes Lúcio Ferreira Fontes. Co-advisors: Irene Maria Cardoso and Lucas de Carvalho Gomes.

Climate changes caused by anthropic activities have been one of the main environmental issues around the world. The soil can act as a moderator of these changes, as it stores two thirds of all terrestrial carbon and small changes in this stock can increase the concentrations of greenhouse gases (GHG) and aggravate global warming. Therefore, the objective was to model and map the soil organic carbon stock in the state of Rondônia, at present and in future scenarios until 2100. Data from almost three thousand soil profiles from the study were used for the second approximation of the Socioeconomic Ecological Zoning of Rondônia. SOC stock were stratified at depths of 0-5, 5-15, 15-30, 30-60 and 60-100 cm and for the selection of variables two methods were adopted, one by correlation |< 0.95 | and another by importance (recursive feature elimination - RFE). Seven different machine learning algorithms were tested, and the selection and modeling process (training, validation, testing, and prediction) was performed 100 times for each depth with different subsets of samples. For the predictions of changes in the SOC stock between the years 2020 and 2100, we used the climate change data projected in contrasting CMIP6 scenarios (ssp126 and ssp585), predicted by the CanESM5 and MIROC6 models. At the end, average final maps and uncertainty maps were elaborated. The algorithm with the best performance and selected for the prediction was Random Forest (rf). The better performance of rf is explained by its ability to detect non-linear and hierarchical relationships between SOC stock and environmental predictors. The variable selection process reduced by 50% the number of predictors per correlation (r > |0.95|), reducing processing time. Soil class was one of the predictors that most influenced SOC stock, especially in the superficial layers (0-5, 5-15 and 15-30 cm). Furthermore, climate was another important predictor, especially in the subsurface layers (30-60, 60-100 cm). More than half of the 2,530.91 Tg of SOC present in one meter of depth are concentrated in the first 30 cm of soil. The largest SOC stock are in the southern region of the state, at low altitudes and places with seasonal flooding dynamics, and the smallest were found in the central region of the state, especially on the surface. The lower SOC values are associated with the high weathering of the Latossolos and the deep anthropization of the area. The highest density of SOC is found in Gleissolos and Organossolos, a consequence of their genesis and limitation of decomposition by hydromorphism. Higher

altitude regions will suffer a high reduction in SOC stock in future scenarios, especially in the ssp585 scenario. However, in areas where SOC stock is limited, projections indicate an increase. In the two climate models (CanESM5 and MIROC6) the differences in impact on the SOC will be in areas with more carbon, with CanESM5 with the greatest variations in climate and in SOC demonstrating how much the SOC of Rondônia can be climate dependent. The dispersion of observed and predicted points of the adjusted model compared to other studies at national and international level, shows the importance of modeling in smaller extensions to achieve greater precision. Although there is a strong relationship between climate predictors and the SOC, caution is recommended when interpreting the results, especially in areas where the SOC tends to increase in future projections. Rondônia is in a region with low climatic amplitude, therefore, the adjusted model made predictions in scenarios beyond its training range.

Keywords: machine learning. Amazon. climate changes. random forest. tropical soil. pedometrics

SUMÁRIO

Introdução geral

Na última década intensificou-se a preocupação com o impacto da perturbação antropogênica nos ecossistemas naturais, ocasionado pelas atividades agrícolas, sobretudo pelo modelo hegemônico de agricultura, conhecido como agricultura convencional (Altieri and Nicholls, 2020). As alterações da paisagem, fruto das atividades antropogênicas relacionadas ao modelo agrícola hegemônico, provocam diversos problemas ambientais, que alteram as dinâmicas globais e locais do clima e têm contribuído para uma acelerada reconfiguração da biosfera (Lohbeck et al., 2018).

A temperatura global média aumentou em 0,93 ± 0,07 °C na década de 2009-2018, comparado a linha de base pré-industrial (1850-1900), efeito do acréscimo na concentração dos gases de efeito estufa (GEE) (Olsson et al., 2019). O solo, pode atuar como moderador dessas alterações, uma vez que armazena dois terços de todo carbono terrestre e a depender das atribuições de uso, relevo e clima que estão condicionados pode agir como fonte ou dreno dos GEE (Rumpel et al., 2020; Smith, 2012). Pequenas mudanças no carbono orgânico do solo (SOC) podem impactar de forma significativa a concentração dos GEE e consequentemente alterar as projeções de feedback do clima globalmente (Gonçalves et al., 2021). Além do clima, o SOC está diretamente relacionado a segurança alimentar, qualidade da água e fertilidade do solo, o que destaca ainda mais, a necessidade de conhecer sua distribuição espaço-temporal em escalas regionais e globais (Li et al., 2021; Ramesh et al., 2019).

Modelos de sistema terrestre são importantes estimadores do SOC presente e de simulações futuras (Luo et al., 2016). No entanto, estes modelos podem não representar de forma adequada as características ambientais inerentes a formação do solo em regiões específicas, o que pode aumentar a incerteza das previsões e não representar adequadamente a heterogeneidade espacial do SOC (Todd-Brown et al., 2013). Mais de uma década de pesquisa com modelagem de SOC, evidenciam a importância de considerar variações regionais nos modelos preditivos (Jenkinson, 1990; Trumbore, 2009).

A utilização do mapeamento digital de solos com a aplicação de técnicas de machine learning pode ser uma maneira de reduzir estas incertezas. Nesta abordagem, são estabelecidas relações entre observações de campo e preditores ambientais representativos a área e a partir disso realizam-se predições do SOC para áreas não amostradas (Fathizad et al., 2020; McBratney et al., 2003). As técnicas de machine learning são capazes de detectar relações não lineares entre os preditores ambientais e o SOC, o que permite realizar predições de mapas de SOC mais plausíveis a realidade de campo. Com o avanço no poder de processamento computacional e softwares, existe uma diversidade de modelos de machine learning

disponíveis, que podem ser mais ou menos adequados a depender das condições de estudo e disponibilidade de dados a serem trabalhados (Khaledian and Miller, 2020). Isto mostra, a importância de testar diferentes algoritmos para avaliar o mais adequado a situação de estudo.

Outros fatores são determinantes na redução de incertezas relacionadas a distribuição do SOC. Estratificar as camadas preditas por exemplo, pode auxiliar na compreensão da dinâmica do SOC em superfície, que é mais sensível a evolução na forma de CO_2, comparado ao SOC em profundidade (Jansson and Hofmockel, 2020). A densidade amostral, pode ser outro fator limitante nesta técnica, pois, com um baixo quantitativo de amostras de SOC o modelo pode não identificar os padrões adequados e aumentar a incerteza das predições (Somarathna et al., 2017).

A Amazônia é um bioma brasileiro que contém as florestas tropicais mais extensas do mundo e que armazenam altas quantidades de carbono, desempenhando importante função na atenuação das mudanças climáticas (Barros Ferraz et al., 2005). Devido à dificuldade operacional de acesso à região amazônica, este bioma contém uma das menores densidade de pontos amostrais de solos no Brasil. Estimativas do inventário de estoque de SOC para essa região estão inseridas em abordagens nacionais (Gomes et al., 2019; Vasques et al., 2017) e globais (Hengl et al., 2017), com baixas resoluções espaciais e densidade amostral.

No entanto, Rondônia, é um dos estados inseridos na Amazônia que contém um denso levantamento de solos realizado na década de 1990 e que não foi utilizado na construção dos modelos supracitados (Cochrane and Cochrane, 2006). O estado sofreu intensa supressão de suas florestas nativas nos últimos 35 anos (Souza et al., 2020), resultantes principalmente dos programas de colonização propostos pelo Estado Brasileiro para ocupação da área e redução de conflitos agrários em outras regiões brasileiras (Goza, 1994; Millikan, 1992). Devido à pressão internacional diante da destruição da floresta, o detalhado levantamento dos solos, geologia, geomorfologia e fauna da região amostrou quase 3.000 perfis de solo em todo o estado, com principal objetivo de manejar de forma sustentável a natureza (Cochrane and Cochrane, 2006).

Inventariar o estoque de SOC presente e avaliar suas mudanças temporais em menor escala espacial é fundamental no planejamento de uso da terra bem como nos impactos do SOC em cenários de mudanças climáticas. Portanto, estruturou-se a dissertação em introdução geral, dois capítulos e considerações finais. O capítulo 1, intitulado “Spatial distribution of soil organic carbon stock in Rondônia predicted with machine learning”, com objetivo em modelar e mapear o estoque de carbono orgânico do solo em Rondônia. O capítulo 2, intitulado “Dynamics of soil organic carbon stock in future climate change scenarios in southwestern

Amazonia", objetivou-se modelar e mapear o estoque de carbono orgânico do solo no estado de Rondônia em cenários futuros até 2100.

Referências

Altieri, M.A., Nicholls, C.I., 2020. Agroecology and the reconstruction of a post-COVID-19 agriculture. Journal of Peasant Studies 47, 881–898. https://doi.org/10.1080/03066150.2020.1782891

Barros Ferraz, S.F., Vettorazzi, C.A., Theobald, D.M., Ballester, M.V.R., 2005. Landscape dynamics of Amazonian deforestation between 1984 and 2002 in central Rondônia, Brazil: Assessment and future scenarios. Forest Ecology and Management 204, 69–85. https://doi.org/10.1016/j.foreco.2004.07.073

Cochrane, T.T., Cochrane, T.A., 2006. Diversity of the land resources in the Amazonian state of Rondônia, Brazil. Acta Amazonica 36, 91–102. https://doi.org/10.1590/s0044-59672006000100011

Fathizad, H., Ardakani, M.A.H., Heung, B., Sodaiezadeh, H., Rahmani, A., Fathabadi, A., Scholten, T., Taghizadeh-Mehrjardi, R., 2020. Spatio-temporal dynamic of soil quality in the central Iranian desert modeled with machine learning and digital soil assessment techniques. Ecological Indicators 118, 106736. https://doi.org/10.1016/j.ecolind.2020.106736

Gomes, L.C., Faria, R.M., Souza, E., Veloso, G.V., Schaefer, C.E.G.R., Fernandes Filho, E.I., 2019. Modelling and mapping soil organic carbon stocks in Brazil. Geoderma 340, 337–350. https://doi.org/10.1016/j.geoderma.2019.01.007

Gonçalves, D.R.P., Mishra, U., Wills, S., Gautam, S., 2021. Regional environmental controllers influence continental scale soil carbon stocks and future carbon dynamics. Scientific Reports 11, 6474. https://doi.org/10.1038/s41598-021-85992-y

Goza, F., 1994. Brazilian frontier settlement: The case of Rondônia. Population and Environment 16, 37–60. https://doi.org/10.1007/BF02208002

Hengl, T., 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., 2017. SoilGrids250m: Global gridded soil information based on machine learning. PLoS ONE 12, e0169748. https://doi.org/10.1371/journal.pone.0169748

Jansson, J.K., Hofmockel, K.S., 2020. Soil microbiomes and climate change. Nature Reviews Microbiology 18, 35–46. https://doi.org/10.1038/s41579-019-0265-7

Jenkinson, D.S., 1990. The turnover of organic carbon and nitrogen in soil. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 329, 361–368. https://doi.org/10.1098/rstb.1990.0177

Khaledian, Y., Miller, B.A., 2020. Selecting appropriate machine learning methods for digital soil mapping. Applied Mathematical Modelling 81, 401–418. https://doi.org/10.1016/j.apm.2019.12.016

Li, H., Wu, Y., Chen, J., Zhao, F., Wang, F., Sun, Y., Zhang, G., Qiu, L., 2021. Responses of soil organic carbon to climate change in the Qilian Mountains and its future projection. Journal of Hydrology 596, 126110. https://doi.org/10.1016/j.jhydrol.2021.126110

Lohbeck, M., Winowiecki, L., Aynekulu, E., Okia, C., Vågen, T.G., 2018. Trait-based approaches for guiding the restoration of degraded agricultural landscapes in East Africa. Journal of Applied Ecology 55, 59–68. https://doi.org/10.1111/1365-2664.13017

Luo, Y., Ahlström, A., Allison, S.D., Batjes, N.H., Brovkin, V., Carvalhais, N., Chappell, A., Ciais, P., Davidson, E.A., Finzi, A., Georgiou, K., Guenet, B., Hararuk, O., Harden, J.W., He, Y., Hopkins, F., Jiang, L., Koven, C., Jackson, R.B., Jones, C.D., Lara, M.J., Liang, J., McGuire, A.D., Parton, W., Peng, C., Randerson, J.T., Salazar, A., Sierra, C.A., Smith, M.J., Tian, H., Todd-Brown, K.E.O., Torn, M., van Groenigen, K.J., Wang, Y.P., West, T.O., Wei, Y., Wieder, W.R., Xia, J., Xu, Xia, Xu, Xiaofeng, Zhou, T., 2016. Toward more realistic projections of soil carbon dynamics by Earth system models. Global Biogeochemical Cycles 30, 40–56. https://doi.org/10.1002/2015GB005239

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4

Millikan, B.H., 1992. Tropical deforestation, land degradation, and society: lessons from Rondonia, Brazil. Latin American Perspectives 19, 45–72. https://doi.org/10.1177/0094582X9201900103

Olsson, L., Barbosa, Humberto, Bhadwal, Suruchi, Manuel Moreno, J., Vera, C., Salisu Barau, A., Barbosa, H, Bhadwal, S, Cowie, A., Delusca, K., Flores-Renteria, D., Hermans, K., Jobbagy, E., Kurz, W., Li, D., Shukla, P., Skea, J., Calvo Buendia, E., Masson-Delmotte, V., Pörtner, H., Roberts, D.C., Zhai, P., Slade, R., Connors, S., van Diemen, R., Ferrat, M., Haughey, E., Luz, S., Neogi, S., Pathak, M., Petzold, J., Portugal Pereira, J., Vyas, P., Huntley, E., Kissick, K., Belkacemi, M., Malley, J., 2019. 4 Land degradation. Climate Change and Land: an IPCC special report on climate change, desertification, land

degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems.

Ramesh, T., Bolan, N.S., Kirkham, M.B., Wijesekara, H., Kanchikerimath, M., Srinivasa Rao, C., Sandeep, S., Rinklebe, J., Ok, Y.S., Choudhury, B.U., Wang, H., Tang, C., Wang, X., Song, Z., Freeman, O.W., 2019. Soil organic carbon dynamics: Impact of land use changes and management practices: A review, in: Advances in Agronomy. Academic Press Inc., pp. 1–107. https://doi.org/10.1016/bs.agron.2019.02.001

Rumpel, C., Amiraslani, F., Chenu, C., Garcia Cardenas, M., Kaonga, M., Koutika, L.S., Ladha, J., Madari, B., Shirato, Y., Smith, P., Soudi, B., Soussana, J.F., Whitehead, D., Wollenberg, E., 2020. The 4p1000 initiative: Opportunities, limitations and challenges for implementing soil organic carbon sequestration as a sustainable development strategy. Ambio 49, 350–360. https://doi.org/10.1007/s13280-019-01165-2

Smith, P., 2012. Soils and climate change. Current Opinion in Environmental Sustainability 4, 539–544. https://doi.org/10.1016/j.cosust.2012.06.005

Somarathna, P.D.S.N., Minasny, B., Malone, B.P., 2017. More Data or a Better Model? Figuring Out What Matters Most for the Spatial Prediction of Soil Carbon. Soil Science Society of America Journal 81, 1413–1426. https://doi.org/10.2136/sssaj2016.11.0376

Souza, C.M., Shimbo, J.Z., Rosa, M.R., Parente, L.L., Alencar, A.A., Rudorff, B.F.T., Hasenack, H., Matsumoto, M., Ferreira, L.G., Souza-Filho, P.W.M., de Oliveira, S.W., Rocha, W.F., Fonseca, A. v., Marques, C.B., Diniz, C.G., Costa, D., Monteiro, D., Rosa, E.R., Vélez-Martin, E., Weber, E.J., Lenti, F.E.B., Paternost, F.F., Pareyn, F.G.C., Siqueira, J. v., Viera, J.L., Neto, L.C.F., Saraiva, M.M., Sales, M.H., Salgado, M.P.G., Vasconcelos, R., Galano, S., Mesquita, V. v., Azevedo, T., 2020. Reconstructing three decades of land use and land cover changes in brazilian biomes with landsat archive and earth engine. Remote Sensing 12, 2735. https://doi.org/10.3390/RS12172735

Todd-Brown, K.E.O., Randerson, J.T., Post, W.M., Hoffman, F.M., Tarnocai, C., Schuur, E.A.G., Allison, S.D., 2013. Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. Biogeosciences 10, 1717–1736. https://doi.org/10.5194/bg-10-1717-2013

Trumbore, S., 2009. Radiocarbon and soil carbon dynamics. Annual Review of Earth and Planetary Sciences 37, 47–66. https://doi.org/10.1146/annurev.earth.36.031207.124300

Vasques, G.M., Dart, R.O., Baca, J.F.M., Ceddia, M.B., Santos, M. de L.M., 2017. Soil Organic Carbon Stock at 0-30 cm Map for Brazil, Embrapa. Rio de Janeiro.

Chapter 1: Spatial distribution of soil organic carbon stock in Rondônia predicted with machine learning

Cássio Marques Moquedace dos Santos[1], Lucas de Carvalho Gomes[1], Marcos Guedes de Lana[1], Gustavo Vieira Veloso[1], Elpídio Inácio Fernandes-Filho[1], Irene Maria Cardoso[1], Emanuel Fernando Maia de Souza[2], Renildes Lúcio Ferreira Fontes[1]

[1]Department of Soils, Universidade Federal de Viçosa, Brazil; [2]Department of Forestry Engineering, Federal University of Rondônia, Brazil

Abstract

Soil stores two-thirds of all terrestrial carbon, and depending on land use and management, it can be a source or sink of greenhouse gases. Therefore, the objective was to model and map the soil organic carbon (SOC) stock in Rondônia. Data from almost three thousand soil profiles from the second approximation of the Socioeconomic Ecological Zoning of Rondônia were used in this study. SOC stock were stratified at depths of 0-5, 5-15, 15-30, 30-60 and 60-100 cm and for the selection of environmental predictors two methods were adopted, one per correlation |< 0.95| and another by importance (recursive feature elimination - RFE). To reduce uncertainties and select the best model, seven different machine learning algorithms were tested. The selection and modeling process (training, testing and prediction) were performed 100 times for each depth with different subsets of samples. In the end, mean and quartile maps (Q05 and Q95) and uncertainty maps (Coefficient of variation) were elaborated. In addition, the SOC stock was stratified by soil class and protected areas. The model with the best performance and selected for the prediction was the rf. The variable selection process reduced processing time. The soil class was one of the predictors that most influenced the SOC stock, especially in the superficial layers (0-30 cm). Furthermore, the climate was another important predictor, especially in the subsurface layers (30-100 cm). Regarding vegetation, the enhanced vegetation index (EVI) influenced the SOC stock in the 0-5 cm layer, which shows the importance of vegetation for the surface SOC. More than half of the 2,530.91 Tg of SOC in one meter of soil is in the first 30 cm. The largest SOC stock are in the southern region of Rondônia, at low altitudes and places with seasonal flooding dynamics, and the smallest are found in the central region, associated with the high weathering Latossolos and the deep anthropization of the area. The highest density of SOC stock is found in Gleissolos and Organossolos, a consequence of their genesis and limitation of decomposition by hydromorphism. The highest SOC stock are in unprotected areas due to their size, while the highest density is in protected

areas. This reaffirms the importance of promoting more sustainable production systems. The data used are from 1996/1997 and from this period to the present, the state has undergone strong anthropization, suggesting that the data presented here may be overestimated. The dispersion of observed and predicted points of the fitted model for this study compared to global and national models shows the importance of modeling in smaller extensions to achieve greater precision.

Keywords: Amazon, random forest, protected areas, recursive feature elimination

1. Introduction

Soil stores two-thirds of all terrestrial carbon and, depending on land use and management, it can be a source or sink of greenhouse gases (Rumpel et al., 2020; Smith, 2012), playing an essential role in the climate change mitigation (Lal, 2004; Minasny et al., 2017). However, accurate spatial quantification of soil organic carbon (SOC) stock is still scarce, especially in developing countries with low density of soil samples. More accurate quantification of SOC stock can help to refine carbon inventories at national and international levels, but also improve our understanding it spatial distribution, as well as contribute to environmental conservation in the management of areas with high concentrations subject to human activities. The spatial distribution of SOC stock can be performed using digital soil mapping techniques, using remote sensing associated with machine learning (Gomes et al., 2019; McBratney et al., 2003).

Digital soil mapping has advanced in the last decades and works with SOC stock have been carried out from local to global scales (Chen et al., 2018; Y. Li et al., 2021). For SOC stock mapping, local samples are needed to build the models, and the representativeness of these samples is vital to fit a model with good generalization. Studies show that sample density influences performance and error reduction, and models fitted with higher sample density can produce more plausible maps with the reality (Ng et al., 2020; Taghizadeh-Mehrjardi et al., 2016; Yang et al., 2020).

Machine learning is capable of modeling non-linear relationships between environmental predictors and the SOC stock, which can improve the accuracy of the mapping and enable the quantification of uncertainties associated with prediction. However, there are a variety of models available, with different characteristics in the data analysis process, which can accumulate gains and/or losses depending on the study situation (Khaledian and Miller, 2020). Therefore, it is important to test different algorithms to identify the most suitable one

for the respective analysis conditions. In addition to inventorying the total SOC stock, the stratification of these data in depth to understand the SOC dynamics, since carbon stored at depth is less likely to be lost as CO_2 than surface carbon (Jansson and Hofmockel, 2020).

The Amazon is one of the most extensive tropical forest areas in the world, which in addition to harboring a high diversity of species, also stores carbon in the soil and biomass, playing an important role in mitigating climate change (Barros Ferraz et al., 2005). Due to the operational difficulty of accessing the Amazon region, this biome contains one of the lowest densities of soil sampling points in Brazil, which generates great uncertainty in the prediction of SOC stock for this region (Gomes et al., 2019; Vasques et al., 2017). Quantifying and spatializing the SOC stock in these areas is essential for understanding carbon inputs and outputs from deforestation. The state of Rondônia, for example, is located in the Amazon biome and has lost more than 30% of its native coverage in the last 35 years (Souza et al., 2020), as a result of deforestation, fires, expansion of the agricultural frontier, especially pasture and monoculture (Bullock et al., 2020).

Despite the lack of sampling points for the Amazon biome, the state of Rondônia carried out a detailed survey of the region's soils in the 1990s, with more than 2,900 profiles, distributed throughout the state's territorial extension. This database is now available online but has not been used in current SOC stock projections at the national (Gomes et al., 2019; Vasques et al., 2017) and international (Hengl et al., 2017) levels. Then, mapping the SOC stock at the regional level with a broader database has the potential to improve it quantification and distribution and better guide public policy makers in the context of fostering strategies for the conservation of SOC stock and discourage activities that promote emissions.

The objective of this study was to model and map the SOC stock in Rondônia. To this end, it was specifically aimed at: i) selecting and evaluating the performance of different machine learning algorithms; ii) stratify stock into layers and prepare maps of average, upper quartile (Q95), lower quartile (Q05) and coefficient of variation of SOC stock; iii) stratify SOC stock into protected, unprotected and soil types; and iv) compare the adjusted model to other SOC stock mappings covering the territorial extension of the state.

2. Material and methods

The state of Rondônia is located in the northern region of Brazil, along the border with Bolivia (Fig. 1). Rondônia originally held most of its territory covered by the Amazon rainforest, with exceptions in the southern region of the state, where there are patches of Cerrado (Eiten, 1972). From 1985 to 2019 the state lost more than 67,804 km^2 of its native

formation, equivalent to the suppression of more than 31% of its native cover (Souza et al., 2020).

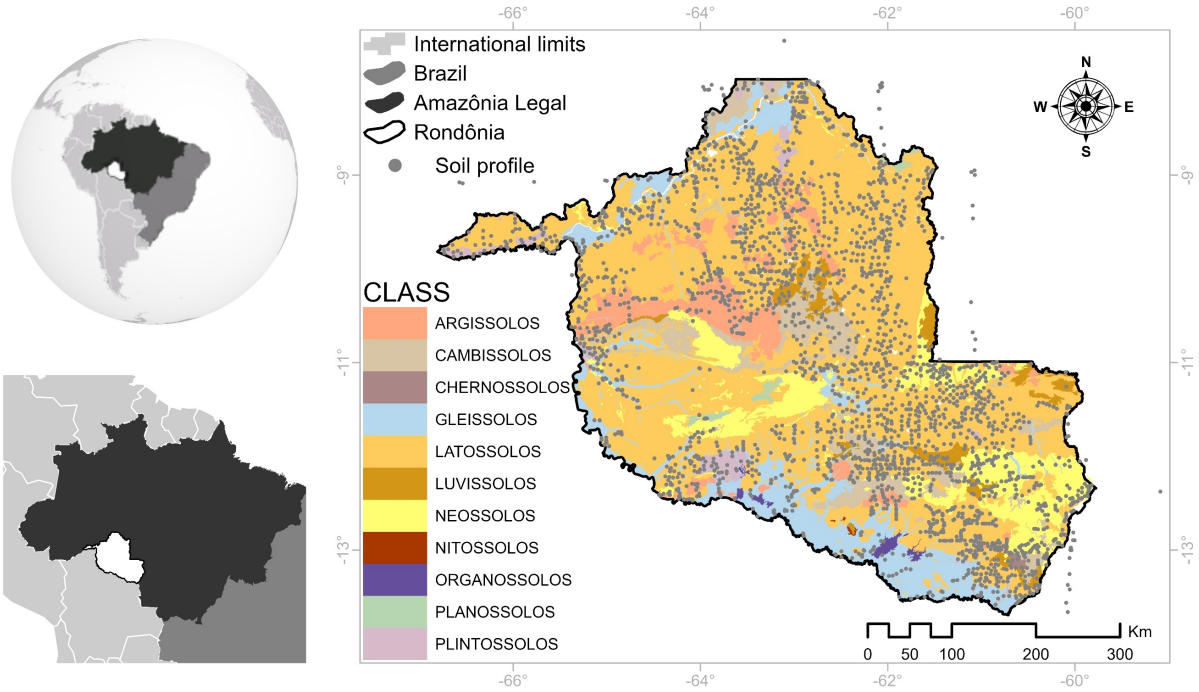

Fig. 1. Distribution of soil classes and profiles in the state of Rondônia.

The state of Rondônia is subdivided into 52 municipalities, distributed over an approximate area of 225,076.83 km^2 and represents approximately 5.3% of the Legal Amazon (IBGE, 2012; Pereira and Goncalves, 2021). The state's climate is of the Am type (tropical monsoon), with a well-defined dry season, between May and October (Alvares et al., 2013). Rains are intense between October and April, with an average annual precipitation of 2,300 mm; the relative humidity of the air is between 80% and 90% e; the average temperature is approximately 26°C. The altitude in the state varies between 33 and 1,100 m, with an approximate average of 200 m (NASA JPL, 2020). The spatial distribution of the state covers time zones 19, 20, and 21, and for the study in question the projected coordinate system was adopted: South America Lambert Conformal Conic (ESRI:102015).

2.1. Soil data

The soil data used are public, coming from the study for the second approximation of the Socioeconomic Ecological Zoning of Rondônia (ZEERO), available at the State Secretariat for Environmental Development (SEDAM) of Rondônia. Due to the low level of information at the time (the 1990s) for carrying out the second ZEERO approximation, there was an intensive fieldwork, surveying 2,914 soil profiles (Fig. 1), with physical, chemical and organic

carbon content of the soil (SOC) at different depths, covering the entire territory of the state (Cochrane and Cochrane, 2006; SEDAM, 2021). This dataset was collected over one year (1996/1997) and was analyzed by the same laboratory, accumulating advantages by reducing sources of error and providing SOC contents with short intervals between collections and analyses. SOC was measured following the Walkley-Black methodology in air-dried fine earth fraction (< 2 mm) (Cochrane and Cochrane, 2006; Nelson and Sommers, 1982). The data from soil profiles were used to calculate the SOC stock through equation (1) (Veldkamp, 1994).

$$\text{SOC stock} = \frac{\text{SOC} \times \text{BD} \times \text{e}}{10} \quad (1)$$

Where: SOC stock = Stock of soil organic carbon (Mg ha^{-1}); SOC = Organic carbon content in the soil (g kg^{-1}); BD = Bulk density of soil (g cm^{-3}); e = Thickness of the layer considered (cm).

The data set did not contain all parameters at all sampled points, such as soil density, necessary for calculating the SOC stock (Equation 1). Soil density can be relatively difficult to determine, since the presence of plant roots, and soil moisture, texture, relief, and stony conditions can make it impossible to collect undisturbed samples. Therefore, when necessary, densities were estimated using a pedotransfer function adjusted from other soil parameters available in all profiles. This function was adjusted using machine learning models and the selected model was the svmRadial (Karatzoglou et al., 2004) with root mean squared error - RMSE = 0.080 g cm^{-3}, coefficient of determination - R^2 = 0.751 and mean absolute error - MAE = 0.052 g cm^{-3} (Table S1 and Table S2, Supplementary material).

SOC stock were stratified at depths of 0-5, 5-15, 15-30, 30-60 and 60-100 cm as proposed by GlobalSoilMap (Hengl et al., 2017). For subdivision of SOC stock at these depths, data were harmonized from splines using the "spline" function of the "GSIF" package (Hengl, 2020) in the free software R (R Core Team, 2021).

2.2. Environmental predictors

Initially, 112 environmental predictors closely related to SOC dynamics were used, capable of altering the carbon cycle in the system (Akpa et al., 2016). The environmental predictors (Table S3, Supplementary material) used in the model represent the different elements of the SCORPAN model, quantitatively expressing the relationships with the SOC stock, being an extension of the five soil formation factors already well established in the literature (McBratney et al., 2003; Pendleton and Jenny, 1945).

The digital elevation model (DEM) used was NASADEM (NASA JPL, 2020). This model has an approximate spatial resolution of 30 m and is derived from telemetry data from the Shuttle Radar Topography Mission (SRTM). From the DEM, 48 morphometric variables were derived using the packages “RSAGA” (Brenning et al., 2018) and “rgrass7” (Bivand, 2021) of the free software R (R Core Team, 2021).

The historical variables of temperature, precipitation and bioclimatic were originated from Worldclim with a spatial resolution of approximately 1 km (Fick and Hijmans, 2017). The geology and soil class data were taken from the ZEERO second approximation survey (Cochrane and Cochrane, 2006; SEDAM, 2021). Soil class data were subdivided into the third categorical level, rare were grouped soils according to SOC samples.

The enhanced vegetation index (EVI), modified soil adjusted vegetation index (MSAVI), normalized burn ratio (NBR), normalized burn ratio 2 (NBR2), normalized difference moisture index (NDMI), normalized difference vegetation index (NDVI) and soil adjusted vegetation index (SAVI) were calculated in the Google Earth Engine platform from images from the “USGS Landsat 5 Surface Reflectance Tier 1” collection (Gorelick et al., 2017). The images have a spatial resolution of 30 m and a temporal resolution of 16 days. An image was generated for each index composed of the average of the pixels of each image available between the period of 01/01/1984 and 12/31/1984 with a cloud filter smaller than 10%. The indices were produced for the first year of satellite imagery available for the region (1984) to generate a predictor that more likely corresponded to the characteristics of the native vegetation of Rondônia.

All environmental predictors were harmonized to 200 m spatial resolution using the cubic spline method of the “gdalUtilities” package (O’Brien, 2021) and redesigned for the datum: South America Lambert Conformal Conic (ESRI:102015) with the aid of the package “raster” (Hijmans, 2021).

2.3. Variable selection

Variable selection was used in this study to avoid overfitting or underfitting or to reduce computational effort (Cole et al., 2014). Overfitting is when the model is adjusted to the point of capturing even the noise of the training data, and underfitting is when the model cannot find patterns of relationship between predictor and predicted variables, achieving low performance already in training, in both situations the model loses generalizability. For the selection of variables, two methods were adopted, one by correlation and the other by importance. In the method by correlation, Spearman's correlation was used, establishing a cutoff limit of |0.95|, for

this, the “findcorrelation” function of the “caret” package was applied (Kuhn, 2021). In this method, pairs of variables that have values |0.95| greater than this limit were evaluated, one of the two being removed by the multicollinearity method. Sequentially, the variables that were not removed by the previously established criteria were submitted to evaluation using the importance method.

The remaining variables were selected through the interactive recursive feature elimination (RFE) algorithm, contained in the “caret” package (Kuhn, 2021). The RFE is a back-forward selection technique that eliminates the variables that contribute least to the model, based on a measure of importance (Kuhn and Johnson, 2013). For application, we used discrete subsets ranging from 2 to 40 variables and then 5 in 5 to the total of remaining variables, and the set with the smallest number of variables (more parsimonious) and with the best predictive performance was adopted.

2.4. Machine learning algorithms

Machine learning (ML) techniques are increasingly used for predictive geospatial analyses, especially for soils (Behrens et al., 2018; Wadoux et al., 2020). To achieve less uncertainty in the estimates and select the best model for the final prediction, seven different algorithms were tested: i) cubist, ii) multivariate adaptive regression spline (earth), iii) stochastic gradient boosting (gbm), iv) elastic-net regularized generalized linear models (glmnet), v) k-nearest neighbors (k-knn), vi) random forest (rf) and vii) support vector machines radial kernel (svmRadial).

2.4.1. Cubist

The cubist model was initially based on the M5 model tree (Kuhn and Johnson, 2013). The method of this model creates “if it's after and then” rules, each of the rules contains an associated multivariate linear model, therefore, meeting the rules, the corresponding model calculates the predicted value. The number of neighbors is used to change rule-based predictions (Zhou et al., 2019).

Due to rule conditions, cubist is advantageous because it automatically handles interactions, allowing several linear models to describe linearity in various parts of the predictor (Zeraatpisheh et al., 2019). The cubist model has been used successfully in soil mapping through modeling (Gomes et al., 2019; Nguyen et al., 2019; Peng et al., 2019). The algorithm was applied with the aid of the “cubist” package (Kuhn and Quinlan, 2021) and the hyperparameters: “committees” and “neighbors” were optimized using the “caret”.

2.4.2. Multivariate adaptive regression spline (earth)

Multivariate adaptive regression (earth) was proposed by Friedman (1991) and has recently been applied to geospatial modeling (Amiri et al., 2020; Goh et al., 2018). Earth explains an easy-to-interpret model for data with high-dimensional implicit function reconstruction with reasonable efficiency and precision, not depending on the previous assumption of the functional relationship (Wang et al., 2020).

Earth is a non-parametric method that depicts the functional relationship with various parts of linear segments, where the function of each segment and the total number of segments are automatically determined. Each of the regions generated is called “nodes” while the term “basic function” is used to indicate each distinct range of predictors (Conoscenti et al., 2015).

The model is operated in two steps, in the first step it generates a complex and very fine-tuned model, adding all possible basic functions, in the second step it removes the less important basic functions. Earth reduces model complexity without degradation in overall performance (Felicísimo et al., 2013). The algorithm was adjusted with the “earth” package (Milborrow, 2020) and its two hyperparameters (“degree” and “nprune”) were optimized with the aid of the “caret”.

2.4.3. Stochastic gradient boosting (gbm)

The gbm is a hybrid technique that adopts a combination of boosting and bagging. Unlike random forest, gbm builds sequential trees, where subsequent ones are built from the residuals of the previous tree, not focusing on the complete data set. The gbm performs a boost by selecting a random sample of data at each stage, without replacement (Friedman, 2002; Rahman et al., 2020). gbm holds four hyperparameters: i) "n.trees;" ii) “interaction.depth”; iii) “shrinkage”; and iv) "n.minobsinnode". The model was adjusted using the “gbm” package (Greenwell et al., 2020). Of the four hyperparameters “interaction.depth” and “n.trees” were optimized using “caret” and the others were kept fixed.

2.4.4. Elastic-net regularized generalized linear models (glmnet)

The glmnet proposed by Friedman et al. (2010), fits a generalized linear model with penalized maximum likelihood, mixing generalized linear models with elastic network models. The model is efficient in working with large datasets, accumulating advantages when using the sparsity between the data (Kodikara and McHenry, 2020).

In glmnet each parameter is optimized by minimizing the objective function, on the other hand the remaining parameters remain fixed. The algorithm was applied with the aid of the “glmnet” package (Friedman et al., 2010) and the hyperparameters: “alpha” and “lambda” were optimized using “caret”.

2.4.5. K-nearest neighbors (k-knn)

Algorithms that use the “nearest neighbor” technique are non-parametric and intuitive in the application. Unlike knn general and k-knn in particular, which in addition to measuring distance, weights the closest observations. The distances, on which the search for nearest neighbors in the first step is based, are transformed into similarity measures, which are used as weights (Hechenbichler and Schliep, 2004).

The model was adjusted using the “kknn” package (Schliep and Hechenbichler, 2016) and has two hyperparameters, the “distance” was kept constant and the “kmax” was optimized with the “caret” package.

2.4.6. Random forest (rf)

The rf is one of the ML techniques, based on tree predictions, built from an initialization sample of data, used for training. Tree predictions are calculated as averages and these averages are used as the final predictions. In the training of the algorithm, the rf introduces an additional random perturbation to reduce the chances of overfitting during the division of a tree, progressively reducing it to smaller subsets (Wadoux et al., 2020).

This technique was applied with the aid of the “randomForest” package (Liaw and Wiener, 2002). The model has three hyperparameters for optimization, i) “ntree”, ii) “nodesize” and iii) "mtry". Of these, the “mtry” was optimized with the aid of the “caret”, testing different values, and selecting the one with the best performance, the other hyperparameters were kept fixed (Breiman, 2001; Janitza and Hornung, 2018).

2.4.7. Support vector machines radial kernel (svmRadial)

svmRadial is based on statistical learning and the maximum margin principle. In this sense, svmRadial seeks the ideal that maximizes the separation between the support vectors (Cortes and Vapnik, 1995). The kernel used for the adjustment was the radial, applied with the “kernlab” package (Karatzoglou et al., 2004; Meyer et al., 2021). The model has two hyperparameters, in which the “sigma” was kept fixed, and the “C” was optimized using the “caret”.

2.5. Modeling and prediction

SOC stock data were separated for each depth by 75% for training and cross-validation with 10 folds and 25% for external testing (Table S4, Supplementary material). The model was selected following the principle of parsimony, that is, better performance with fewer explanatory predictors. For this purpose, the models were individually evaluated with their respective sets of predictors. Performance was evaluated based on: i) mean absolute error - MAE (1); ii) coefficient of determination - R^2 (2); and iii) root mean squared error - RMSE (3).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|P_i - O_i| \quad (1)$$

$$R^2 = \frac{\sum_{i=1}^{n}(P_i - \bar{O}_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O}_i)^2} \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \quad (3)$$

Where: n represents the number of samples; P_i and O_i represent the predicted and observed values at location i, respectively.

Predictor selection, training, validation, testing, and prediction processes were performed 100 runs for each depth with different sets of samples (Fig. 2). This sequential repetition process under different possibilities of random sets of samples in training and testing is important in determining the variability of the prediction, since different arrangements must give rise to different results, impacting on accuracy (Gomes et al., 2019; Kuhn and Johnson, 2013). All analyzes were performed in the R software (R Core Team, 2021), and for training, validation, testing and prediction, the “caret” (Kuhn, 2021) and “labgeo” (Fernandes Filho, 2020) packages were used.

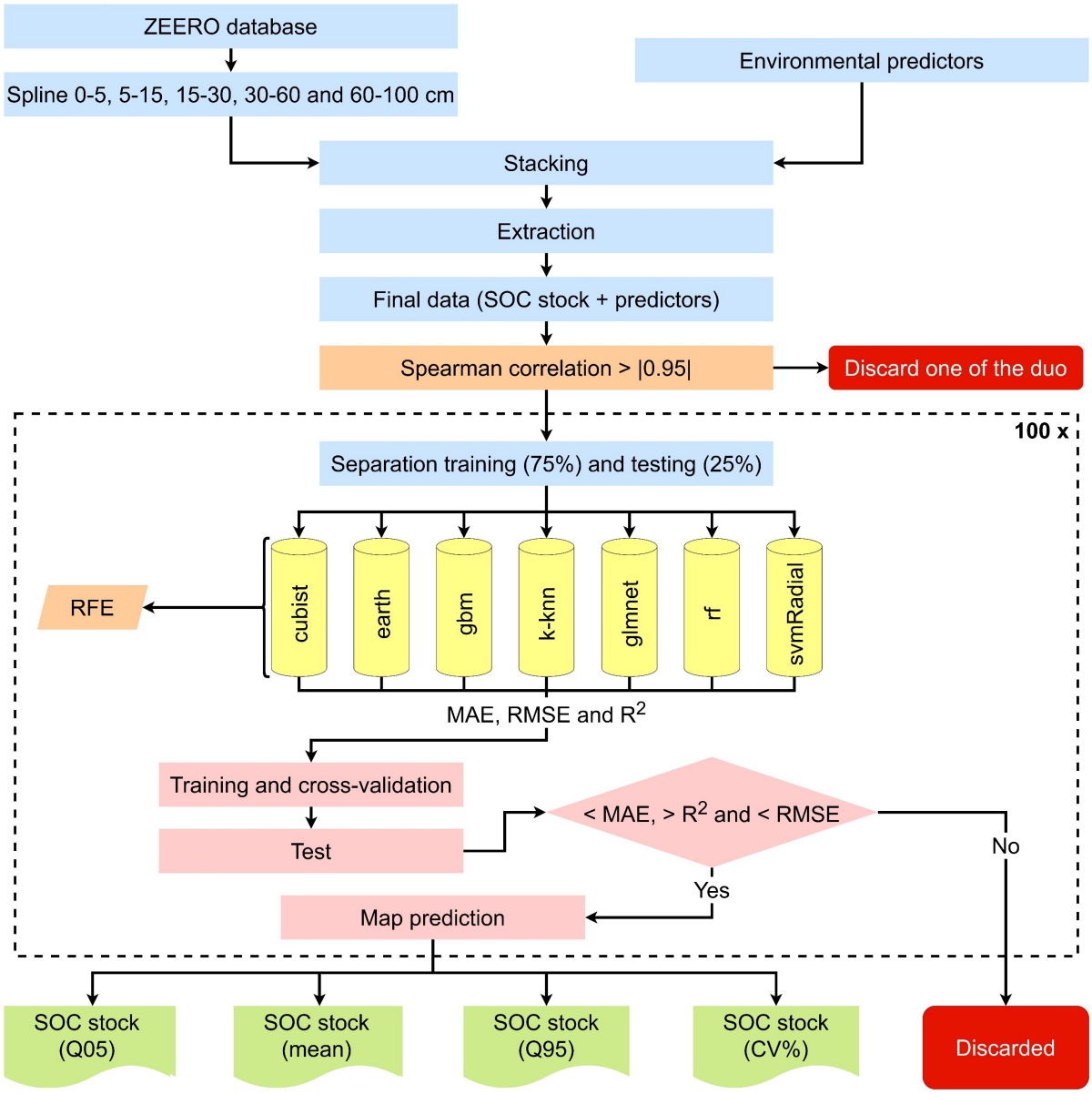

Fig. 2. Sequence of the method for modeling and prediction of soil organic carbon stock in the state of Rondônia. ZEERO = Rondônia's ecological socio-economic zoning; SOC = soil organic carbon; earth = multivariate adaptive regression spline; gbm = stochastic gradient boosting; glmnet = elastic-net regularized generalized linear models; k-knn = K-nearest neighbors; rf = random forest; svmRadial = support vector machines radial kernel; RFE = recursive feature elimination; R^2 = coefficient of determination; RMSE = root mean squared error; MAE = mean absolute error; Q5 = lower percentile; Q95 = upper percentile; CV% = coefficient of variation.

Sequentially after model selection, 100 maps were generated for each depth and from these, final maps were created for each of the depths: i) upper percentile (Q95); ii) lower percentile (Q05); iii) mean map and iv) coefficient of variation (CV) map (Arrouays et al., 2014; Gomes et al., 2019).

SOC stock were stratified by soil classes using the classes mapped in the ZEERO survey and stratified by the protected and unprotected areas classes. Protected areas were imported

from the "geobr" package (Pereira and Goncalves, 2021). The "geobr" compiles official Brazilian geospatial data on protected areas from the Ministry of the Environment and the National Indian Foundation (FUNAI). The unprotected class consisted of all land use coverage in the state that is not included as a protected area (conservation units and/or indigenous lands). The result of the 0-30 cm estimates was compared to estimates made by other authors with forecasts on a larger territorial scale and lower sample/area densities (Batjes et al., 2019; Hengl et al., 2017; Souza et al., 2020; Vasques et al., 2017).

3. Results

3.1. SOC stock data

The density of the SOC stocks of the 2,914 profiles were distributed differently at each depth, all of which had a long tail to the right, with most values smaller than the average, however, the distribution tended to flatten the curve as the depth has increased (Fig. 3). This evidences the large concentration of SOC stock values between 0 and 25 Mg ha^{-1} with natural "outliers" above 100 Mg ha^{-1}.

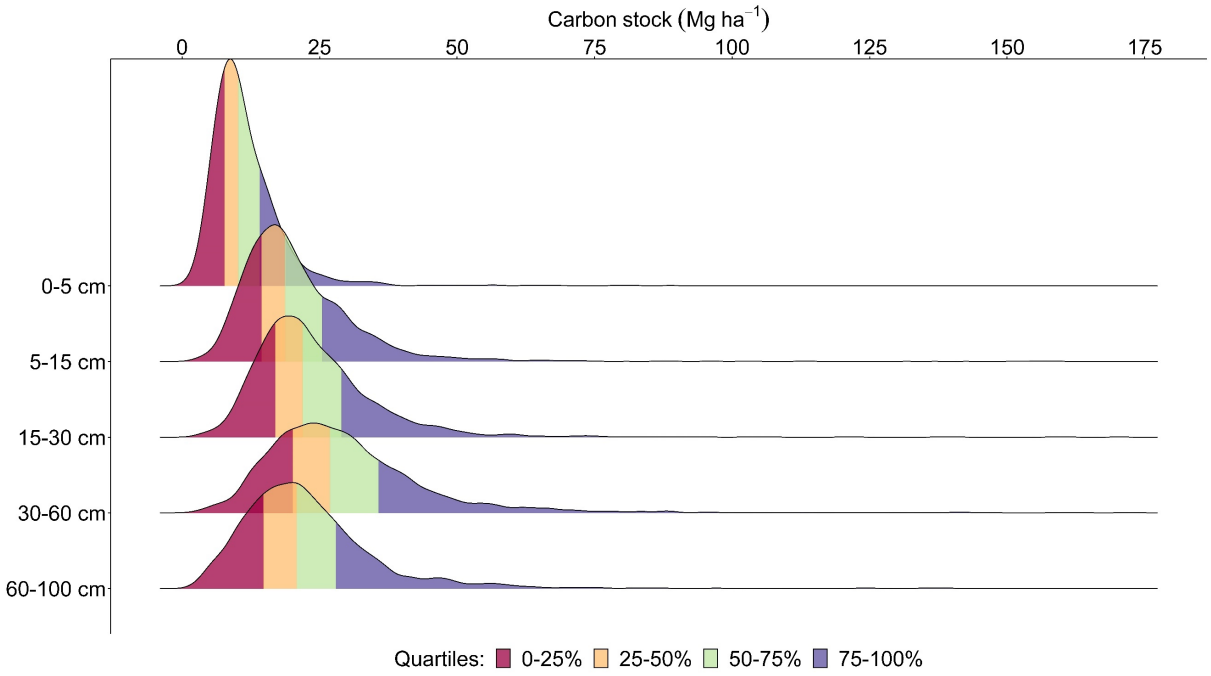

Fig. 3. Density plots of the soil organic carbon stocks in the soil of Rondônia up to 1 m depth, Rondônia State, Brazil.

3.2. Performance and model selection

There was great variation in the performance of the algorithms in the 100 runs for all depths (Fig. 4). Random forest had the highest R^2 and lowest MAE and RMSE at all depths,

while the lowest R^2 were from earth, glmnet and svmRadial varying according to depth. svmRadial was the algorithm with the least variation among all the ones evaluated.

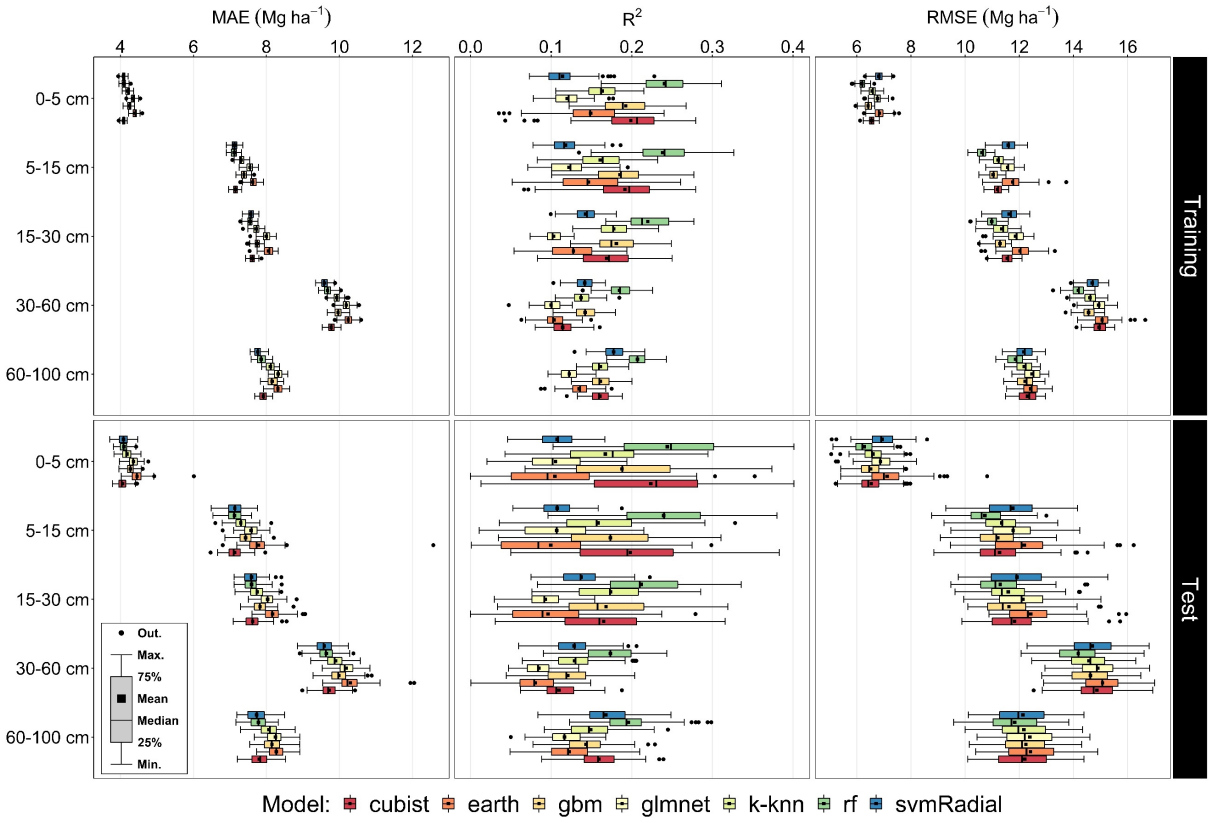

Fig. 4. Variation of the performance of the different algorithms tested to predict the soil organic carbon stock of at different depths, Rondônia State, Brazil. R^2 = coefficient of determination; RMSE = root mean squared error; MAE = mean absolute error.

Therefore, the rf was chosen to predict SOC stock for all studied depths. In general, the results of the performance of the training and test data did not vary significantly among themselves, indicating a balance between bias and variance in the adjusted models.

Except for the earth model, which showed some outliers in the quantification of errors, the algorithms were homogeneous relative to MAE and RMSE, varying less than 5 Mg C ha^{-1}. The errors increased as the depth increased, except for the 60-100 cm layer which was smaller than the 30-60 cm layer, this behavior did not follow R^2 as it was reduced while the soil depth increased.

3.3. Selection and importance of predictors

The process of eliminating predictors by Spearman correlation resulted in the selection of 65 predictors for the depths of 0-5 and 5-15 cm; 66 predictors for depths of 15-30 and 30-60 cm; and 64 predictors for the 60-100 cm depth (Table S5, Table S6, Table S7, Table S8 and

Table S9, Supplementary material). Sequentially, the use of RFE indicated an asymptotic performance trend from 12 predictors, for all depths (Figure S1, Figure S2, Figure S3, Figure S4 and Figure S5, Supplementary material). From 12 predictors, for a depth of 0-5 cm and 5-15 cm the median R^2, resulting from the algorithm, was above 0.2 median above 0.2 for a depth of 0-5 cm and 5-15 cm. While at other depths (15-30, 30-60 and 60-100 cm) the R^2 was above 0.15. Tangent to the errors, from 12 predictors the RFE was below 10 Mg ha^{-1} and 15 Mg ha^{-1} of SOC for the MAE and RMSE respectively. The average range of variation for all depths was 0.2, 1 Mg ha^{-1} and 1 Mg ha^{-1} for the R^2, MAE and RMSE respectively.

As each execution of the algorithm resulted in the selection of subgroups of different predictors for prediction, Fig. 5 is the result of the average of the 15 most important predictors for the prediction of the SOC stock. Soil classes and bioclimatic variables were the ones that most contributed to the prediction of SOC stock. In the superficial layers (0-5, 5-15 and 15-30 cm) the soil class, the precipitation of the warmest quarter (BIO18) and the precipitation of the driest quarter (BIO17) controlled more significantly the SOC stock. At depths of 30-60 and 60-100 cm, BIO18 was more important, followed by temperature seasonality (BIO4), BIO17 and soil class. The improved vegetation index (EVI), variable corresponding to native vegetation, corresponded to a little more than 40% of relative importance in predicting SOC stock in the most superficial layer (0-5 cm), not being selected for the other layers as a predictor. Predictors derived from relief expressed greater importance in the 30-60 cm layer and were less expressive in the other layers (< 35%).

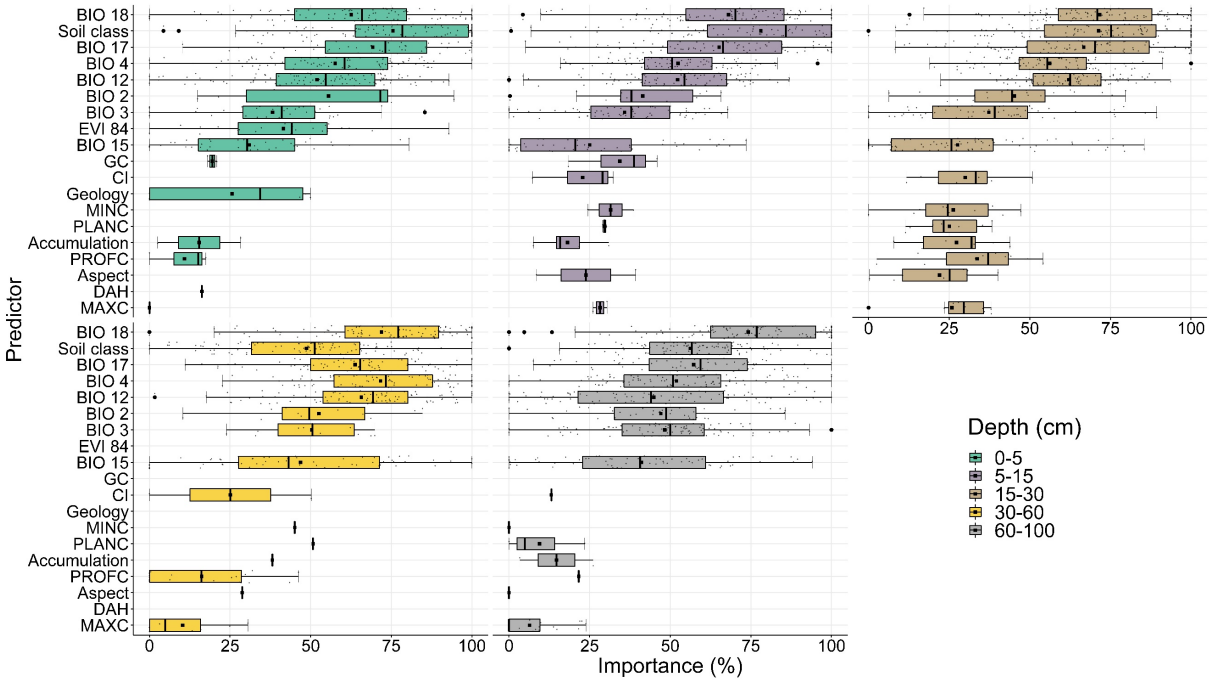

Fig. 5. Boxplot of the 15 most important predictors for the random forest algorithm in the prediction of SOC stock in Rondônia State, Brazil. BIO18 = precipitation of warmest quarter; BIO17 = precipitation of driest quarter; BIO4 = temperature seasonality; BIO12 = annual precipitation; BIO2 = mean diurnal range; BIO3 = isothermality; EVI 84 = enhanced vegetation index 1984; BIO15 = precipitation seasonality; PLANC = plan curvature; GC = general curvature; MINC = minimal curvature; CI = convergence index; PROFC = profile curvature; DAH = diurnal anisotropic heating; MAXC = maximal curvature.

3.4. Soil organic carbon stock in Rondônia

At a depth of one meter, the soils of Rondônia store approximately 2,530.91 Tg and more than 50% of this carbon is found in the superficial layers (0-30 cm). The largest stock of SOC were found in the southern and southeastern regions of the state, where there is a predominance of Gleissolos, low altitudes and seasonal flooding conditions (Fig. 6). The smallest SOC stock are in the central region of the state under Latossolos and where much of the native forest cover has been suppressed. The spatial pattern of distribution of SOC stock was similar between the evaluated layers and heterogeneous among the SOC stock in the same layer. It is possible to observe more abrupt spatial transitions in the stock of layers 0-5 and 5-15 cm, when compared to the other layers.

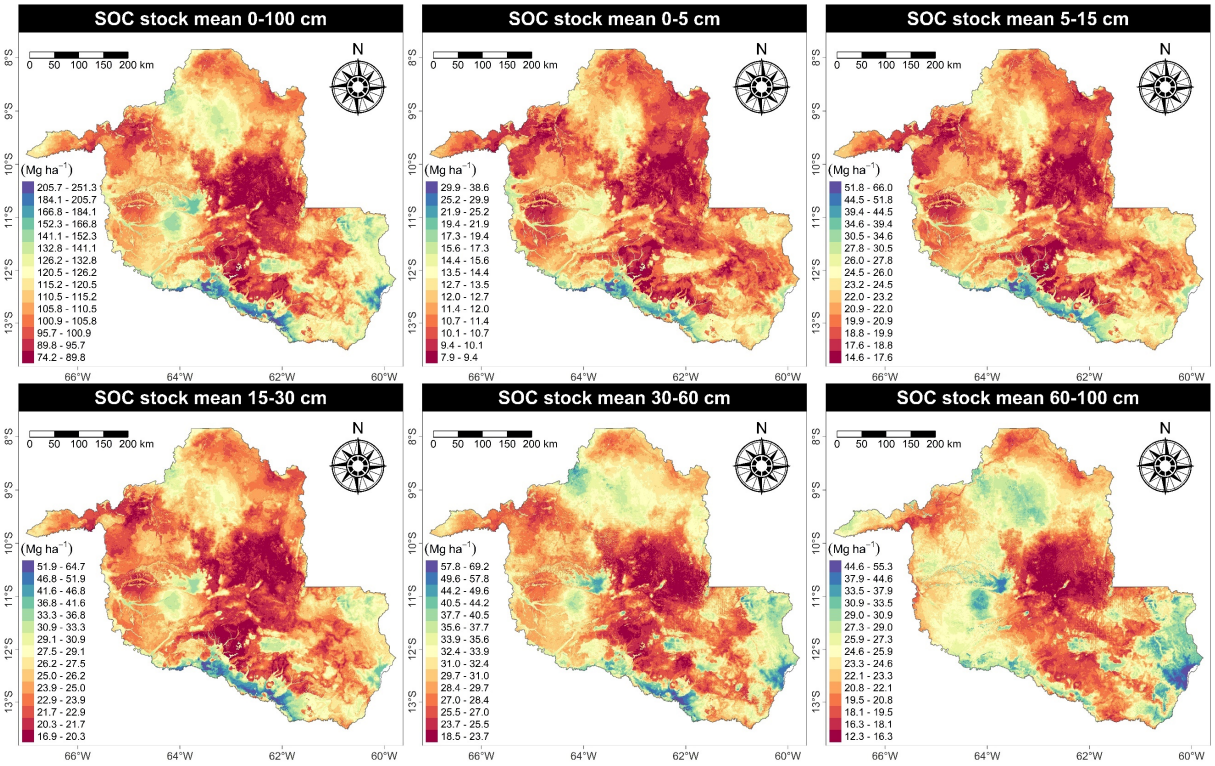

Fig. 6. Average of 100 soil organic carbon stock maps at different depths, Rondônia State, Brazil.

The values of the mean maps (Fig. 6) did not differ from the lower (Q05) (Fig. 7) and upper (Q95) (Fig. 8) intervals, showing a similar spatial pattern. The coefficient of variation maps (Fig. 9) corroborates this statement, as the percentages of pixel variation were less than 15% in a large part of the state. Higher prediction uncertainties were associated with higher SOC stock at all depths, especially in the southern region of the state. The variation was from 1.8% to 35.1% depending on the depth, with the layer of 30-60 cm < 15-30 cm < 60-100 cm < 5-15 cm < 0-5 cm.

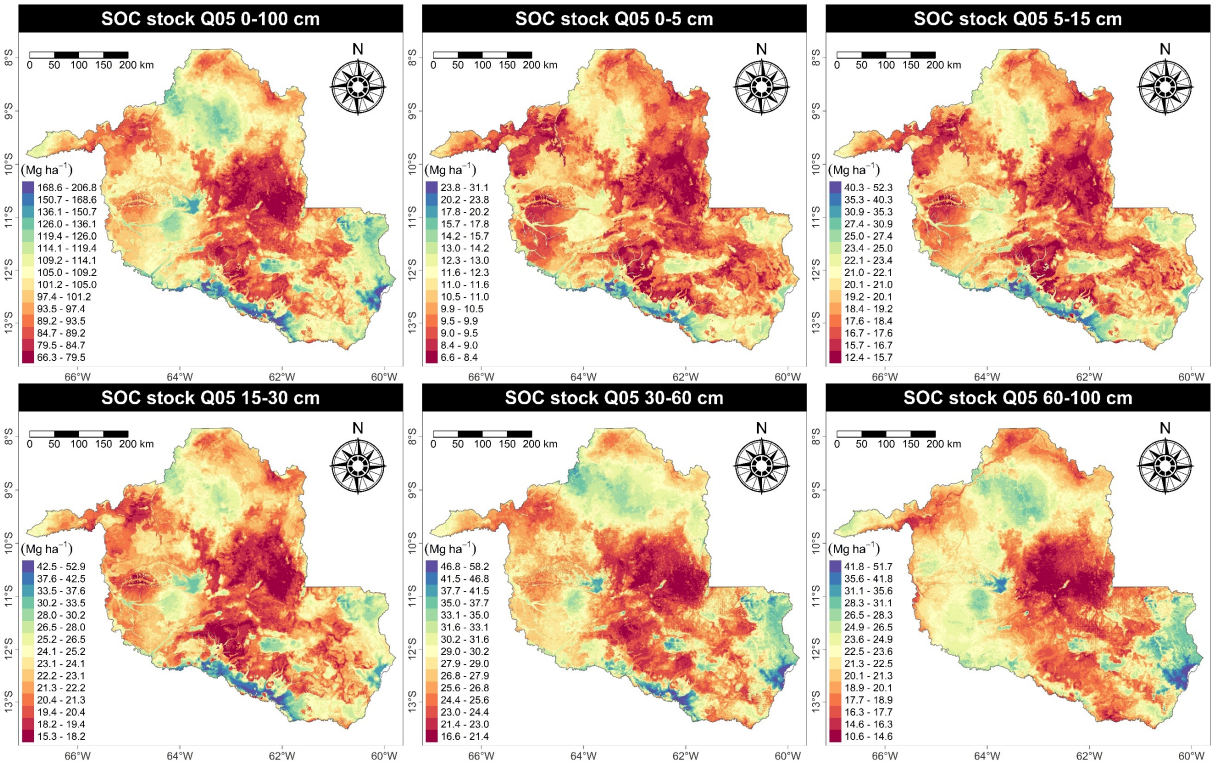

Fig. 7. Lower interval (Q05) of the 100 maps of soil organic carbon stock at different depths, Rondônia State, Brazil.

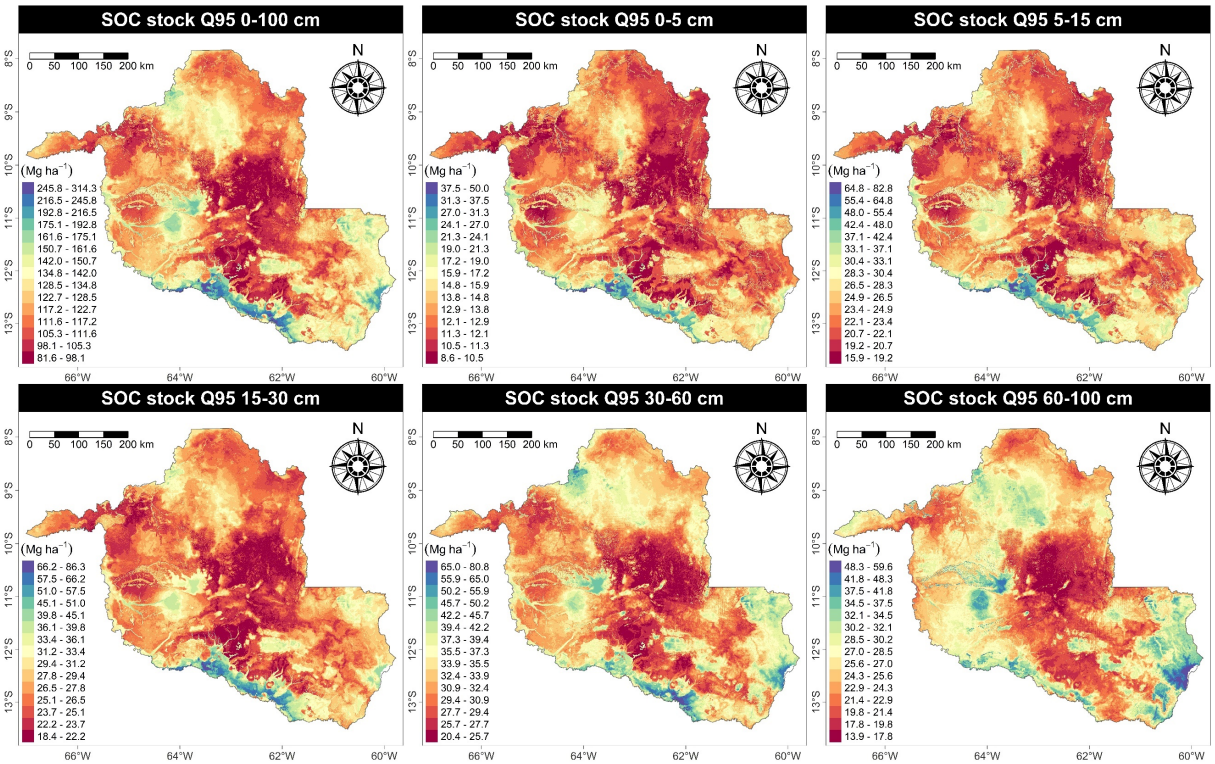

Fig. 8. Upper range (Q95) of the 100 maps of soil organic carbon stock at different depths, Rondônia State, Brazil.

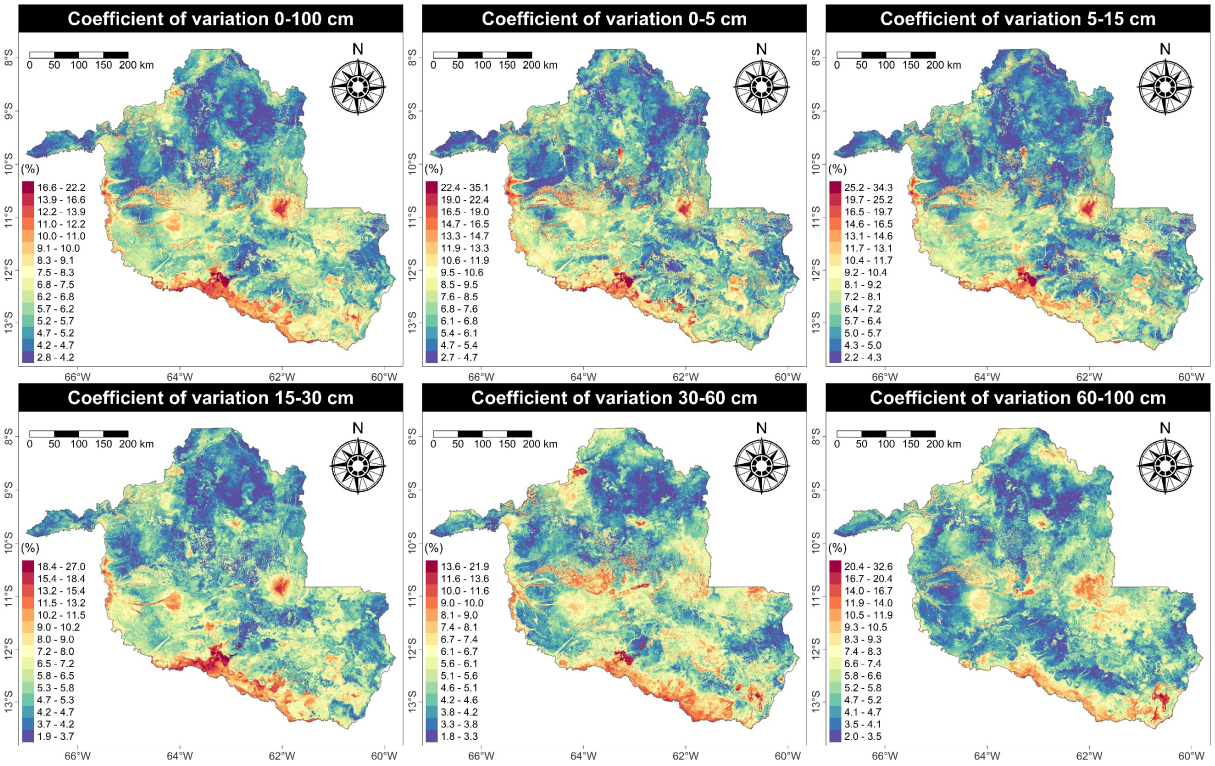

Fig. 9. Coefficient of variation of the 100 maps of soil organic carbon stock at different depths, Rondônia State, Brazil.

3.5. Stratification of SOC stock

At 1 m depth (Fig. 10), the largest total SOC stock were found in Latossolos (> 1,450 Tg C) and the smallest in Nitossolos (< 1 Tg C). Latossolos are the most representative soils and Nitossolos (0.04%) are the least represented in the area. Latossolos represent 57.8% and Nitosols 0.04% of SOC. The behavior in the stratified layers was similar to 1 meter, except in Gleissolos, with SOC stock always higher than the others, except for the 60-100 cm layer, with SOC stock smaller than in Neossolos. Regarding the density of SOC stock, Gleissolos and Organossolos had the largest stock (> 129 Mg ha^{-1}) at 1 m depth and the smallest in Cambissolos (< 108 Mg ha^{-1}). In the distribution between layers, the largest stock were maintained between Gleissolos and Organossolos, except for the 60-100 cm layer where the largest SOC stock are in Neossolos. The lowest SOC densities (ranging from 11.5 to 22 Mg ha^{-1}) were found among the Latossolos (0-5, 5-15 and 15-30 cm), Luvissolos (30-60 cm) and Nitossolos (60-100 cm). Total values varied greatly between soil types when compared to SOC densities. SOC densities varied more between layers than between soil types.

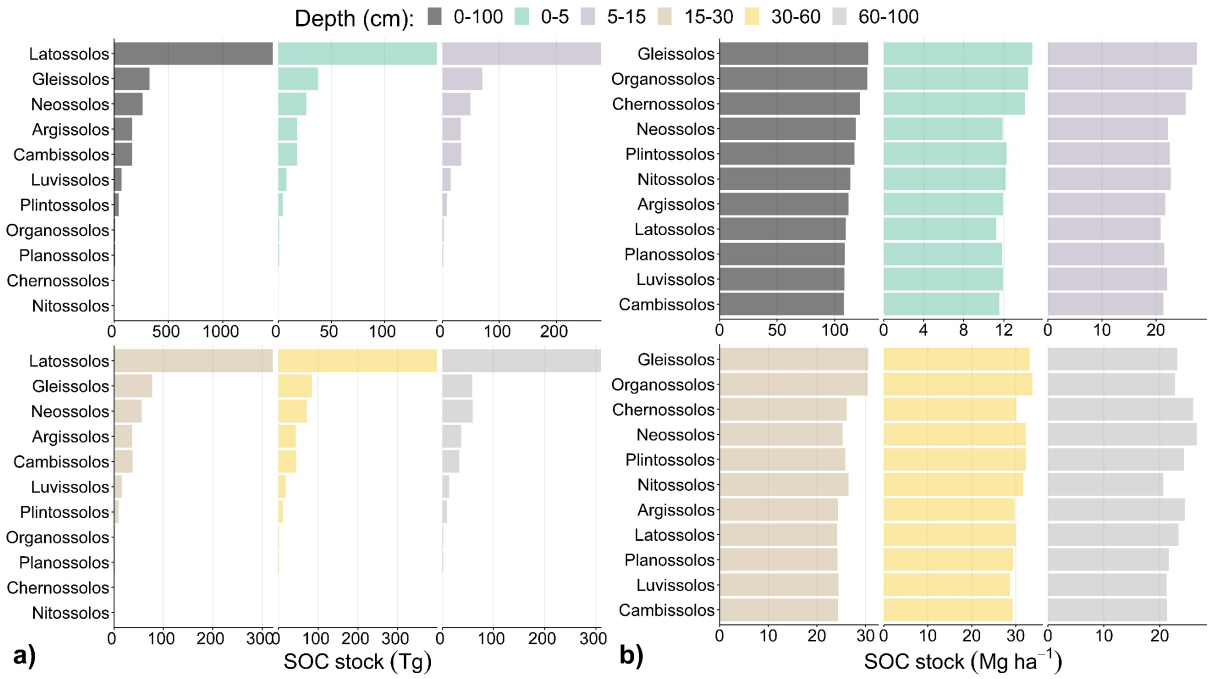

Fig. 10. Soil organic carbon (SOC) stock (a) and density of SOC stock (b) in soil classes at different depths, Rondônia State, Brazil.

Total SOC stock in the first 100 cm were highest in unprotected areas and indigenous lands (Fig. 11). The two together total 77% of the state's SOC stock, with unprotected areas holding more than 2.5 times the SOC stock of indigenous lands. SOC stock, from zero to 100 cm in depth, were similar between the stratified soil depths. Unlike total SOC stock, SOC densities (ranging from 10 to 29 Mg ha^{-1}) were higher in protected areas, and lower in private natural heritage reserves (PNHR) and unprotected areas, in all evaluated layers (Fig. 11). The highest SOC densities were in PNHR, environmental protection area (EPA), parks and biological reserves.

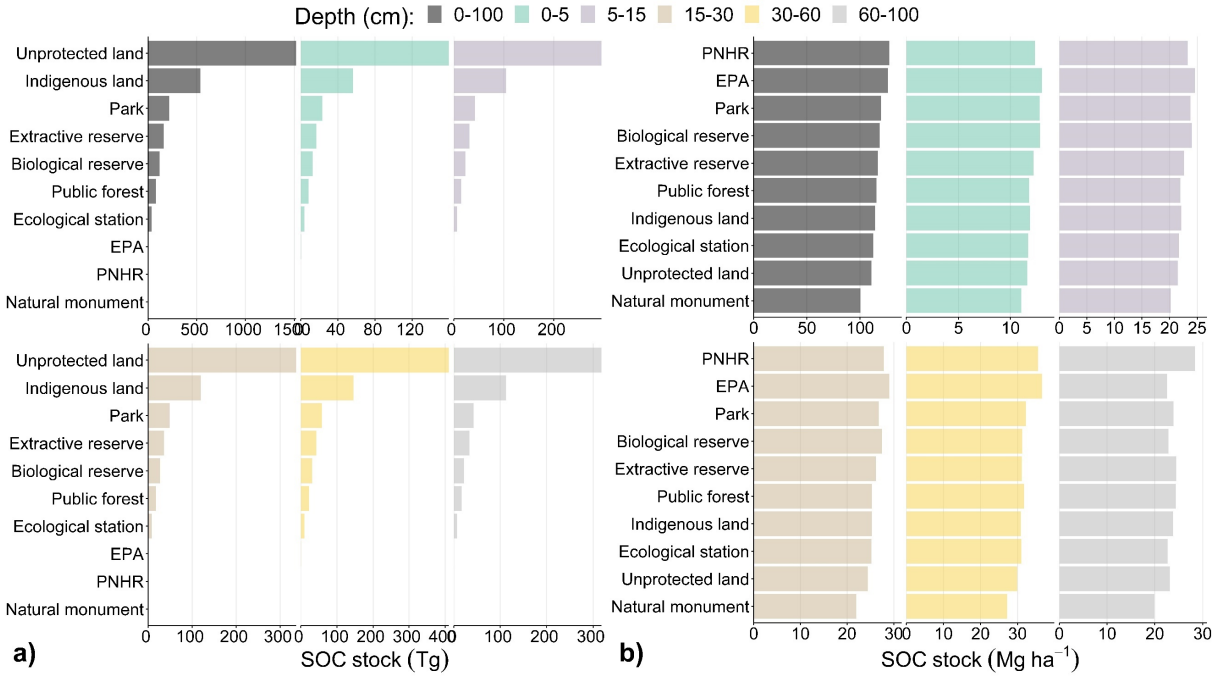

Fig. 11. Soil organic carbon (SOC) stock (a) and density of SOC stock (b) in different types of protected land (EPA = Environmental protection area; PNHR = Private natural heritage reserve) and unprotected land at different depths**,** Rondônia State, Brazil.

The dispersion of observed values and predicted values indicate that both the rf adjusted in this work and the models adjusted by other authors underestimate the SOC stock (Fig. 12 and Fig. 13a). However, the rf showed less dispersion compared to the others. Furthermore, in the distribution of values (Fig. 13), the rf was the closest to the observed values, especially in the mean. Gomes et al. (2019) fitted the model with 155 profiles (0.00069 profiles km^{-2}) distributed over the state of Rondônia, Vasques et al. (2017) with 7,002 profiles (0.00085 profiles km^{-2}) distributed throughout Brazil and Hengl et al. (2017) 368 profiles (0.00163 profiles km^{-2}) distributed throughout the state of Rondônia. The model was fitted in the research presented here with 2,557 profiles distributed in the territorial limits of Rondônia, therefore with a density greater than that carried out by the other authors (0.01225 profiles km^{-2}).

Other authors adjusted national (Gomes et al., 2019; Vasques et al., 2017) and international (Hengl et al., 2017) models, but with coarser resolutions, unlike the model adjusted in this research. Gomes et al. (2019), Hengl et al. (2017) and Vasques et al. (2017) fitted models with approximate areas of each pixel ranging from 57,838 m^2 to 945,005 m^2, the rf adjusted here has an area of 40,000 m^2 per pixel. Furthermore, the rf resulted in lower MAE (< 20 Mg C ha^{-1}) and RMSE (< 30 Mg C ha^{-1}) compared to the other models (Fig. 13b).

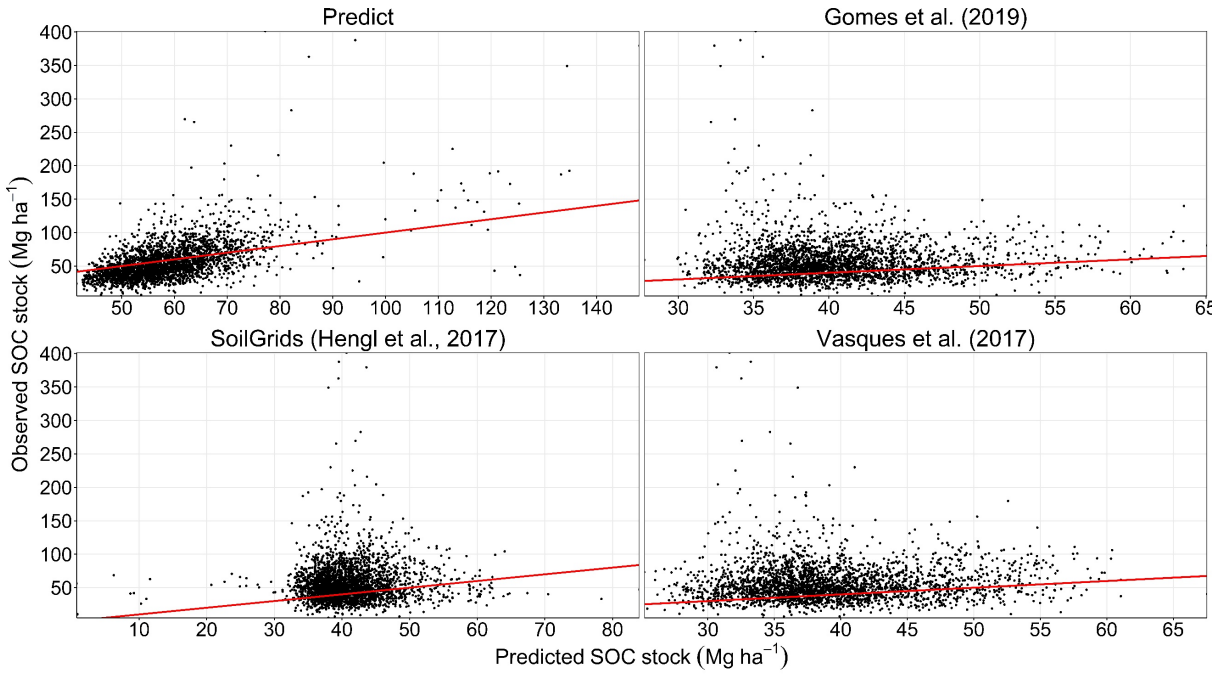

Fig. 12. Dispersion of soil organic carbon (SOC) stock values (0-30 cm) observed and estimated by rf in this study (predicted), and by Gomes et al. (2019), SoilGrids (Hengl et al., 2017) and (Vasques et al., 2017).

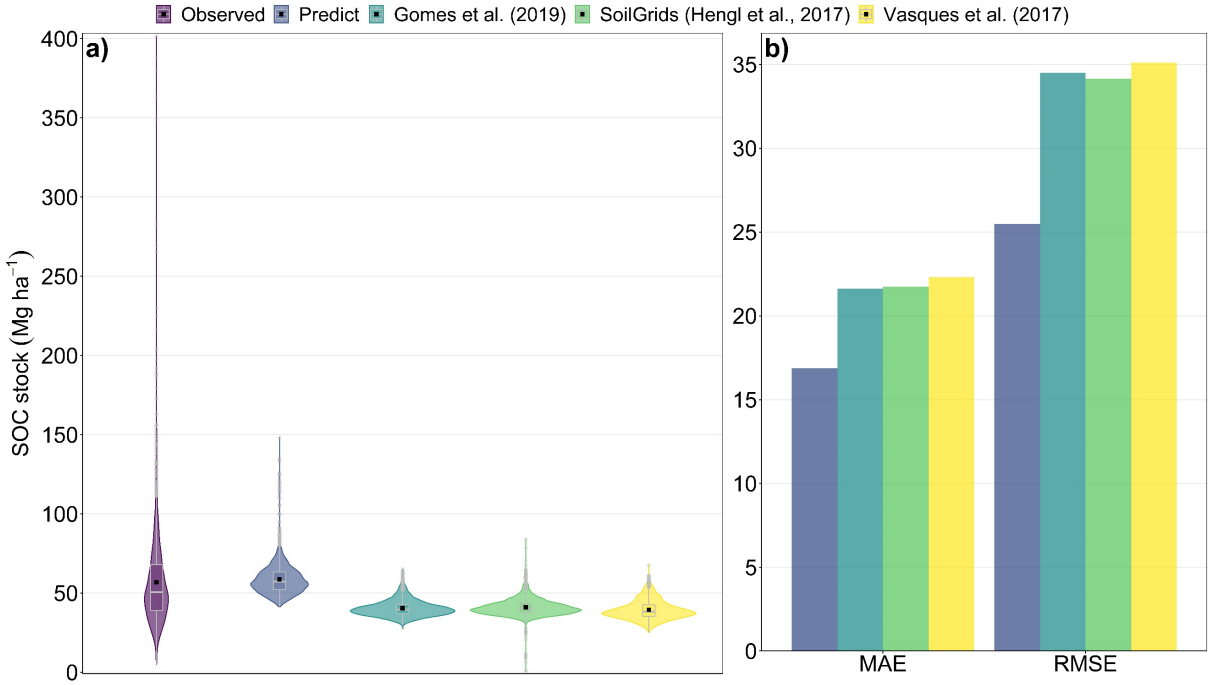

Fig. 13. Violinplot and boxplot (a) of the density of SOC stock (0-30 cm) of the observed, predicted (rf), Gomes et al. (2019), SoilGrids (Hengl et al., 2017) and Vasques et al. (2017) values and their respective mean absolute error - MAE and root mean square error - RMSE (b).

4. Discussion

4.1. Performance and model selection

The variation in performance metrics observed demonstrates the importance of repetition in the adjustment processes in machine learning, under different subsets, avoiding optimistic or pessimistic results that are not under with the reality of the predicted phenomenon. This is especially important in digital SOC mapping as SOC stock can be highly variable (Y. Li et al., 2021; Victor et al., 2021). Therefore, the methodological structure adopted is efficient for digital soil mapping from environmental predictors using machine learning.

Among the seven algorithms evaluated, the rf was the model with the best average performance, which, compared to the second best (cubist), resulted in 10% more performance (> R2; < MAE and RMSE). This corroborates other studies that use machine learning to predict soil attributes, which point to rf as the algorithm with the best results when contrasted with others (Nabiollahi et al., 2021; Silatsa et al., 2020; Zhang et al., 2021). The better performance given to rf may be due to its ability to detect non-linear and hierarchical relationships between SOC stock and environmental predictors (Wadoux et al., 2020). However, Wei et al. (2021) when studying soil salinity from spectral data in China, obtained divergent results, with responses from the performance of an algorithm from the boosting family higher than rf, indicating that different algorithms can better suit certain situations, emphasizing the importance in the selection of models, as done in the research presented here.

4.2. Selection and importance of predictors

Almost 50% of the total predictors of all layers (Table S5, Table S6, Table S7, Table S8 and Table S9, Supplementary material) were potentially redundant ($r > |0.95|$) and removed in the initial correlation removal process. This led to significant gains in processing time, both for the execution of the RFE and for the adjustment of the rf (Li et al., 2018). The RFE results show the importance of selecting predictors in the use of machine learning to predict SOC stock. The methodological structure used made it possible to reduce the predictors from 65 to an average range of 9 to 19 variables at each depth, added to the optimization of the hyperparameters and the 100 repetitions of execution required less time, power, and computational effort, which in contrast to this, if not executed, it could make the analysis unfeasible, given the size of the studied area. Although the application of RFE does not necessarily provide performance gains, there was a quantitative reduction in predictors without significant loss of performance, this behavior corroborates other authors who used RFE in digital soil mapping via machine learning (Gomes et al., 2019; Meier et al., 2018; Mosavi et al., 2020).

Soil class was one of the predictors that most influenced SOC stock at all depths, especially in the superficial layers (0-5, 5-15 and 15-30 cm). This relationship is mainly described by texture and structure, as these properties are decisive in the balance between carbon output and input, as well as the persistence of carbon in the soil (Swetha and Chakraborty, 2021; Tan et al., 2004). Well-structured and fine-textured soils tend to store more carbon due to the organo-mineral interaction between soil particles with a larger specific area and organic particles deposited under the surface (Basile-Doelsch et al., 2020). Furthermore, soil genesis involves complex processes that go beyond physicochemical characteristics, reflecting the integrative importance of soil classes in the variability of SOC stock (Soucémarianadin et al., 2018).

Another important factor that controlled SOC stock at all depths was the bioclimatic predictors, especially in the subsurface layers (30-60, 60-100 cm), with the soil classes standing out in importance. The influence of climatic factors on SOC stock is well documented in the literature, either at local (X. Li et al., 2021) and regional (Duan et al., 2020) or national scales (D. Chen et al., 2020) and global (S. Chen et al., 2020).

Temperature mainly moderates plant carbon deposition under the soil and decomposition promoted by soil biota. The increase in temperature increases the decomposition of the carbon consumed by the biota, however, this increase in temperature also provides higher productivity for plants, which in turn deposit more C, balancing carbon inputs and outputs. However, this balance is complex, due to the dynamics of ecosystems and the covariations of other factors, which can, for example, given the variation in precipitation, decompose C at more or less accelerated rates compared to biomass replacement by plants. This explains places with low SOC stock even under high temperatures (Tan et al., 2020).

This close relationship between SOC and climate explains the high importance found between predictors BIO4, BIO17 and BIO18 and SOC stock. After all, BIO4 describes the seasonality of temperature, BIO17 with precipitation from the driest quarter and BIO18 with precipitation from the warmest quarter reproduce the interactions between precipitation and temperature.

Of the predictors representing the vegetation, only the EVI was selected for the model adjustment and the others were discarded in the predictor selection process. The EVI is a recommended vegetation index to describe areas with a dense forest as it is more responsive to structural changes in the canopy, including plant physiognomy, canopy type and architecture (Huete et al., 2002). This may explain its permanence compared to the others, after the selection process, as Rondônia, in its majority, is covered by dense Amazon Forest. EVI influenced SOC

stock only in the 0-5 cm layer, evidencing the importance of vegetation for surface SOC stock. The soil surface is more sensitive to changes, generally containing less perennial carbon and dependent on the dynamics of vegetation input, which influences the amount of litter and, consequently, the surface SOC stock (Hu et al., 2020), elucidating the relevance of EVI in the construction of the model for predicting superficial SOC stock.

4.3. Soil organic carbon stock in Rondônia

Estimates that found more than half of the 2,530.91 Tg of SOC stored at a depth of 100 cm are found in the 0-30 cm layer is corroborated by other studies that performed digital mapping of the SOC. For instance, Gomes et al. (2019), using machine learning to estimate SOC stock in Brazil up to 100 cm deep, found 50.34% of the stock in the surface layer (0-30 cm) and Silatsa et al. (2020) using the same technique in Cameroon found 50% of the 100 cm deep SOC stock on the surface.

Except for studies on a national and global scale, which estimate SOC stock at coarser scales (Batjes, 2006; Gomes et al., 2019), no previous research was found that estimated SOC stock for Rondônia at scale detailed as performed in this study. Due to the global importance of the Amazon Biome in climate regulation (Area Leão Pereira et al., 2020), approaches that provide information on soil carbon, as presented here, are fundamental for sustainable land management and climate change mitigation. The information presented demonstrates that the loss of forest cover can intensify the problem of climate change (Fearnside, 2006), both through the emission of carbon stored in the soil and through the emission of carbon from forest biomass.

In addition to vegetation cover, climate also influences SOC stock. The largest SOC stock found in the southern region of the state for all depths can be explained by local edaphoclimatic characteristics, low altitudes, and seasonal flooding dynamics. This, together with the permanence of adapted native vegetation and soils with high clay contents, make it difficult to convert these areas for intensive use and condition the accumulation of SOC (Hounkpatin et al., 2018; Magha et al., 2021). However, the data must be interpreted with caution due to the high coefficients of variation in the SOC modeling of these areas, which indicates greater uncertainties in these predictions.

The lower SOC contents found in the central region of the state, especially on the surface, may be associated both with the high weathering of the Latossolos, as well as with the use and land cover. Latossolos are soils in the advanced weathering stage, with an absence of less resistant primary and secondary minerals, commonly deep, alic and dystrophic (Santos et

al., 2018). This results in soils with SOC stock dependent on aboveground vegetation and input dynamics. This region had profound changes in the landscape, resulting from the colonization process (since the 1960s) and the expansion of extensive cattle raising, facilitated by the construction of the main access road to the state. As a result, the land conversion process has given way to degraded pastures or at risk of degradation, which can considerably reduce SOC stock (Maia et al., 2009; Numata et al., 2007).

4.4. Stratification of SOC stock

Understanding the relationship between soil class and the distribution of total stock and SOC density is important. Latossolos, with almost 60% of the total stock, are predominant not only in the state of Rondônia, but also in the entire Brazilian territory (IBGE and Embrapa, 2011). The geographic predominance of Latossolos, added to its position in the relief and desirable physical characteristics (permeability and structure), make it one of the most used soil classes in agriculture, even intensively (Ker, 1997). These factors can increase pressure on natural areas and increase erosion in agricultural areas, which consequently reduces the SOC stock of these soils.

The higher density of SOC stock in Gleissolos and Organossolos is related to the genesis of these soils. Gleissolos are inserted in hydromorphic conditions which directly impact the greater accumulation and stability of SOC (Magha et al., 2021). Soil under water saturation conditions is subjected to reducing processes, where microbial activity is reduced by the limitation of free O_2, which hinders the decomposition of organic matter and favors the accumulation of SOC (Amendola et al., 2018). The genesis of Organossolos also contributes to the accumulation of soil organic matter and explains the high density of SOC found here. These soils, unlike the other classes, suffer little influence from mineral materials derived from rocks, consisting mainly of organic materials. Arising from litter accumulation added to humification processes, delayed decomposition by edaphoclimatic conditions and/or paludization (accumulation of organic matter, in a predominantly anaerobic environment) due to relief conditions and floristic composition (Buol et al., 2011; Pereira et al., 2005).

The largest SOC stock in Neossolos in the 60-100 cm layer occurred due to the topographic and climatic conditions where these soils occur, as they are inserted at higher altitudes (> 350 m) and with a milder weather pattern, which can reduce the weathering intensity and SOC evolution in the form of CO_2 (Ma and Chang, 2019).

Higher stock of SOC in unprotected areas compared to protected areas were already expected. This is because unprotected areas are predominant in territorial extension in the state

(58.2%). Indigenous lands contain the second largest stock of SOC (>500 Tg C) in the state. This is due to a sum of factors, firstly, the greater territorial extension, compared to the other classes of protected areas. Furthermore, these populations have managed the soil since the pre-Columbian era and use forest mechanisms to their advantage. This allows these populations to lead productive systems that increase the SOC. One of the most conclusive evidences of the influence of these peoples on the environment, and especially on the soil, is the "terra preta de índio", these soils are widely studied anthropogenic soils (Bezerra, 2015; Glaser et al., 2000; Pagano et al., 2016; Pandey et al., 2020). "Terra preta de índio" is known for its high chemical fertility and high SOC contents, much of its composition is related to carbon-based materials (Oliveira et al., 2020). Part of these soils is found within some indigenous lands of Rondônia (Teixeira et al., 2009), which, in addition to the territorial extension, can justify the high amounts of SOC in these areas.

The higher stock densities of PNHR in the 1 m layer are due to their places of occurrence in the state. They are located mainly under Neossolos with eutrophic characteristics and, as mentioned above, estimates point to high SOC stock of 60-100 cm in Neossolos of Rondônia. The high density of SOC in protected areas to the detriment of non-protected areas reaffirms the importance of protecting these areas and other areas composed of forests, even if not protected. The forest is important for biodiversity and other ecosystem services (Garland et al., 2021), and stored SOC stock, as evidenced by our data.

Not all unprotected areas in Rondônia are composed of agriculture or livestock, part of these areas is still composed of natural forest and is important to maintain the high amount of SOC. Especially when considering climate change scenarios, conserving these forests, and increasing them, recovering degraded areas, and encouraging sustainable production systems, such as agroforestry systems, for example, contribute to increasing soil carbon and mitigating the effects of climate change.

However, actions have not been aimed at mitigating such effects. The data used for modeling are from 1996/1997. From that period to the current period, the state had much of its forest cover removed and converted to pasture, which may indicate that the data presented here are overestimated.

The dispersion of observed and predicted points highlights the importance of modeling SOC stock at lower territorial levels to achieve greater precision. Not only the size of the area, but also the density of samples may have impacted the estimates, giving greater precision to the rf compared to the others (Ng et al., 2020). Somarathna et al. (2017) for example, tested different sampling intensities and types of models for mapping the SOC stock in Lower Hunter

Valley, NSW, Australia and concluded that the sample density more significantly impacted the performance of the models than the different types of models, which increased the generalization capacity as the sample density increased.

5. Conclusion

The use of machine learning algorithms was efficient in mapping SOC stock in Rondônia. Among the evaluated algorithms, random forest offered better performance. The methodological framework reduced the computational cost and allowed the quantification of uncertainty in predictions.

Comparison of the model fitted in the work and other models with the observed data contributed to an additional validation of the model and the SOC stock maps generated for Rondônia. The greater density of profiles used to model the SOC conferred greater precision on the model compared to existing estimates.

Soil classes, temperature and precipitation were the main drivers of SOC stock distribution in Rondônia. EVI was an important controller of SOC stock in the topsoil. The SOC stock in the state of Rondônia at 1 m depth is approximately 2,530.91 Tg.

The combined Latossolos and Gleissolos contain 70.80% of the state's total SOC stock. Protected areas store the highest density of SOC in the state, however, the largest total stock are found in unprotected areas, due to their area of occurrence. Therefore, fostering more sustainable production systems that mimic protected areas is important to conserve and/or increase SOC stock and mitigate climate change.

References

Akpa, S.I.C., Odeh, I.O.A., Bishop, T.F.A., Hartemink, A.E., Amapu, I.Y., 2016. Total soil organic carbon and carbon sequestration potential in Nigeria. Geoderma 271, 202–215. https://doi.org/10.1016/j.geoderma.2016.02.021

Alvares, C.A., Stape, J.L., Sentelhas, P.C., Moraes Gonçalves, J.L., Sparovek, G., 2013. Köppen's climate classification map for Brazil. Meteorologische Zeitschrift 22, 711–728. https://doi.org/10.1127/0941-2948/2013/0507

Amendola, D., Mutema, M., Rosolen, V., Chaplot, V., 2018. Soil hydromorphy and soil carbon: A global data analysis. Geoderma 324, 9–17. https://doi.org/10.1016/j.geoderma.2018.03.005

Amiri, M., Tarkesh, M., Jafari, R., Jetschke, G., 2020. Bioclimatic variables from precipitation and temperature records vs. remote sensing-based bioclimatic variables: Which side can

perform better in species distribution modeling? Ecological Informatics 57, 101060. https://doi.org/10.1016/j.ecoinf.2020.101060

Area Leão Pereira, E.J., Santana Ribeiro, L.C., Silva Freitas, L.F., Barros Pereira, H.B., 2020. Brazilian policy and agribusiness damage the Amazon rainforest. Land Use Policy 92, 104491. https://doi.org/10.1016/j.landusepol.2020.104491

Arrouays, D., Grundy, M.G., Hartemink, A.E., Hempel, J.W., Heuvelink, G.B.M., Hong, S.Y., Lagacherie, P., Lelyk, G., McBratney, A.B., McKenzie, N.J., Mendonca-Santos, M. d. L., Minasny, B., Montanarella, L., Odeh, I.O.A., Sanchez, P.A., Thompson, J.A., Zhang, G.L., 2014. GlobalSoilMap. Toward a Fine-Resolution Global Grid of Soil Properties, in: Advances in Agronomy. Academic Press Inc., pp. 93–134. https://doi.org/10.1016/B978-0-12-800137-0.00003-0

Barros Ferraz, S.F., Vettorazzi, C.A., Theobald, D.M., Ballester, M.V.R., 2005. Landscape dynamics of Amazonian deforestation between 1984 and 2002 in central Rondônia, Brazil: Assessment and future scenarios. Forest Ecology and Management 204, 69–85. https://doi.org/10.1016/j.foreco.2004.07.073

Basile-Doelsch, I., Balesdent, J., Pellerin, S., 2020. Reviews and syntheses: The mechanisms underlying carbon storage in soil. Biogeosciences 17, 5223–5242. https://doi.org/10.5194/bg-17-5223-2020

Batjes, N.H., 2006. Organic carbon stocks in the soils of Brazil. Soil Use and Management 21, 22–24. https://doi.org/10.1111/j.1475-2743.2005.tb00102.x

Batjes, N.H., Ribeiro, E., van Oostrum, A.J.M., 2019. Standardised soil profile data to support global mapping and modelling (WoSIS snapshot 2019). Earth System Science Data Discussions 1–46. https://doi.org/10.5194/essd-2019-164

Behrens, T., Schmidt, K., Viscarra Rossel, R.A., Gries, P., Scholten, T., MacMillan, R.A., 2018. Spatial modelling with Euclidean distance fields and machine learning. European Journal of Soil Science 69, 757–770. https://doi.org/10.1111/ejss.12687

Bezerra, J., 2015. Terra Preta de Índio and Amazonian History. Springer, Cham, pp. 15–58. https://doi.org/10.1007/978-3-319-23030-6_2

Bivand, R., 2021. rgrass7: Interface Between GRASS 7 Geographical Information System and R.

Breiman, L., 2001. Random forests, Machine Learning. https://doi.org/10.1023/A:1010933404324

Brenning, A., Bangs, D., Becker, M., 2018. RSAGA: SAGA Geoprocessing and Terrain Analysis.

Bullock, E.L., Woodcock, C.E., Olofsson, P., 2020. Monitoring tropical forest degradation using spectral unmixing and Landsat time series analysis. Remote Sensing of Environment 238, 110968. https://doi.org/10.1016/j.rse.2018.11.011

Buol, S.W., Southard, R.J., Graham, R.C., McDaniel, P.A., 2011. Soil Genesis and Classification: Sixth Edition, Soil Genesis and Classification: Sixth Edition. Wiley-Blackwell, Oxford, UK. https://doi.org/10.1002/9780470960622

Chen, D., Wei, W., Daryanto, S., Tarolli, P., 2020. Does terracing enhance soil organic carbon sequestration? A national-scale data analysis in China. Science of the Total Environment 721, 137751. https://doi.org/10.1016/j.scitotenv.2020.137751

Chen, S., Richer-de-Forges, A.C., Saby, N.P.A., Martin, M.P., Walter, C., Arrouays, D., 2018. Building a pedotransfer function for soil bulk density on regional dataset and testing its validity over a larger area. Geoderma 312, 52–63. https://doi.org/10.1016/j.geoderma.2017.10.009

Chen, S., Zou, J., Hu, Z., Lu, Y., 2020. Temporal and spatial variations in the mean residence time of soil organic carbon and their relationship with climatic, soil and vegetation drivers. Global and Planetary Change 195, 103359. https://doi.org/10.1016/j.gloplacha.2020.103359

Cochrane, T.T., Cochrane, T.A., 2006. Diversity of the land resources in the Amazonian state of Rondônia, Brazil. Acta Amazonica 36, 91–102. https://doi.org/10.1590/s0044-59672006000100011

Cole, B., McMorrow, J., Evans, M., 2014. Empirical Modelling of Vegetation Abundance from Airborne Hyperspectral Data for Upland Peatland Restoration Monitoring. Remote Sensing 6, 716–739. https://doi.org/10.3390/rs6010716

Conoscenti, C., Ciaccio, M., Caraballo-Arias, N.A., Gómez-Gutiérrez, Á., Rotigliano, E., Agnesi, V., 2015. Assessment of susceptibility to earth-flow landslide using logistic regression and multivariate adaptive regression splines: A case of the Belice River basin (western Sicily, Italy). Geomorphology 242, 49–64. https://doi.org/10.1016/j.geomorph.2014.09.020

Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine Learning 20, 273–297. https://doi.org/10.1007/BF00994018

Duan, L., Li, Zhenwei, Xie, H., Yuan, H., Li, Zhiming, Zhou, Q., 2020. Regional pattern of soil organic carbon density and its influence upon the plough layers of cropland. Land Degradation and Development 31, 2461–2474. https://doi.org/10.1002/ldr.3610

Eiten, G., 1972. The cerrado vegetation of Brazil. The Botanical Review 38, 201–341. https://doi.org/10.1007/BF02859158

Fearnside, P.M., 2006. Deforestation in Amazonia: Dynamics, impacts and control. Acta Amazonica 36, 395–400. https://doi.org/10.1590/s0044-59672006000300018

Fernandes Filho, E.I., 2020. labgeo: Collection of functions to fit models with emphasis in land use and soil mapping.

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology 37, 4302–4315. https://doi.org/10.1002/joc.5086

Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software 33, 1–22. https://doi.org/10.18637/jss.v033.i01

Friedman, J.H., 2002. Stochastic gradient boosting. Computational Statistics and Data Analysis 38, 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2

Friedman, J.H., 1991. Multivariate Adaptive Regression Splines. The Annals of Statistics 19, 1–67. https://doi.org/10.1214/aos/1176347963

Garland, G., Banerjee, S., Edlinger, A., Miranda Oliveira, E., Herzog, C., Wittwer, R., Philippot, L., Maestre, F.T., van der Heijden, M.G.A., 2021. A closer look at the functions behind ecosystem multifunctionality: A review. Journal of Ecology 109, 600–613. https://doi.org/10.1111/1365-2745.13511

Glaser, B., Balashov, E., Haumaier, L., Guggenberger, G., Zech, W., 2000. Black carbon in density fractions of anthropogenic soils of the Brazilian Amazon region. Organic Geochemistry 31, 669–678. https://doi.org/10.1016/S0146-6380(00)00044-9

Goh, A.T.C., Zhang, W., Zhang, Y., Xiao, Y., Xiang, Y., 2018. Determination of earth pressure balance tunnel-related maximum surface settlement: a multivariate adaptive regression splines approach. Bulletin of Engineering Geology and the Environment 77, 489–500. https://doi.org/10.1007/s10064-016-0937-8

Gomes, L.C., Faria, R.M., Souza, E., Veloso, G.V., Schaefer, C.E.G.R., Fernandes Filho, E.I., 2019. Modelling and mapping soil organic carbon stocks in Brazil. Geoderma 340, 337–350. https://doi.org/10.1016/j.geoderma.2019.01.007

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Greenwell, B., Boehmke, B., Cunningham, J., GBM Developers, 2020. gbm: Generalized Boosted Regression Models.

Hechenbichler, K., Schliep, K., 2004. Weighted k-Nearest-Neighbor Techniques and Ordinal Classification, Molecular Ecology.

Hengl, T., 2020. GSIF: Global Soil Information Facilities.

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., 2017. SoilGrids250m: Global gridded soil information based on machine learning. PLoS ONE 12, e0169748. https://doi.org/10.1371/journal.pone.0169748

Hijmans, R.J., 2021. raster: Geographic Data Analysis and Modeling.

Hounkpatin, O.K.L., Op Hipt, F., Bossa, A.Y., Welp, G., Amelung, W., 2018. Soil organic carbon stocks and their determining factors in the Dano catchment (Southwest Burkina Faso). CATENA 166, 298–309. https://doi.org/10.1016/j.catena.2018.04.013

Hu, P., Xiao, J., Zhang, W., Xiao, L., Yang, R., Xiao, D., Zhao, J., Wang, K., 2020. Response of soil microbial communities to natural and managed vegetation restoration in a subtropical karst region. Catena 195, 104849. https://doi.org/10.1016/j.catena.2020.104849

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment 83, 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2

IBGE, 2012. Anuário estatístico do Brasil. Wasmália Bivar 72, 56.

IBGE, Embrapa, 2011. Soil map of Brazil, 1:5,000,000. Rio de Janeiro.

Janitza, S., Hornung, R., 2018. On the overestimation of random forest's out-of-bag error. PLoS ONE 13, e0201904. https://doi.org/10.1371/journal.pone.0201904

Jansson, J.K., Hofmockel, K.S., 2020. Soil microbiomes and climate change. Nature Reviews Microbiology 18, 35–46. https://doi.org/10.1038/s41579-019-0265-7

Karatzoglou, A., Hornik, K., Smola, A., Zeileis, A., 2004. kernlab - An S4 package for kernel methods in R. Journal of Statistical Software 11, 1–20. https://doi.org/10.18637/jss.v011.i09

Ker, J.C., 1997. Latossolos do Brasil: Uma Revisão. Geonomos. https://doi.org/10.18285/geonomos.v5i1.187

Khaledian, Y., Miller, B.A., 2020. Selecting appropriate machine learning methods for digital soil mapping. Applied Mathematical Modelling 81, 401–418. https://doi.org/10.1016/j.apm.2019.12.016

Kodikara, G.R.L., McHenry, L.J., 2020. Machine learning approaches for classifying lunar soils. Icarus 345, 113719. https://doi.org/10.1016/j.icarus.2020.113719

Kuhn, M., 2021. caret: Classification and Regression Training. https://doi.org/10.1887/0750303123/b365c43

Kuhn, M., Johnson, K., 2013. Applied predictive modeling, Applied Predictive Modeling. Springer New York, New York, NY. https://doi.org/10.1007/978-1-4614-6849-3

Kuhn, M., Quinlan, R., 2021. Cubist: Rule - And Instance-Based Regression Modeling.

Lal, R., 2004. Soil carbon sequestration impacts on global climate change and food security. Science 304, 1623–1627. https://doi.org/10.1126/science.1097396

Li, Hongzhi, Li, W., Pan, X., Huang, J., Gao, T., Hu, L.H., Li, Hui, Lu, Y., 2018. Correlation and redundancy on machine learning performance for chemical databases. Journal of Chemometrics 32, e3023. https://doi.org/10.1002/cem.3023

Li, X., Ding, J., Liu, J., Ge, X., Zhang, J., 2021. Digital mapping of soil organic carbon using sentinel series data: A case study of the ebinur lake watershed in xinjiang. Remote Sensing 13, 1–19. https://doi.org/10.3390/rs13040769

Li, Y., Duan, X., Li, Ya, Li, Yuxiang, Zhang, L., 2021. Interactive effects of land use and soil erosion on soil organic carbon in the dry-hot valley region of southern China. Catena 201, 105187. https://doi.org/10.1016/j.catena.2021.105187

Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest, R News.

Ma, M., Chang, R., 2019. Temperature drive the altitudinal change in soil carbon and nitrogen of montane forests: Implication for global warming. Catena 182, 104126. https://doi.org/10.1016/j.catena.2019.104126

Magha, A.M., Azinwi Tamfuh, P., Mamdem, L.E., Shey Yefon, M.C., Kenzong, B., Bitom, D., 2021. Soil Water Characteristics of Gleysols in the Bamenda (Cameroon) Wetlands and Implications for Agricultural Management Strategies. Applied and Environmental Soil Science 2021, 1–15. https://doi.org/10.1155/2021/6643208

Maia, S.M.F., Ogle, S.M., Cerri, C.E.P., Cerri, C.C., 2009. Effect of grassland management on soil carbon sequestration in Rondônia and Mato Grosso states, Brazil. Geoderma 149, 84–91. https://doi.org/10.1016/j.geoderma.2008.11.023

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4

Meier, M., Souza, E., Francelino, M.R., Fernandes Filho, E.I., Schaefer, C.E.G.R., 2018. Digital soil mapping using machine learning algorithms in a tropical mountainous area. Revista Brasileira de Ciencia do Solo 42, 1–22. https://doi.org/10.1590/18069657rbcs20170421

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., 2021. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien.

Milborrow, S., 2020. earth: Multivariate Adaptive Regression Splines.

Minasny, B., Malone, B.P., McBratney, A.B., Angers, D.A., Arrouays, D., Chambers, A., Chaplot, V., Chen, Z.S., Cheng, K., Das, B.S., Field, D.J., Gimona, A., Hedley, C.B., Hong, S.Y., Mandal, B., Marchant, B.P., Martin, M., McConkey, B.G., Mulder, V.L., O'Rourke, S., Richer-de-Forges, A.C., Odeh, I., Padarian, J., Paustian, K., Pan, G., Poggio, L., Savin, I., Stolbovoy, V., Stockmann, U., Sulaeman, Y., Tsui, C.C., Vågen, T.G., van Wesemael, B., Winowiecki, L., 2017. Soil carbon 4 per mille. Geoderma 292, 59–86. https://doi.org/10.1016/j.geoderma.2017.01.002

Mosavi, A., Hosseini, F.S., Choubin, B., Goodarzi, M., Dineva, A.A., 2020. Groundwater Salinity Susceptibility Mapping Using Classifier Ensemble and Bayesian Machine Learning Models. IEEE Access 8, 145564–145576. https://doi.org/10.1109/ACCESS.2020.3014908

Nabiollahi, K., Taghizadeh-Mehrjardi, R., Shahabi, A., Heung, B., Amirian-Chakan, A., Davari, M., Scholten, T., 2021. Assessing agricultural salt-affected land using digital soil mapping and hybridized random forests. Geoderma 385, 114858. https://doi.org/10.1016/j.geoderma.2020.114858

NASA JPL, 2020. NASADEM Merged DEM Global 1 arc second V001. NASA EOSDIS Land Processes DAAC. https://doi.org/https://doi.org/10.5067/MEaSUREs/NASADEM/NASADEM_HGT.001

Nelson, D.W., Sommers, L.E., 1982. Methods of Soil Analysis. Part 2. Chemical and Microbiological Properties Methods of Soil Analysis. Part 2. Chemical and Microbiological Properties., Madison, WI: American Society of Agronomy, Soil Science Society of America. Madison.

Ng, W., Minasny, B., de Sousa Mendes, W., Melo Demattê, J.A., 2020. The influence of training sample size on the accuracy of deep learning models for the prediction of soil properties with near-infrared spectroscopy data. Soil 6, 565–578. https://doi.org/10.5194/soil-6-565-2020

Nguyen, H., Bui, X.N., Tran, Q.H., Mai, N.L., 2019. A new soft computing model for estimating and controlling blast-produced ground vibration based on Hierarchical K-means clustering and Cubist algorithms. Applied Soft Computing Journal 77, 376–386. https://doi.org/10.1016/j.asoc.2019.01.042

Numata, I., Chadwick, O.A., Roberts, D.A., Schimel, J.P., Sampaio, F.F., Leonidas, F.C., Soares, J. v., 2007. Temporal nutrient variation in soil and vegetation of post-forest pastures as a function of soil order, pasture age, and management, Rondônia, Brazil. Agriculture, Ecosystems and Environment 118, 159–172. https://doi.org/10.1016/j.agee.2006.05.019

O'Brien, J., 2021. gdalUtilities: Wrappers for "GDAL" Utilities Executables.

Oliveira, E.A., Marimon-Junior, B.H., Marimon, B.S., Iriarte, J., Morandi, P.S., Maezumi, S.Y., Nogueira, D.S., Aragão, L.E.O.C., Silva, I.B., Feldpausch, T.R., 2020. Legacy of Amazonian Dark Earth soils on forest structure and species composition. Global Ecology and Biogeography 29, 1458–1473. https://doi.org/10.1111/geb.13116

Pagano, M.C., Ribeiro-Soares, J., Cançado, L.G., Falcão, N.P.S., Gonçalves, V.N., Rosa, L.H., Takahashi, J.A., Achete, C.A., Jorio, A., 2016. Depth dependence of black carbon structure, elemental and microbiological composition in anthropic Amazonian dark soil. Soil and Tillage Research 155, 298–307. https://doi.org/10.1016/j.still.2015.09.001

Pandey, S.D., Rocha, L.C., Pereira, G., Deschamps, C., Campos, J.L.E., Falcão, N., Prous, A., Jorio, A., 2020. Properties of carbon particles in archeological and natural Amazon rainforest soils. Catena 194, 104687. https://doi.org/10.1016/j.catena.2020.104687

Pendleton, R.L., Jenny, H., 1945. Factors of Soil Formation: A System of Quantitative Pedology. Geographical Review 35, 336. https://doi.org/10.2307/211491

Peng, J., Biswas, A., Jiang, Q., Zhao, R., Hu, J., Hu, B., Shi, Z., 2019. Estimating soil salinity from remote sensing and terrain data in southern Xinjiang Province, China. Geoderma 337, 1309–1319. https://doi.org/10.1016/j.geoderma.2018.08.006

Pereira, M.G., Anjos, L.H.C., Valladares, G.S., 2005. Organossolos: Ocorrência, Gênese, Classificação, Alterações pelo Uso e Manejo, in: Tópicos Ciências Do Solo. Sociedade Brasileira de Ciência do Solo, pp. 233–276.

Pereira, R.H.M., Goncalves, C.N., 2021. geobr: Download Official Spatial Data Sets of Brazil.

R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

Rahman, M.M., Zhang, X., Ahmed, I., Iqbal, Z., Zeraatpisheh, M., Kanzaki, M., Xu, M., 2020. Remote sensing-based mapping of senescent leaf C:N ratio in the sundarbans reserved

forest using machine learning techniques. Remote Sensing 12, 1375. https://doi.org/10.3390/RS12091375

Rumpel, C., Amiraslani, F., Chenu, C., Garcia Cardenas, M., Kaonga, M., Koutika, L.S., Ladha, J., Madari, B., Shirato, Y., Smith, P., Soudi, B., Soussana, J.F., Whitehead, D., Wollenberg, E., 2020. The 4p1000 initiative: Opportunities, limitations and challenges for implementing soil organic carbon sequestration as a sustainable development strategy. Ambio 49, 350–360. https://doi.org/10.1007/s13280-019-01165-2

Santos, H.G., Jacomine, P.K.T., Anjos, L.H.C., Oliveira, V.Á., Lumbreras, J.F., Coelho, M.R., Almeida, J.A., Cunha, T.J.F., Oliveira, J.B., 2018. Sistema brasileiro de classificação de solos, Embrapa Solos. EMBRAPA, Distrito Federal.

Schliep, K., Hechenbichler, K., 2016. kknn: Weighted k-Nearest Neighbors.

SEDAM, 2021. Coordenadoria de Geociências [WWW Document]. SEDAM. URL http://cogeo.sedam.ro.gov.br/ (accessed 7.18.20).

Silatsa, F.B.T., Yemefack, M., Tabi, F.O., Heuvelink, G.B.M., Leenaars, J.G.B., 2020. Assessing countrywide soil organic carbon stock using hybrid machine learning modelling and legacy soil data in Cameroon. Geoderma 367, 114260. https://doi.org/10.1016/j.geoderma.2020.114260

Smith, P., 2012. Soils and climate change. Current Opinion in Environmental Sustainability 4, 539–544. https://doi.org/10.1016/j.cosust.2012.06.005

Somarathna, P.D.S.N., Minasny, B., Malone, B.P., 2017. More Data or a Better Model? Figuring Out What Matters Most for the Spatial Prediction of Soil Carbon. Soil Science Society of America Journal 81, 1413–1426. https://doi.org/10.2136/sssaj2016.11.0376

Soucémarianadin, L.N., Cécillon, L., Guenet, B., Chenu, C., Baudin, F., Nicolas, M., Girardin, C., Barré, P., 2018. Environmental factors controlling soil organic carbon stability in French forest soils. Plant and Soil 426, 267–286. https://doi.org/10.1007/s11104-018-3613-x

Souza, C.M., Shimbo, J.Z., Rosa, M.R., Parente, L.L., Alencar, A.A., Rudorff, B.F.T., Hasenack, H., Matsumoto, M., Ferreira, L.G., Souza-Filho, P.W.M., de Oliveira, S.W., Rocha, W.F., Fonseca, A. v., Marques, C.B., Diniz, C.G., Costa, D., Monteiro, D., Rosa, E.R., Vélez-Martin, E., Weber, E.J., Lenti, F.E.B., Paternost, F.F., Pareyn, F.G.C., Siqueira, J. v., Viera, J.L., Neto, L.C.F., Saraiva, M.M., Sales, M.H., Salgado, M.P.G., Vasconcelos, R., Galano, S., Mesquita, V. v., Azevedo, T., 2020. Reconstructing three decades of land use and land cover changes in brazilian biomes with landsat archive and earth engine. Remote Sensing 12, 2735. https://doi.org/10.3390/RS12172735

Swetha, R.K., Chakraborty, S., 2021. Combination of soil texture with Nix color sensor can improve soil organic carbon prediction. Geoderma 382, 114775. https://doi.org/10.1016/j.geoderma.2020.114775

Taghizadeh-Mehrjardi, R., Nabiollahi, K., Kerry, R., 2016. Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran. Geoderma 266, 98–110. https://doi.org/10.1016/j.geoderma.2015.12.003

Tan, Q., Han, W., Li, X., Wang, G., 2020. Clarifying the response of soil organic carbon storage to increasing temperature through minimizing the precipitation effect. Geoderma 374, 114398. https://doi.org/10.1016/j.geoderma.2020.114398

Tan, Z.X., Lal, R., Smeck, N.E., Calhoun, F.G., 2004. Relationships between surface soil organic carbon pool and site variables. Geoderma 121, 187–195. https://doi.org/10.1016/j.geoderma.2003.11.003

Teixeira, W.G., Kern, D.C., Madari, B.E., Lima, H.N., Woods, W., 2009. As terras pretas de índio da Amazônia: sua caracterização e uso deste conhecimento na criação de novas áreas. Embrapa Amazônia Ocidental, Manaus.

Vasques, G.M., Dart, R.O., Baca, J.F.M., Ceddia, M.B., Santos, M. de L.M., 2017. Soil Organic Carbon Stock at 0-30 cm Map for Brazil, Embrapa. Rio de Janeiro.

Veldkamp, E., 1994. Organic Carbon Turnover in Three Tropical Soils under Pasture after Deforestation. Soil Science Society of America Journal 58, 175–180. https://doi.org/10.2136/sssaj1994.03615995005800010025x

Victor, A.D., Valery, N.N., Francois, A.I., Vanissa, T.D.C., Paulidore, M., Louis, Z., 2021. Dynamics of soil organic carbon stock under different types of Savannah agrosystems in the Sudano-Sahelian zone of Cameroon. Eurasian Journal of Soil Science 10, 51–60. https://doi.org/10.18393/ejss.809272

Wadoux, A.M.J.C., Samuel-Rosa, A., Poggio, L., Mulder, V.L., 2020. A note on knowledge discovery and machine learning in digital soil mapping. European Journal of Soil Science 71, 133–136. https://doi.org/10.1111/ejss.12909

Wang, L., Wu, C., Gu, X., Liu, H., Mei, G., Zhang, W., 2020. Probabilistic stability analysis of earth dam slope under transient seepage using multivariate adaptive regression splines. Bulletin of Engineering Geology and the Environment 79, 2763–2775. https://doi.org/10.1007/s10064-020-01730-0

Wei, Y., Ding, J., Yang, S., Yang, X., Wang, F., 2021. Comparisons of random forest and stochastic gradient treeboost algorithms for mapping soil electrical conductivity with multiple subsets using Landsat OLI and DEM/GIS-based data at a type oasis in Xinjiang,

China. European Journal of Remote Sensing 54, 158–181. https://doi.org/10.1080/22797254.2021.1888657

Yang, L., Li, X., Shi, J., Shen, F., Qi, F., Gao, B., Chen, Z., Zhu, A.X., Zhou, C., 2020. Evaluation of conditioned Latin hypercube sampling for soil mapping based on a machine learning method. Geoderma 369, 114337. https://doi.org/10.1016/j.geoderma.2020.114337

Zeraatpisheh, M., Ayoubi, S., Jafari, A., Tajik, S., Finke, P., 2019. Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. Geoderma 338, 445–452. https://doi.org/10.1016/j.geoderma.2018.09.006

Zhang, S., Liu, G., Chen, S., Rasmussen, C., Liu, B., 2021. Assessing soil thickness in a black soil watershed in northeast China using random forest and field observations. International Soil and Water Conservation Research 9, 49–57. https://doi.org/10.1016/j.iswcr.2020.09.004

Zhou, J., Li, E., Wei, H., Li, C., Qiao, Q., Armaghani, D.J., 2019. Random forests and cubist algorithms for predicting shear strengths of rockfill materials. Applied Sciences (Switzerland) 9, 1621. https://doi.org/10.3390/app9081621

Supplementary material

Table S1. Summary of performance metrics for model selection for predicting soil density in Rondônia.

Model	Tranning			Test		
	RMSE (g cm^{-3})	MAE (g cm^{-3})	R^2	RMSE (g cm^{-3})	MAE (g cm^{-3})	R^2
avNNet	0,32265	0,29202	0,39530	0,32902	0,29876	0,25979
brnn	0,08288	0,06351	0,70250	0,10655	0,08268	0,55716
cubist	0,00000	0,00000	1,00000	0,08647	0,05280	0,71521
dnn	0,14178	0,11380	0,27781	0,15453	0,12593	0,21859
earth	0,66330	0,38938	0,00531	1,10226	0,48041	0,00018
extraTrees	0,01194	0,00756	0,99504	0,08423	0,06111	0,73456
gbm	0,17629	0,14036	0,00005	0,17570	0,14107	0,01117
gcvEarth	0,08768	0,06838	0,66667	0,10742	0,08492	0,54912
pcaNNet	0,32265	0,29201	0,38014	0,32899	0,29871	0,38933
rf	0,03566	0,02608	0,96104	0,09540	0,07250	0,65812
svmLinear	0,09538	0,07284	0,60848	0,10589	0,08010	0,56186
svmPoly	0,09045	0,06789	0,64539	0,11125	0,08313	0,51496
svmRadial	0,03153	0,02057	0,95857	0,07984	0,05177	0,75110
xgbDART	0,01216	0,00907	0,99443	0,10117	0,07461	0,60193
xgbTree	0,00093	0,00069	0,99996	0,10066	0,07255	0,60255

Where: R^2 = coefficient of determination; RMSE = root mean squared error; MAE = mean absolute error.

Table S2. Importance and description of the predictors used to predict soil density in Rondônia using the svmRadial model.

Predictor	Importance	Description
SAND	100,00 %	Sand
SOC	93,86 %	Organic carbon content, Walkley and Black method
CLAY	82,02 %	Clay
HEX	61,44 %	Hydrogen extracted $cmol_c\ kg^{-1}$ (extracted with calcium acetate solution)
N	55,07 %	Total nitrogen
CTC	51,35 %	Exchangeable cation capacity (Ca + Mg + K + Na) $cmol_c\ kg^{-1}$
ACEX	50,02 %	H + Al cmol kg^{-1}
Mg	41,74 %	Mg^{2+} exchangeable in $cmol_c\ kg^{-1}$
SILT	33,79 %	Silt
CTCC	24,56 %	CTC x 100 / % clay
ECEC	22,49 %	Effective exchange capacity in $cmol_c\ kg^{-1}$
Na	19,97 %	Na^{+} exchangeable in $cmol_c\ kg^{-1}$
PHW	19,68 %	pH-H2O determined in suspension of the soil-water mixture in a 1:2.5 ratio
EC	16,55 %	Electrical conductivity determined in suspension of the soil-water mixture in a 1:2.5 ratio
FEEX	15,15 %	Extracted Fe (double acid) mg kg^{-1}
VEX	14,72 %	% of CTC base saturation
CA	14,41 %	Ca^{2+} exchangeable in $cmol_c\ kg^{-1}$
H	10,08 %	H^{+} exchangeable in $cmol_c\ kg^{-1}$
Al	7,76 %	Al^{3+} exchangeable in $cmol_c\ kg^{-1}$
CASC	6,55 %	Gravel

Table S3. Environmental predictors used as predictors of soil organic carbon stock for the state of Rondônia.

Morphometric	T	Bioclimatic	Precipitation	Soil	Vegetation
Accumulation	Tmin 1	Bio 1	Precip. 1	Geology	EVI
Aspect	Tmin 2	Bio 2	Precip. 2	Soil class	MSAVI
Convergence Index	Tmin 3	Bio 3	Precip. 3	-	NBR
Cross-Sectional Curvature	Tmin 4	Bio 4	Precip. 4	-	NBR2
Curvature Classification	Tmin 5	Bio 5	Precip. 5	-	NDVI
Difference	Tmin 6	Bio 6	Precip. 6	-	NMDI
Diffuse Insolation	Tmin 7	Bio 7	Precip. 7	-	SAVI
Direct Insolation	Tmin 8	Bio 8	Precip. 8	-	-
Direct to Diffuse Ratio	Tmin 9	Bio 9	Precip. 9	-	-
Diurnal Anisotropic Heating	Tmin 10	Bio 10	Precip. 10	-	-
Duration of Insolation	Tmin 11	Bio 11	Precip. 11	-	-
Easterness	Tmin 12	Bio 12	Precip. 12	-	-
Flow Line Curvature	Tmax 1	Bio 13	-	-	-
General Curvature	Tmax 2	Bio 14	-	-	-
Hill Height	Tmax 3	Bio 15	-	-	-
Hill Index	Tmax 4	Bio 16	-	-	-
Hillslope Index	Tmax 5	Bio 17	-	-	-
Landforms	Tmax 6	Bio 18	-	-	-
Longitudinal Curvature	Tmax 7	Bio 19	-	-	-
Mass Balance Index	Tmax 8	-	-	-	-
Maximal Curvature	Tmax 9	-	-	-	-
DEM	Tmax 10	-	-	-	-
Mid-Slope Positon	Tmax 11	-	-	-	-
Minimal Curvature	Tmax 12	-	-	-	-
Multiresolution Index of Valley Bottom Flatness (MRVBF)	-	-	-	-	-
Normalized Height	-	-	-	-	-
Northerness					
Plan Curvature	-	-	-	-	-
Profile Curvature	-	-	-	-	-
Protection Index	-	-	-	-	-
Real Surface Area	-	-	-	-	-
Slope	-	-	-	-	-
Slope Height	-	-	-	-	-
Standardized Height	-	-	-	-	-
Sunrise	-	-	-	-	-
Sunset	-	-	-	-	-
Surface Specific Points	-	-	-	-	-
Tangential Curvature	-	-	-	-	-
Terrain Ruggedness Index (TRI)	-	-	-	-	-
Terrain Surface Convexity	-	-	-	-	-
Terrain Surface Texture	-	-	-	-	-
Topographic Position Index	-	-	-	-	-
Topographic Wetness Index	-	-	-	-	-
Total Curvature	-	-	-	-	-
Total Insolation	-	-	-	-	-
Valley Depth	-	-	-	-	-
Valley Index	-	-	-	-	-
Vector Terrain Ruggedness (VRM)	-	-	-	-	-

Where: DEM = Digital elevation model.

T = Temperature; Tmin = Minimum temperature, Tmax = Maximum temperature (1, 2, 3, ..., 12 = Jan., Feb., Mar., ..., Dec.).

Bio 1 = Annual mean temperature; Bio 2 = Mean diurnal range (Mean of monthly (max temp - min temp)); Bio 3 = Isothermality (Bio 2 / Bio 7) (×100); Bio 4 = Temperature seasonality (standard deviation × 100); Bio 5 = Max temperature of warmest month; Bio 6 = Min temperature of coldest month; Bio 7 = Temperature annual range (Bio 5 – Bio 6); Bio 8 = Mean temperature of wettest quarter; Bio 9 = Mean temperature of driest quarter; Bio 10 = Mean temperature of warmest quarter; Bio 11 = Mean temperature of coldest quarter; Bio 12 = Annual

precipitation; Bio 13 = Precipitation of wettest month; Bio 14 = Precipitation of driest month; Bio 15 = Precipitation seasonality (coefficient of variation); Bio 16 = Precipitation of wettest quarter; Bio 17 = Precipitation of Driest Quarter; Bio 18 = Precipitation of Warmest Quarter; Bio 19 = Precipitation of coldest quarter.
Precip. = Precipitation (1, 2, 3, ..., 12 = Jan, Feb., Mar., ..., Dec.).
EVI = Enhanced vegetation index; MSAVI = Modified soil adjusted vegetation index; NBR = Normalized burn ratio; NBR2 = Normalized burn ratio 2; NDMI = Normalized difference moisture index; NDVI = Normalized difference vegetation index; SAVI = Soil adjusted vegetation index.

Table S4. Number of profiles used in the modeling process of soil organic carbon stock in Rondônia.

Depth (cm)	Total profiles	Training profiles	Test profiles
0-5	2905	2179	726
5-15	2901	2176	725
15-30	2854	2141	714
30-60	2799	2099	700
60-100	2564	1923	641

Table S5. Correlations of Spearman (r) above |0.95| of the removed predictor (P1) with its respective pair (P2) for data in the depth of 0-5 cm.

P1	P2	r
BIO 1	BIO 10	0,98329804
BIO 1	BIO 8	0,98397922
BIO 1	TMAX 2	0,96742248
BIO 1	TMIN 1	0,97823889
BIO 1	TMIN 10	0,97511297
BIO 1	TMIN 11	0,97691251
BIO 1	TMIN 12	0,97592476
BIO 1	TMIN 2	0,98067061
BIO 1	TMIN 3	0,98474179
BIO 1	TMIN 4	0,98304471
BIO 1	TMIN 9	0,98680086
BIO 10	BIO 1	0,98329804
BIO 10	BIO 8	0,99675181
BIO 10	TMAX 2	0,96950408
BIO 10	TMAX 4	0,96985239
BIO 10	TMIN 1	0,97556550
BIO 10	TMIN 10	0,97608984
BIO 10	TMIN 11	0,96989282
BIO 10	TMIN 12	0,97507420
BIO 10	TMIN 2	0,98649743
BIO 10	TMIN 3	0,97153104
BIO 10	TMIN 4	0,95251882
BIO 10	TMIN 9	0,96965089
BIO 11	BIO 6	0,98340249
BIO 11	BIO 9	0,99710426
BIO 11	TMIN 4	0,95486381
BIO 11	TMIN 5	0,98875722
BIO 11	TMIN 6	0,98461810
BIO 11	TMIN 7	0,98340249
BIO 11	TMIN 8	0,97532650

P1	P2	r
BIO 13	BIO 16	0,95792739
BIO 13	Precipitation 1	0,98593091
BIO 14	Precipitation 7	0,99577684
BIO 16	BIO 13	0,95792739
BIO 19	Precipitation 5	0,96309887
BIO 5	TMAX 8	0,99878088
BIO 6	BIO 11	0,98340249
BIO 6	BIO 9	0,97723914
BIO 6	TMIN 4	0,95723478
BIO 6	TMIN 5	0,99089407
BIO 6	TMIN 6	0,99378014
BIO 6	TMIN 8	0,98969091
BIO 7	BIO 2	0,98940796
BIO 8	BIO 1	0,98397922
BIO 8	BIO 10	0,99675181
BIO 8	TMAX 2	0,97025443
BIO 8	TMAX 4	0,96964567
BIO 8	TMIN 1	0,97691306
BIO 8	TMIN 10	0,97603554
BIO 8	TMIN 11	0,96965497
BIO 8	TMIN 12	0,97548750
BIO 8	TMIN 2	0,99015500
BIO 8	TMIN 3	0,97501037
BIO 8	TMIN 4	0,95577299
BIO 8	TMIN 9	0,97107043
BIO 9	BIO 11	0,99710426
BIO 9	BIO 6	0,97723914
BIO 9	TMIN 5	0,98354046
BIO 9	TMIN 6	0,97589683
BIO 9	TMIN 7	0,97723914
BIO 9	TMIN 8	0,97582245
CL	PROFC	0,98574354
CSC	CTANG	0,97921755
Diffuse inso. 1	DEM	-0,99666350
Diffuse inso. 1	Diffuse inso. 2	0,99998048
Diffuse inso. 2	DEM	-0,99618011
Diffuse inso. 2	Diffuse inso. 1	0,99998048
Direct inso. 1	Radiation total 1	0,98199233
Direct inso. 2	Radiation total 2	0,99406089
Direction	Drainage	0,99997168
Duration inso. 1	Sunrise 1	-0,97326574
MBI	GC	0,99996813
NBR 84	NBR2 84	0,95364653
NBR 84	NMDI 84	0,97077311
NBR2 84	NBR 84	0,95364653
NDVI 84	MSAVI 84	0,99930847
NDVI 84	SAVI 84	0,99999991
RSA	Slope	0,99999993
SAVI 84	MSAVI 84	0,99930935

P1	P2	r
SAVI 84	NDVI 84	0,99999991
TMAX 1	TMAX 11	0,97051949
TMAX 1	TMAX 12	0,97289970
TMAX 1	TMAX 3	0,97364138
TMAX 10	TMAX 11	0,98112488
TMAX 10	TMAX 9	0,96339989
TMAX 11	TMAX 1	0,97051949
TMAX 11	TMAX 10	0,98112488
TMAX 11	TMAX 12	0,96148396
TMAX 11	TMAX 3	0,95516676
TMAX 11	TMAX 9	0,95942670
TMAX 2	BIO 1	0,96742248
TMAX 2	BIO 10	0,96950408
TMAX 2	BIO 8	0,97025443
TMAX 2	TMAX 4	0,98383168
TMAX 2	TMAX 5	0,96522872
TMAX 3	TMAX 1	0,97364138
TMAX 3	TMAX 11	0,95516676
TMAX 3	TMAX 12	0,97927751
TMAX 9	TMAX 10	0,96339989
TMAX 9	TMAX 11	0,95942670
TMIN 1	BIO 1	0,97823889
TMIN 1	BIO 10	0,97556550
TMIN 1	BIO 8	0,97691306
TMIN 1	TMIN 10	0,99834997
TMIN 1	TMIN 11	0,99805839
TMIN 1	TMIN 12	0,99921116
TMIN 1	TMIN 2	0,98910873
TMIN 1	TMIN 3	0,99360292
TMIN 1	TMIN 4	0,98332029
TMIN 1	TMIN 9	0,99156226
TMIN 10	BIO 1	0,97511297
TMIN 10	BIO 10	0,97608984
TMIN 10	BIO 8	0,97603554
TMIN 10	TMIN 1	0,99834997
TMIN 10	TMIN 11	0,99724098
TMIN 10	TMIN 12	0,99878699
TMIN 10	TMIN 2	0,98798310
TMIN 10	TMIN 3	0,98932350
TMIN 10	TMIN 4	0,97806843
TMIN 10	TMIN 9	0,99058635
TMIN 11	BIO 1	0,97691251
TMIN 11	BIO 10	0,96989282
TMIN 11	BIO 8	0,96965497
TMIN 11	TMIN 1	0,99805839
TMIN 11	TMIN 10	0,99724098
TMIN 11	TMIN 12	0,99863095
TMIN 11	TMIN 2	0,98421763
TMIN 11	TMIN 3	0,99333306

P1	P2	r
TMIN 11	TMIN 4	0,98641891
TMIN 11	TMIN 9	0,99147762
TMIN 2	BIO 1	0,98067061
TMIN 2	BIO 10	0,98649743
TMIN 2	BIO 8	0,99015500
TMIN 2	TMIN 1	0,98910873
TMIN 2	TMIN 10	0,98798310
TMIN 2	TMIN 11	0,98421763
TMIN 2	TMIN 12	0,98864175
TMIN 2	TMIN 3	0,99048699
TMIN 2	TMIN 4	0,97189421
TMIN 2	TMIN 9	0,98338704
TMIN 3	BIO 1	0,98474179
TMIN 3	BIO 10	0,97153104
TMIN 3	BIO 8	0,97501037
TMIN 3	TMIN 1	0,99360292
TMIN 3	TMIN 10	0,98932350
TMIN 3	TMIN 11	0,99333306
TMIN 3	TMIN 12	0,99263895
TMIN 3	TMIN 2	0,99048699
TMIN 3	TMIN 4	0,99271526
TMIN 3	TMIN 9	0,99245074
TMIN 4	BIO 1	0,98304471
TMIN 4	BIO 10	0,95251882
TMIN 4	BIO 11	0,95486381
TMIN 4	BIO 6	0,95723478
TMIN 4	BIO 8	0,95577299
TMIN 4	TMIN 1	0,98332029
TMIN 4	TMIN 10	0,97806843
TMIN 4	TMIN 11	0,98641891
TMIN 4	TMIN 12	0,98147879
TMIN 4	TMIN 2	0,97189421
TMIN 4	TMIN 3	0,99271526
TMIN 4	TMIN 6	0,95112406
TMIN 4	TMIN 7	0,95723478
TMIN 4	TMIN 9	0,99181386
TMIN 5	BIO 11	0,98875722
TMIN 5	BIO 6	0,99089407
TMIN 5	BIO 9	0,98354046
TMIN 5	TMIN 6	0,99631444
TMIN 5	TMIN 7	0,99089407
TMIN 5	TMIN 8	0,98549311
TMIN 6	BIO 11	0,98461810
TMIN 6	BIO 6	0,99378014
TMIN 6	BIO 9	0,97589683
TMIN 6	TMIN 4	0,95112406
TMIN 6	TMIN 5	0,99631444
TMIN 6	TMIN 7	0,99378014
TMIN 6	TMIN 8	0,98069268

P1	P2	r
TMIN 7	BIO 11	0,98340249
TMIN 7	BIO 9	0,97723914
TMIN 7	TMIN 4	0,95723478
TMIN 7	TMIN 5	0,99089407
TMIN 7	TMIN 6	0,99378014
TMIN 7	TMIN 8	0,98969091
TMIN 9	BIO 1	0,98680086
TMIN 9	BIO 10	0,96965089
TMIN 9	BIO 8	0,97107043
TMIN 9	TMIN 1	0,99156226
TMIN 9	TMIN 10	0,99058635
TMIN 9	TMIN 11	0,99147762
TMIN 9	TMIN 12	0,98947058
TMIN 9	TMIN 2	0,98338704
TMIN 9	TMIN 3	0,99245074
TMIN 9	TMIN 4	0,99181386

Table S6. Correlations of Spearman (r) above |0.95| of the removed predictor (P1) with its respective pair (P2) for data in the depth of 5-15 cm.

P1	P2	r
BIO 1	BIO 10	0,98324886
BIO 1	BIO 8	0,98393619
BIO 1	TMAX 2	0,96739796
BIO 1	TMIN 1	0,97820259
BIO 1	TMIN 10	0,97506220
BIO 1	TMIN 11	0,97686888
BIO 1	TMIN 12	0,97587826
BIO 1	TMIN 2	0,98066399
BIO 1	TMIN 3	0,98473876
BIO 1	TMIN 4	0,98301154
BIO 1	TMIN 9	0,98678548
BIO 10	BIO 1	0,98324886
BIO 10	BIO 8	0,99675113
BIO 10	TMAX 2	0,96951682
BIO 10	TMAX 4	0,96979604
BIO 10	TMIN 1	0,97549695
BIO 10	TMIN 10	0,97602842
BIO 10	TMIN 11	0,96980876
BIO 10	TMIN 12	0,97500365
BIO 10	TMIN 2	0,98649656
BIO 10	TMIN 3	0,97146902
BIO 10	TMIN 4	0,95238446
BIO 10	TMIN 9	0,96957487
BIO 11	BIO 6	0,98338802
BIO 11	BIO 9	0,99709770
BIO 11	TMIN 4	0,95485616
BIO 11	TMIN 5	0,98875120
BIO 11	TMIN 6	0,98462993
BIO 11	TMIN 7	0,98338802

P1	P2	r
BIO 11	TMIN 8	0,97530186
BIO 13	BIO 16	0,95778246
BIO 13	Precipitation 1	0,98587188
BIO 14	Precipitation 7	0,99576379
BIO 16	BIO 13	0,95778246
BIO 19	Precipitation 5	0,96308387
BIO 5	TMAX 8	0,99877711
BIO 6	BIO 11	0,98338802
BIO 6	BIO 9	0,97720590
BIO 6	TMIN 4	0,95715058
BIO 6	TMIN 5	0,99089355
BIO 6	TMIN 6	0,99379944
BIO 6	TMIN 8	0,98967363
BIO 7	BIO 2	0,98942072
BIO 8	BIO 1	0,98393619
BIO 8	BIO 10	0,99675113
BIO 8	TMAX 2	0,97030258
BIO 8	TMAX 4	0,96961528
BIO 8	TMIN 1	0,97683637
BIO 8	TMIN 10	0,97596971
BIO 8	TMIN 11	0,96956929
BIO 8	TMIN 12	0,97540446
BIO 8	TMIN 2	0,99013964
BIO 8	TMIN 3	0,97493724
BIO 8	TMIN 4	0,95563411
BIO 8	TMIN 9	0,97099947
BIO 9	BIO 11	0,99709770
BIO 9	BIO 6	0,97720590
BIO 9	TMIN 5	0,98352688
BIO 9	TMIN 6	0,97590041
BIO 9	TMIN 7	0,97720590
BIO 9	TMIN 8	0,97578825
CL	PROFC	0,98573567
CSC	CTANG	0,97916914
Diffuse inso. 1	DEM	-0,99668710
Diffuse inso. 1	Diffuse inso. 2	0,99998060
Diffuse inso. 2	DEM	-0,99620697
Diffuse inso. 2	Diffuse inso. 1	0,99998060
Direct inso. 1	Radiation total 1	0,98191159
Direct inso. 2	Radiation total 2	0,99404542
Direction	Drainage	0,99997156
Duration inso. 1	Sunrise 1	-0,97313907
MBI	GC	0,99996808
NBR 84	NBR2 84	0,95356666
NBR 84	NMDI 84	0,97070679
NBR2 84	NBR 84	0,95356666
NDVI 84	MSAVI 84	0,99930817
NDVI 84	SAVI 84	0,99999991
RSA	Slope	0,99999993

P1	P2	r
SAVI 84	MSAVI 84	0,99930905
SAVI 84	NDVI 84	0,99999991
TMAX 1	TMAX 11	0,97052369
TMAX 1	TMAX 12	0,97293486
TMAX 1	TMAX 3	0,97361031
TMAX 10	TMAX 11	0,98109132
TMAX 10	TMAX 9	0,96338803
TMAX 11	TMAX 1	0,97052369
TMAX 11	TMAX 10	0,98109132
TMAX 11	TMAX 12	0,96143945
TMAX 11	TMAX 3	0,95509065
TMAX 11	TMAX 9	0,95943210
TMAX 2	BIO 1	0,96739796
TMAX 2	BIO 10	0,96951682
TMAX 2	BIO 8	0,97030258
TMAX 2	TMAX 4	0,98383758
TMAX 2	TMAX 5	0,96513820
TMAX 3	TMAX 1	0,97361031
TMAX 3	TMAX 11	0,95509065
TMAX 3	TMAX 12	0,97928502
TMAX 9	TMAX 10	0,96338803
TMAX 9	TMAX 11	0,95943210
TMIN 1	BIO 1	0,97820259
TMIN 1	BIO 10	0,97549695
TMIN 1	BIO 8	0,97683637
TMIN 1	TMIN 10	0,99835915
TMIN 1	TMIN 11	0,99806491
TMIN 1	TMIN 12	0,99920960
TMIN 1	TMIN 2	0,98908057
TMIN 1	TMIN 3	0,99358731
TMIN 1	TMIN 4	0,98327806
TMIN 1	TMIN 9	0,99155880
TMIN 10	BIO 1	0,97506220
TMIN 10	BIO 10	0,97602842
TMIN 10	BIO 8	0,97596971
TMIN 10	TMIN 1	0,99835915
TMIN 10	TMIN 11	0,99723402
TMIN 10	TMIN 12	0,99879336
TMIN 10	TMIN 2	0,98798501
TMIN 10	TMIN 3	0,98932155
TMIN 10	TMIN 4	0,97801105
TMIN 10	TMIN 9	0,99055796
TMIN 11	BIO 1	0,97686888
TMIN 11	BIO 10	0,96980876
TMIN 11	BIO 8	0,96956929
TMIN 11	TMIN 1	0,99806491
TMIN 11	TMIN 10	0,99723402
TMIN 11	TMIN 12	0,99863541
TMIN 11	TMIN 2	0,98420873

P1	P2	r
TMIN 11	TMIN 3	0,99334083
TMIN 11	TMIN 4	0,98639123
TMIN 11	TMIN 9	0,99145670
TMIN 2	BIO 1	0,98066399
TMIN 2	BIO 10	0,98649656
TMIN 2	BIO 8	0,99013964
TMIN 2	TMIN 1	0,98908057
TMIN 2	TMIN 10	0,98798501
TMIN 2	TMIN 11	0,98420873
TMIN 2	TMIN 12	0,98861573
TMIN 2	TMIN 3	0,99045802
TMIN 2	TMIN 4	0,97183072
TMIN 2	TMIN 9	0,98339402
TMIN 3	BIO 1	0,98473876
TMIN 3	BIO 10	0,97146902
TMIN 3	BIO 8	0,97493724
TMIN 3	TMIN 1	0,99358731
TMIN 3	TMIN 10	0,98932155
TMIN 3	TMIN 11	0,99334083
TMIN 3	TMIN 12	0,99262329
TMIN 3	TMIN 2	0,99045802
TMIN 3	TMIN 4	0,99271062
TMIN 3	TMIN 9	0,99247500
TMIN 4	BIO 1	0,98301154
TMIN 4	BIO 10	0,95238446
TMIN 4	BIO 11	0,95485616
TMIN 4	BIO 6	0,95715058
TMIN 4	BIO 8	0,95563411
TMIN 4	TMIN 1	0,98327806
TMIN 4	TMIN 10	0,97801105
TMIN 4	TMIN 11	0,98639123
TMIN 4	TMIN 12	0,98142593
TMIN 4	TMIN 2	0,97183072
TMIN 4	TMIN 3	0,99271062
TMIN 4	TMIN 6	0,95109218
TMIN 4	TMIN 7	0,95715058
TMIN 4	TMIN 9	0,99180995
TMIN 5	BIO 11	0,98875120
TMIN 5	BIO 6	0,99089355
TMIN 5	BIO 9	0,98352688
TMIN 5	TMIN 6	0,99632213
TMIN 5	TMIN 7	0,99089355
TMIN 5	TMIN 8	0,98548305
TMIN 6	BIO 11	0,98462993
TMIN 6	BIO 6	0,99379944
TMIN 6	BIO 9	0,97590041
TMIN 6	TMIN 4	0,95109218
TMIN 6	TMIN 5	0,99632213
TMIN 6	TMIN 7	0,99379944

P1	P2	r
TMIN 6	TMIN 8	0,98070813
TMIN 7	BIO 11	0,98338802
TMIN 7	BIO 9	0,97720590
TMIN 7	TMIN 4	0,95715058
TMIN 7	TMIN 5	0,99089355
TMIN 7	TMIN 6	0,99379944
TMIN 7	TMIN 8	0,98967363
TMIN 9	BIO 1	0,98678548
TMIN 9	BIO 10	0,96957487
TMIN 9	BIO 8	0,97099947
TMIN 9	TMIN 1	0,99155880
TMIN 9	TMIN 10	0,99055796
TMIN 9	TMIN 11	0,99145670
TMIN 9	TMIN 12	0,98945678
TMIN 9	TMIN 2	0,98339402
TMIN 9	TMIN 3	0,99247500
TMIN 9	TMIN 4	0,99180995

Table S7. Correlations of Spearman (r) above |0.95| of the removed predictor (P1) with its respective pair (P2) for data in the depth of 15-30 cm.

P1	P2	r
BIO 1	BIO 10	0,98302238
BIO 1	BIO 8	0,98362528
BIO 1	TMAX 2	0,96716407
BIO 1	TMIN 1	0,97853181
BIO 1	TMIN 10	0,97546311
BIO 1	TMIN 11	0,97715288
BIO 1	TMIN 12	0,97620243
BIO 1	TMIN 2	0,98057772
BIO 1	TMIN 3	0,98478472
BIO 1	TMIN 4	0,98293718
BIO 1	TMIN 9	0,98698676
BIO 10	BIO 1	0,98302238
BIO 10	BIO 8	0,99675268
BIO 10	TMAX 2	0,96990518
BIO 10	TMAX 4	0,96975140
BIO 10	TMIN 1	0,97529072
BIO 10	TMIN 10	0,97590263
BIO 10	TMIN 11	0,96944389
BIO 10	TMIN 12	0,97479572
BIO 10	TMIN 2	0,98631398
BIO 10	TMIN 3	0,97097223
BIO 10	TMIN 4	0,95153711
BIO 10	TMIN 9	0,96907149
BIO 11	BIO 6	0,98379133
BIO 11	BIO 9	0,99711687
BIO 11	TMIN 4	0,95577160
BIO 11	TMIN 5	0,98884743
BIO 11	TMIN 6	0,98487880

P1	P2	r
BIO 11	TMIN 7	0,98379133
BIO 11	TMIN 8	0,97538729
BIO 13	BIO 16	0,95750749
BIO 13	Precipitation 1	0,98547326
BIO 14	Precipitation 7	0,99567676
BIO 16	BIO 13	0,95750749
BIO 19	Precipitation 5	0,96261884
BIO 5	TMAX 8	0,99875478
BIO 6	BIO 11	0,98379133
BIO 6	BIO 9	0,97766121
BIO 6	TMIN 4	0,95786069
BIO 6	TMIN 5	0,99106627
BIO 6	TMIN 6	0,99381414
BIO 6	TMIN 8	0,98972678
BIO 7	BIO 2	0,98946400
BIO 8	BIO 1	0,98362528
BIO 8	BIO 10	0,99675268
BIO 8	TMAX 2	0,97051746
BIO 8	TMAX 4	0,96952741
BIO 8	TMIN 1	0,97668345
BIO 8	TMIN 10	0,97588593
BIO 8	TMIN 11	0,96924589
BIO 8	TMIN 12	0,97525213
BIO 8	TMIN 2	0,99000175
BIO 8	TMIN 3	0,97447046
BIO 8	TMIN 4	0,95476511
BIO 8	TMIN 9	0,97047477
BIO 9	BIO 11	0,99711687
BIO 9	BIO 6	0,97766121
BIO 9	TMIN 5	0,98359184
BIO 9	TMIN 6	0,97617126
BIO 9	TMIN 7	0,97766121
BIO 9	TMIN 8	0,97583752
CL	PROFC	0,98550567
CSC	CTANG	0,97880126
Diffuse inso. 1	DEM	-0,99704992
Diffuse inso. 1	Diffuse inso. 2	0,99998221
Diffuse inso. 2	DEM	-0,99661685
Diffuse inso. 2	Diffuse inso. 1	0,99998221
Direct inso. 1	Radiation total 1	0,98159707
Direct inso. 2	Radiation total 2	0,99392106
Direction	Drainage	0,99997121
Duration inso. 1	Sunrise 1	-0,97567800
GC	MBI	0,99996748
NBR 84	NBR2 84	0,95361409
NBR 84	NMDI 84	0,97036341
NBR2 84	NBR 84	0,95361409
NDVI 84	MSAVI 84	0,99931894
NDVI 84	SAVI 84	0,99999990

P1	P2	r
RSA	Slope	0,99999993
SAVI 84	MSAVI 84	0,99931985
SAVI 84	NDVI 84	0,99999990
TMAX 1	TMAX 11	0,97125103
TMAX 1	TMAX 12	0,97292547
TMAX 1	TMAX 3	0,97367440
TMAX 10	TMAX 11	0,98134858
TMAX 10	TMAX 9	0,96328703
TMAX 11	TMAX 1	0,97125103
TMAX 11	TMAX 10	0,98134858
TMAX 11	TMAX 12	0,96184780
TMAX 11	TMAX 3	0,95564332
TMAX 11	TMAX 9	0,95935426
TMAX 2	BIO 1	0,96716407
TMAX 2	BIO 10	0,96990518
TMAX 2	BIO 8	0,97051746
TMAX 2	TMAX 4	0,98396846
TMAX 2	TMAX 5	0,96488572
TMAX 3	TMAX 1	0,97367440
TMAX 3	TMAX 11	0,95564332
TMAX 3	TMAX 12	0,97933623
TMIN 1	BIO 1	0,97853181
TMIN 1	BIO 10	0,97529072
TMIN 1	BIO 8	0,97668345
TMIN 1	TMIN 10	0,99836600
TMIN 1	TMIN 11	0,99803437
TMIN 1	TMIN 12	0,99920824
TMIN 1	TMIN 2	0,98898867
TMIN 1	TMIN 3	0,99361872
TMIN 1	TMIN 4	0,98325752
TMIN 1	TMIN 9	0,99148464
TMIN 10	BIO 1	0,97546311
TMIN 10	BIO 10	0,97590263
TMIN 10	BIO 8	0,97588593
TMIN 10	TMIN 1	0,99836600
TMIN 10	TMIN 11	0,99721879
TMIN 10	TMIN 12	0,99879299
TMIN 10	TMIN 2	0,98798367
TMIN 10	TMIN 3	0,98939485
TMIN 10	TMIN 4	0,97799997
TMIN 10	TMIN 9	0,99053059
TMIN 11	BIO 1	0,97715288
TMIN 11	BIO 10	0,96944389
TMIN 11	BIO 8	0,96924589
TMIN 11	TMIN 1	0,99803437
TMIN 11	TMIN 10	0,99721879
TMIN 11	TMIN 12	0,99861956
TMIN 11	TMIN 2	0,98404312
TMIN 11	TMIN 3	0,99337943

P1	P2	r
TMIN 11	TMIN 4	0,98641464
TMIN 11	TMIN 9	0,99141267
TMIN 2	BIO 1	0,98057772
TMIN 2	BIO 10	0,98631398
TMIN 2	BIO 8	0,99000175
TMIN 2	TMIN 1	0,98898867
TMIN 2	TMIN 10	0,98798367
TMIN 2	TMIN 11	0,98404312
TMIN 2	TMIN 12	0,98856303
TMIN 2	TMIN 3	0,99026181
TMIN 2	TMIN 4	0,97136061
TMIN 2	TMIN 9	0,98308670
TMIN 3	BIO 1	0,98478472
TMIN 3	BIO 10	0,97097223
TMIN 3	BIO 8	0,97447046
TMIN 3	TMIN 1	0,99361872
TMIN 3	TMIN 10	0,98939485
TMIN 3	TMIN 11	0,99337943
TMIN 3	TMIN 12	0,99268764
TMIN 3	TMIN 2	0,99026181
TMIN 3	TMIN 4	0,99263931
TMIN 3	TMIN 9	0,99237654
TMIN 4	BIO 1	0,98293718
TMIN 4	BIO 10	0,95153711
TMIN 4	BIO 11	0,95577160
TMIN 4	BIO 6	0,95786069
TMIN 4	BIO 8	0,95476511
TMIN 4	TMIN 1	0,98325752
TMIN 4	TMIN 10	0,97799997
TMIN 4	TMIN 11	0,98641464
TMIN 4	TMIN 12	0,98140305
TMIN 4	TMIN 2	0,97136061
TMIN 4	TMIN 3	0,99263931
TMIN 4	TMIN 6	0,95195286
TMIN 4	TMIN 7	0,95786069
TMIN 4	TMIN 9	0,99178247
TMIN 5	BIO 11	0,98884743
TMIN 5	BIO 6	0,99106627
TMIN 5	BIO 9	0,98359184
TMIN 5	TMIN 6	0,99641278
TMIN 5	TMIN 7	0,99106627
TMIN 5	TMIN 8	0,98538659
TMIN 6	BIO 11	0,98487880
TMIN 6	BIO 6	0,99381414
TMIN 6	BIO 9	0,97617126
TMIN 6	TMIN 4	0,95195286
TMIN 6	TMIN 5	0,99641278
TMIN 6	TMIN 7	0,99381414
TMIN 6	TMIN 8	0,98062696

P1	P2	r
TMIN 7	BIO 11	0,98379133
TMIN 7	BIO 9	0,97766121
TMIN 7	TMIN 4	0,95786069
TMIN 7	TMIN 5	0,99106627
TMIN 7	TMIN 6	0,99381414
TMIN 7	TMIN 8	0,98972678
TMIN 9	BIO 1	0,98698676
TMIN 9	BIO 10	0,96907149
TMIN 9	BIO 8	0,97047477
TMIN 9	TMIN 1	0,99148464
TMIN 9	TMIN 10	0,99053059
TMIN 9	TMIN 11	0,99141267
TMIN 9	TMIN 12	0,98939198
TMIN 9	TMIN 2	0,98308670
TMIN 9	TMIN 3	0,99237654
TMIN 9	TMIN 4	0,99178247

Table S8. Correlations of Spearman (r) above |0.95| of the removed predictor (P1) with its respective pair (P2) for data in the depth of 30-60 cm.

P1	P2	r
BIO 1	BIO 10	0,98263990
BIO 1	BIO 8	0,98317772
BIO 1	TMAX 2	0,96651368
BIO 1	TMIN 1	0,97883316
BIO 1	TMIN 10	0,97578323
BIO 1	TMIN 11	0,97738964
BIO 1	TMIN 12	0,97649488
BIO 1	TMIN 2	0,98039092
BIO 1	TMIN 3	0,98477287
BIO 1	TMIN 4	0,98276415
BIO 1	TMIN 9	0,98711885
BIO 10	BIO 1	0,98263990
BIO 10	BIO 8	0,99674123
BIO 10	TMAX 2	0,96964374
BIO 10	TMAX 4	0,96923922
BIO 10	TMIN 1	0,97507066
BIO 10	TMIN 10	0,97584122
BIO 10	TMIN 11	0,96901786
BIO 10	TMIN 12	0,97461061
BIO 10	TMIN 2	0,98621233
BIO 10	TMIN 3	0,97034525
BIO 10	TMIN 4	0,95035287
BIO 10	TMIN 9	0,96851246
BIO 11	BIO 6	0,98410152
BIO 11	BIO 9	0,99710471
BIO 11	TMIN 4	0,95626459
BIO 11	TMIN 5	0,98899526
BIO 11	TMIN 6	0,98510150
BIO 11	TMIN 7	0,98410152

P1	P2	r
BIO 11	TMIN 8	0,97543428
BIO 13	BIO 16	0,95717073
BIO 13	Precipitation 1	0,98490880
BIO 14	Precipitation 7	0,99554958
BIO 16	BIO 13	0,95717073
BIO 19	Precipitation 5	0,96247897
BIO 5	TMAX 8	0,99878054
BIO 6	BIO 11	0,98410152
BIO 6	BIO 9	0,97793196
BIO 6	TMIN 4	0,95869398
BIO 6	TMIN 5	0,99111322
BIO 6	TMIN 6	0,99377131
BIO 6	TMIN 8	0,98968747
BIO 7	BIO 2	0,98955864
BIO 8	BIO 1	0,98317772
BIO 8	BIO 10	0,99674123
BIO 8	TMAX 2	0,97033883
BIO 8	TMAX 4	0,96912688
BIO 8	TMIN 1	0,97637265
BIO 8	TMIN 10	0,97569699
BIO 8	TMIN 11	0,96873766
BIO 8	TMIN 12	0,97498407
BIO 8	TMIN 2	0,98983815
BIO 8	TMIN 3	0,97376527
BIO 8	TMIN 4	0,95348221
BIO 8	TMIN 9	0,96978110
BIO 9	BIO 11	0,99710471
BIO 9	BIO 6	0,97793196
BIO 9	TMIN 5	0,98361024
BIO 9	TMIN 6	0,97627757
BIO 9	TMIN 7	0,97793196
BIO 9	TMIN 8	0,97589312
CL	PROFC	0,98569364
CSC	CTANG	0,97880196
Diffuse inso. 1	DEM	-0,99732087
Diffuse inso. 1	Diffuse inso. 2	0,99998324
Diffuse inso. 2	DEM	-0,99692049
Diffuse inso. 2	Diffuse inso. 1	0,99998324
Direct inso. 1	Radiation total 1	0,98120122
Direct inso. 2	Radiation total 2	0,99380178
Direction	Drainage	0,99997091
Duration inso. 1	Sunrise 1	-0,97502263
GC	MBI	0,99996688
NBR 84	NBR2 84	0,95283809
NBR 84	NMDI 84	0,97000238
NBR2 84	NBR 84	0,95283809
NDVI 84	MSAVI 84	0,99931266
NDVI 84	SAVI 84	0,99999990
RSA	Slope	0,99999993

P1	P2	r
SAVI 84	MSAVI 84	0,99931362
SAVI 84	NDVI 84	0,99999990
TMAX 1	TMAX 11	0,97125498
TMAX 1	TMAX 12	0,97271689
TMAX 1	TMAX 3	0,97355014
TMAX 10	TMAX 11	0,98134352
TMAX 10	TMAX 9	0,96270715
TMAX 11	TMAX 1	0,97125498
TMAX 11	TMAX 10	0,98134352
TMAX 11	TMAX 12	0,96170476
TMAX 11	TMAX 3	0,95595747
TMAX 11	TMAX 9	0,95869295
TMAX 2	BIO 1	0,96651368
TMAX 2	BIO 10	0,96964374
TMAX 2	BIO 8	0,97033883
TMAX 2	TMAX 4	0,98364259
TMAX 2	TMAX 5	0,96462604
TMAX 3	TMAX 1	0,97355014
TMAX 3	TMAX 11	0,95595747
TMAX 3	TMAX 12	0,97932617
TMIN 1	BIO 1	0,97883316
TMIN 1	BIO 10	0,97507066
TMIN 1	BIO 8	0,97637265
TMIN 1	TMIN 10	0,99835418
TMIN 1	TMIN 11	0,99802519
TMIN 1	TMIN 12	0,99921942
TMIN 1	TMIN 2	0,98888035
TMIN 1	TMIN 3	0,99366680
TMIN 1	TMIN 4	0,98313589
TMIN 1	TMIN 9	0,99144182
TMIN 10	BIO 1	0,97578323
TMIN 10	BIO 10	0,97584122
TMIN 10	BIO 8	0,97569699
TMIN 10	TMIN 1	0,99835418
TMIN 10	TMIN 11	0,99716015
TMIN 10	TMIN 12	0,99878678
TMIN 10	TMIN 2	0,98800096
TMIN 10	TMIN 3	0,98942044
TMIN 10	TMIN 4	0,97777346
TMIN 10	TMIN 9	0,99049162
TMIN 11	BIO 1	0,97738964
TMIN 11	BIO 10	0,96901786
TMIN 11	BIO 8	0,96873766
TMIN 11	TMIN 1	0,99802519
TMIN 11	TMIN 10	0,99716015
TMIN 11	TMIN 12	0,99859901
TMIN 11	TMIN 2	0,98385355
TMIN 11	TMIN 3	0,99346168
TMIN 11	TMIN 4	0,98637120

P1	P2	r
TMIN 11	TMIN 9	0,99138139
TMIN 2	BIO 1	0,98039092
TMIN 2	BIO 10	0,98621233
TMIN 2	BIO 8	0,98983815
TMIN 2	TMIN 1	0,98888035
TMIN 2	TMIN 10	0,98800096
TMIN 2	TMIN 11	0,98385355
TMIN 2	TMIN 12	0,98851812
TMIN 2	TMIN 3	0,98994631
TMIN 2	TMIN 4	0,97055629
TMIN 2	TMIN 9	0,98270405
TMIN 3	BIO 1	0,98477287
TMIN 3	BIO 10	0,97034525
TMIN 3	BIO 8	0,97376527
TMIN 3	TMIN 1	0,99366680
TMIN 3	TMIN 10	0,98942044
TMIN 3	TMIN 11	0,99346168
TMIN 3	TMIN 12	0,99275048
TMIN 3	TMIN 2	0,98994631
TMIN 3	TMIN 4	0,99250383
TMIN 3	TMIN 9	0,99227526
TMIN 4	BIO 1	0,98276415
TMIN 4	BIO 10	0,95035287
TMIN 4	BIO 11	0,95626459
TMIN 4	BIO 6	0,95869398
TMIN 4	BIO 8	0,95348221
TMIN 4	TMIN 1	0,98313589
TMIN 4	TMIN 10	0,97777346
TMIN 4	TMIN 11	0,98637120
TMIN 4	TMIN 12	0,98126039
TMIN 4	TMIN 2	0,97055629
TMIN 4	TMIN 3	0,99250383
TMIN 4	TMIN 6	0,95268778
TMIN 4	TMIN 7	0,95869398
TMIN 4	TMIN 9	0,99162475
TMIN 5	BIO 11	0,98899526
TMIN 5	BIO 6	0,99111322
TMIN 5	BIO 9	0,98361024
TMIN 5	TMIN 6	0,99646918
TMIN 5	TMIN 7	0,99111322
TMIN 5	TMIN 8	0,98505907
TMIN 6	BIO 11	0,98510150
TMIN 6	BIO 6	0,99377131
TMIN 6	BIO 9	0,97627757
TMIN 6	TMIN 4	0,95268778
TMIN 6	TMIN 5	0,99646918
TMIN 6	TMIN 7	0,99377131
TMIN 6	TMIN 8	0,98032819
TMIN 7	BIO 11	0,98410152

P1	P2	r
TMIN 7	BIO 9	0,97793196
TMIN 7	TMIN 4	0,95869398
TMIN 7	TMIN 5	0,99111322
TMIN 7	TMIN 6	0,99377131
TMIN 7	TMIN 8	0,98968747
TMIN 9	BIO 1	0,98711885
TMIN 9	BIO 10	0,96851246
TMIN 9	BIO 8	0,96978110
TMIN 9	TMIN 1	0,99144182
TMIN 9	TMIN 10	0,99049162
TMIN 9	TMIN 11	0,99138139
TMIN 9	TMIN 12	0,98935661
TMIN 9	TMIN 2	0,98270405
TMIN 9	TMIN 3	0,99227526
TMIN 9	TMIN 4	0,99162475

Table S9. Correlations of Spearman (r) above |0.95| of the removed predictor (P1) with its respective pair (P2) for data in the depth of 60-100 cm.

P1	P2	r
BIO 1	BIO 10	0,98216457
BIO 1	BIO 8	0,98215974
BIO 1	TMAX 2	0,96527540
BIO 1	TMIN 1	0,98108840
BIO 1	TMIN 10	0,97834869
BIO 1	TMIN 11	0,97950712
BIO 1	TMIN 12	0,97876420
BIO 1	TMIN 2	0,97995340
BIO 1	TMIN 3	0,98492197
BIO 1	TMIN 4	0,98245790
BIO 1	TMIN 9	0,98802610
BIO 10	BIO 1	0,98216457
BIO 10	BIO 8	0,99681854
BIO 10	TMAX 2	0,97043578
BIO 10	TMAX 4	0,96920515
BIO 10	TMIN 1	0,97530860
BIO 10	TMIN 10	0,97641074
BIO 10	TMIN 11	0,96856477
BIO 10	TMIN 12	0,97480971
BIO 10	TMIN 2	0,98576014
BIO 10	TMIN 3	0,96886597
BIO 10	TMIN 9	0,96714624
BIO 11	BIO 6	0,98588956
BIO 11	BIO 9	0,99717720
BIO 11	TMIN 4	0,95950642
BIO 11	TMIN 5	0,98931261
BIO 11	TMIN 6	0,98603073
BIO 11	TMIN 7	0,98588956
BIO 11	TMIN 8	0,97511929
BIO 13	BIO 16	0,95626128

P1	P2	r
BIO 13	Precipitation 1	0,98309097
BIO 14	Precipitation 7	0,99500271
BIO 16	BIO 13	0,95626128
BIO 16	Precipitation 3	0,95122242
BIO 19	Precipitation 5	0,96193506
BIO 5	TMAX 8	0,99870068
BIO 6	BIO 11	0,98588956
BIO 6	BIO 7	-0,95013120
BIO 6	BIO 9	0,98012340
BIO 6	TMIN 4	0,96339883
BIO 6	TMIN 5	0,99165521
BIO 6	TMIN 6	0,99342052
BIO 6	TMIN 8	0,98989816
BIO 6	TMIN 9	0,95126525
BIO 7	BIO 2	0,98967475
BIO 7	BIO 6	-0,95013120
BIO 7	TMIN 7	-0,95013120
BIO 8	BIO 1	0,98215974
BIO 8	BIO 10	0,99681854
BIO 8	TMAX 2	0,97072602
BIO 8	TMAX 4	0,96898368
BIO 8	TMIN 1	0,97657376
BIO 8	TMIN 10	0,97604476
BIO 8	TMIN 11	0,96807468
BIO 8	TMIN 12	0,97512159
BIO 8	TMIN 2	0,98944419
BIO 8	TMIN 3	0,97216193
BIO 8	TMIN 4	0,95007014
BIO 8	TMIN 9	0,96794182
BIO 9	BIO 11	0,99717720
BIO 9	BIO 6	0,98012340
BIO 9	TMIN 5	0,98362584
BIO 9	TMIN 6	0,97720228
BIO 9	TMIN 7	0,98012340
BIO 9	TMIN 8	0,97575284
CL	PROFC	0,98586764
CSC	CTANG	0,97922434
Diffuse inso. 1	DEM	-0,99750900
Diffuse inso. 1	Diffuse inso. 2	0,99998372
Diffuse inso. 2	DEM	-0,99712891
Diffuse inso. 2	Diffuse inso. 1	0,99998372
Direct inso. 1	Radiation total 1	0,98079819
Direct inso. 2	Radiation total 2	0,99348804
Direction	Drainage	0,99996699
Duration inso. 1	Sunrise 1	-0,98138794
GC	MBI	0,99996594
NBR 84	NBR2 84	0,95094580
NBR 84	NMDI 84	0,96823916
NBR2 84	NBR 84	0,95094580

P1	P2	r
NDVI 84	MSAVI 84	0,99928772
NDVI 84	SAVI 84	0,99999990
Precipitation 3	BIO 16	0,95122242
RSA	Slope	0,99999991
SAVI 84	MSAVI 84	0,99928877
SAVI 84	NDVI 84	0,99999990
TMAX 1	TMAX 11	0,97303605
TMAX 1	TMAX 12	0,97309390
TMAX 1	TMAX 3	0,97449025
TMAX 10	TMAX 11	0,98221506
TMAX 10	TMAX 9	0,96256267
TMAX 11	TMAX 1	0,97303605
TMAX 11	TMAX 10	0,98221506
TMAX 11	TMAX 12	0,96452981
TMAX 11	TMAX 3	0,95911238
TMAX 11	TMAX 9	0,95789570
TMAX 2	BIO 1	0,96527540
TMAX 2	BIO 10	0,97043578
TMAX 2	BIO 8	0,97072602
TMAX 2	TMAX 4	0,98327256
TMAX 2	TMAX 5	0,96340546
TMAX 3	TMAX 1	0,97449025
TMAX 3	TMAX 11	0,95911238
TMAX 3	TMAX 12	0,97943554
TMAX 9	TMAX 10	0,96256267
TMAX 9	TMAX 11	0,95789570
TMIN 1	BIO 1	0,98108840
TMIN 1	BIO 10	0,97530860
TMIN 1	BIO 8	0,97657376
TMIN 1	TMIN 10	0,99821663
TMIN 1	TMIN 11	0,99782716
TMIN 1	TMIN 12	0,99919774
TMIN 1	TMIN 2	0,98889941
TMIN 1	TMIN 3	0,99403655
TMIN 1	TMIN 4	0,98318666
TMIN 1	TMIN 9	0,99131497
TMIN 10	BIO 1	0,97834869
TMIN 10	BIO 10	0,97641074
TMIN 10	BIO 8	0,97604476
TMIN 10	TMIN 1	0,99821663
TMIN 10	TMIN 11	0,99700415
TMIN 10	TMIN 12	0,99871597
TMIN 10	TMIN 2	0,98810483
TMIN 10	TMIN 3	0,98966398
TMIN 10	TMIN 4	0,97776304
TMIN 10	TMIN 9	0,99069365
TMIN 11	BIO 1	0,97950712
TMIN 11	BIO 10	0,96856477
TMIN 11	BIO 8	0,96807468

P1	P2	r
TMIN 11	TMIN 1	0,99782716
TMIN 11	TMIN 10	0,99700415
TMIN 11	TMIN 12	0,99848173
TMIN 11	TMIN 2	0,98323648
TMIN 11	TMIN 3	0,99370818
TMIN 11	TMIN 4	0,98668212
TMIN 11	TMIN 9	0,99147389
TMIN 2	BIO 1	0,97995340
TMIN 2	BIO 10	0,98576014
TMIN 2	BIO 8	0,98944419
TMIN 2	TMIN 1	0,98889941
TMIN 2	TMIN 10	0,98810483
TMIN 2	TMIN 11	0,98323648
TMIN 2	TMIN 12	0,98856620
TMIN 2	TMIN 3	0,98916987
TMIN 2	TMIN 4	0,96813059
TMIN 2	TMIN 9	0,98122336
TMIN 3	BIO 1	0,98492197
TMIN 3	BIO 10	0,96886597
TMIN 3	BIO 8	0,97216193
TMIN 3	TMIN 1	0,99403655
TMIN 3	TMIN 10	0,98966398
TMIN 3	TMIN 11	0,99370818
TMIN 3	TMIN 12	0,99319200
TMIN 3	TMIN 2	0,98916987
TMIN 3	TMIN 4	0,99190711
TMIN 3	TMIN 9	0,99149800
TMIN 4	BIO 1	0,98245790
TMIN 4	BIO 11	0,95950642
TMIN 4	BIO 6	0,96339883
TMIN 4	BIO 8	0,95007014
TMIN 4	TMIN 1	0,98318666
TMIN 4	TMIN 10	0,97776304
TMIN 4	TMIN 11	0,98668212
TMIN 4	TMIN 12	0,98137580
TMIN 4	TMIN 2	0,96813059
TMIN 4	TMIN 3	0,99190711
TMIN 4	TMIN 5	0,95088892
TMIN 4	TMIN 6	0,95637626
TMIN 4	TMIN 7	0,96339883
TMIN 4	TMIN 9	0,99110216
TMIN 5	BIO 11	0,98931261
TMIN 5	BIO 6	0,99165521
TMIN 5	BIO 9	0,98362584
TMIN 5	TMIN 4	0,95088892
TMIN 5	TMIN 6	0,99688617
TMIN 5	TMIN 7	0,99165521
TMIN 5	TMIN 8	0,98367884
TMIN 6	BIO 11	0,98603073

P1	P2	r
TMIN 6	BIO 6	0,99342052
TMIN 6	BIO 9	0,97720228
TMIN 6	TMIN 4	0,95637626
TMIN 6	TMIN 5	0,99688617
TMIN 6	TMIN 7	0,99342052
TMIN 6	TMIN 8	0,97908309
TMIN 7	BIO 11	0,98588956
TMIN 7	BIO 7	-0,95013120
TMIN 7	BIO 9	0,98012340
TMIN 7	TMIN 4	0,96339883
TMIN 7	TMIN 5	0,99165521
TMIN 7	TMIN 6	0,99342052
TMIN 7	TMIN 8	0,98989816
TMIN 7	TMIN 9	0,95126525
TMIN 9	BIO 1	0,98802610
TMIN 9	BIO 10	0,96714624
TMIN 9	BIO 6	0,95126525
TMIN 9	BIO 8	0,96794182
TMIN 9	TMIN 1	0,99131497
TMIN 9	TMIN 10	0,99069365
TMIN 9	TMIN 11	0,99147389
TMIN 9	TMIN 12	0,98934536
TMIN 9	TMIN 2	0,98122336
TMIN 9	TMIN 3	0,99149800
TMIN 9	TMIN 4	0,99110216
TMIN 9	TMIN 7	0,95126525

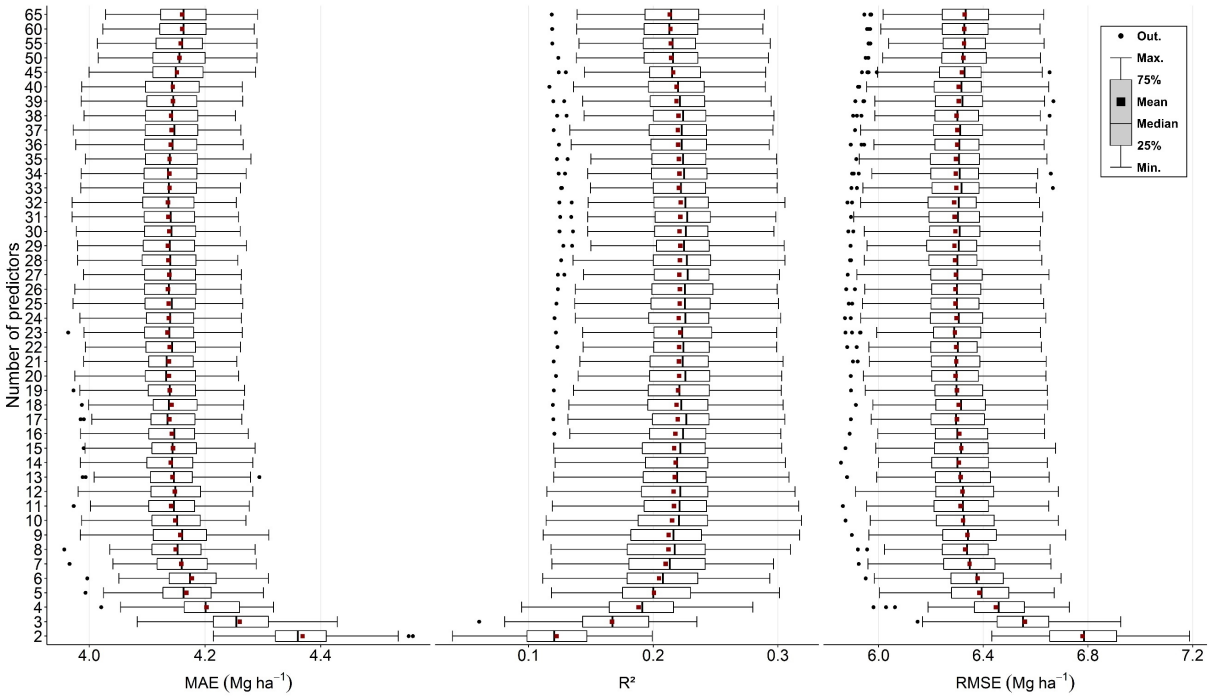

Figure S1. Performance of the recursive feature elimination for the random forest algorithm in the different subsets of predictors of Rondônia's SOC stock in the 0-5 cm layer. R^2 = coefficient of determination; RMSE = root mean squared error; MAE = mean absolute error.

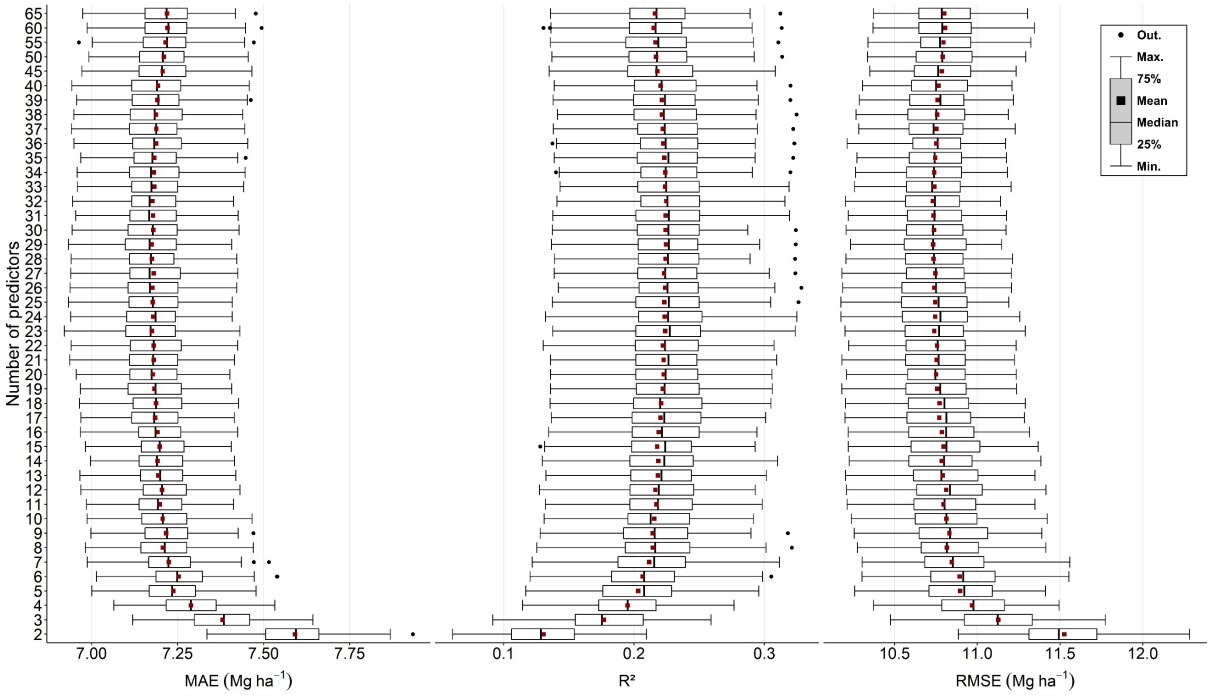

Figure S2. Performance of the recursive feature elimination for the random forest algorithm in the different subsets of predictors of Rondônia's SOC stock in the 5-15 cm layer. R^2 = coefficient of determination; RMSE = root mean squared error; MAE = mean absolute error.

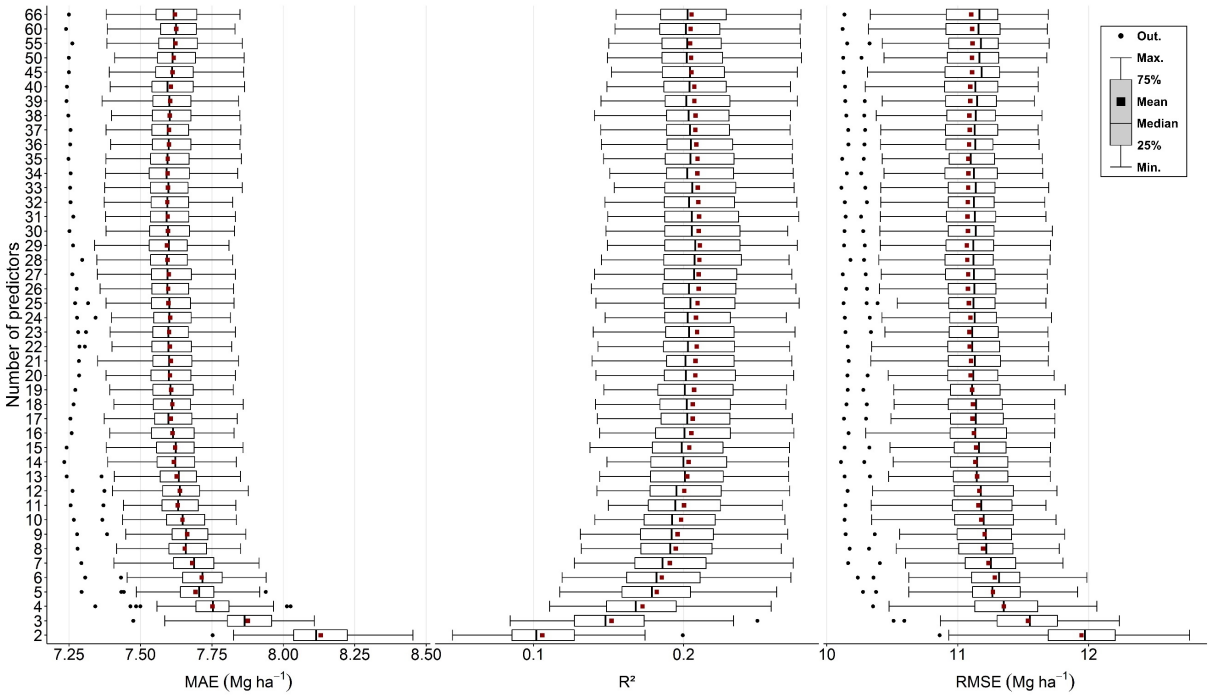

Figure S3. Performance of the recursive feature elimination for the random forest algorithm in the different subsets of predictors of Rondônia's SOC stock in the 15-30 cm layer. R^2 = coefficient of determination; RMSE = root mean squared error; MAE = mean absolute error.

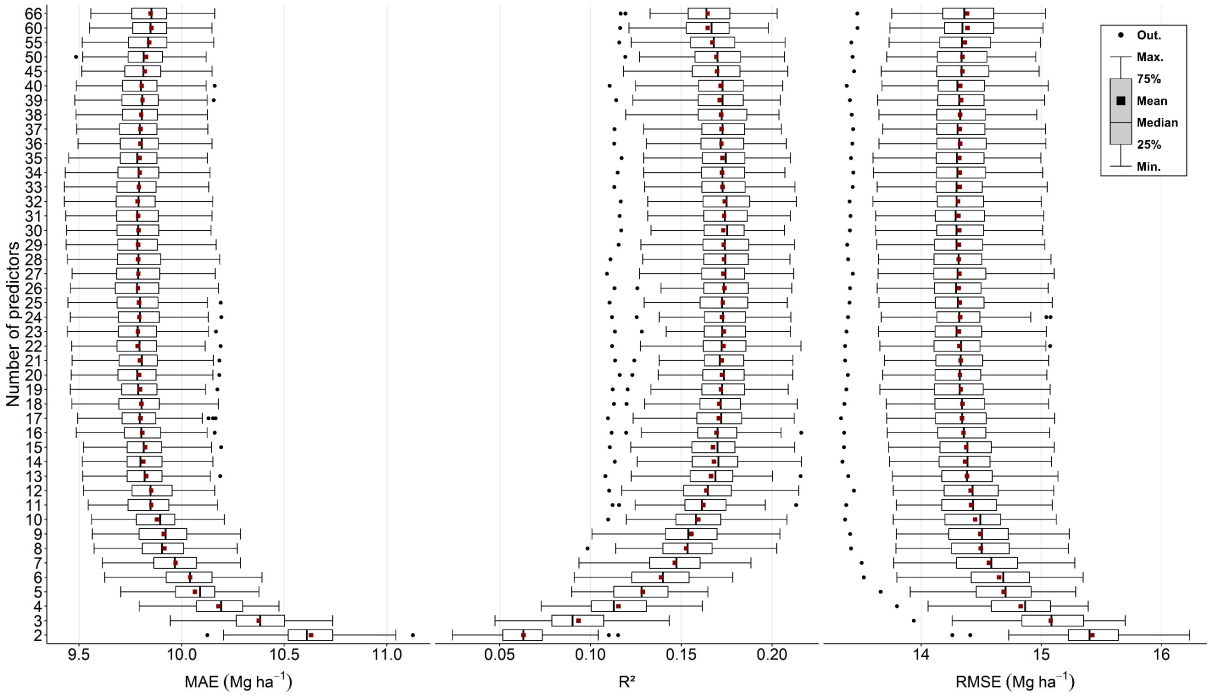

Figure S4. Performance of the recursive feature elimination for the random forest algorithm in the different subsets of predictors of Rondônia's SOC stock in the 30-60 cm layer. R^2 = coefficient of determination; RMSE = root mean squared error; MAE = mean absolute error.

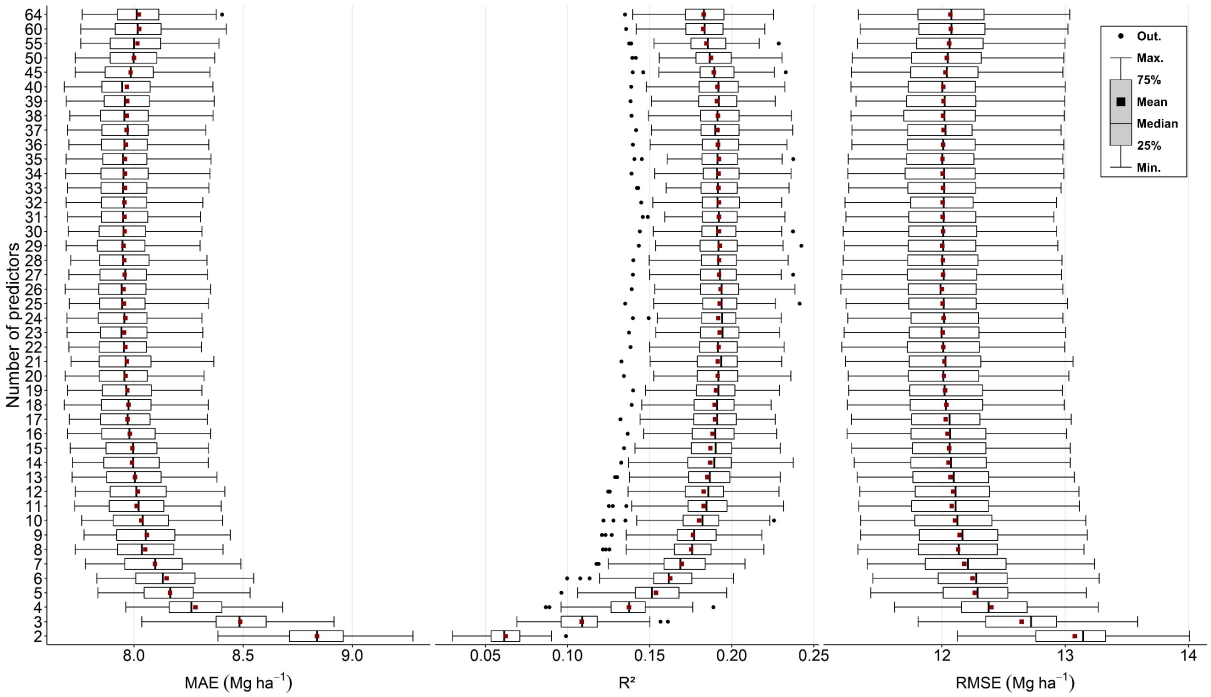

Figure S5. Performance of the recursive feature elimination for the random forest algorithm in the different subsets of predictors of Rondônia's SOC stock in the 60-100 cm layer. R^2 = coefficient of determination; RMSE = root mean squared error; MAE = mean absolute error.

Chapter 2: Dynamics of soil organic carbon stock in future climate change scenarios in southwestern Amazonia

Cássio Marques Moquedace dos Santos[1], Lucas de Carvalho Gomes[1], Marcos Guedes de Lana[1], Gustavo Vieira Veloso[1], Elpídio Inácio Fernandes-Filho[1], Irene Maria Cardoso[1], Emanuel Fernando Maia de Souza[2], Renildes Lúcio Ferreira Fontes[1]

[1]Department of Soils, Universidade Federal de Viçosa, Brazil; [2]Department of Forestry Engineering, Federal University of Rondônia, Brazil

Abstract

Soil is the largest carbon reservoir in the terrestrial biosphere and understanding the relationship between soil organic carbon (SOC) and its environmental drivers is essential to anticipate the impact of environmental changes on SOC, especially under regional conditions. Therefore, this study aimed to model and map the soil organic carbon stock in the state of Rondônia in future scenarios until 2100. We used 2,914 soil profiles distributed throughout Rondônia and environmental predictors such as relief, climate, soil class and geology. The predictors were submitted to variable selection, first by correlation next by the recursive feature elimination (RFE) algorithm. To predict SOC stock changes between 2020 and 2100, we used climate change data projection from contrasting CMIP6 scenarios (ssp126 and ssp585), predicted with the CanESM5 and MIROC6 models. The model selected for SOC prediction was the random forest (rf) with mean R^2 ranging between 0.17 and 0.25. Soil class, temperature and precipitation were the dominant predictors of SOC stock in Rondônia. The central territory and the northern region of the Vale do Guaporé and Zona da Mata contain the smallest stock of SOC and the largest extend from the southwest of the Vale do Guaporé to the south of the Zona da Mata. The lowest SOC stock are in areas with a predominance of Latossolos that suffered strong anthropogenic action. The areas with the largest SOC stock are in places with Gleissolos and Organossolos. Higher altitude regions will suffer a high reduction in SOC stock in future scenarios, especially in ssp585, probably associated with an increase in temperature and water availability, which consequently increases soil biota respiration. In areas where SOC stock are limited, projections indicate an increase. In these areas, there will be an increase in temperature, however, precipitation will be reduced in greater intensity. This water limitation can reduce microbial activity and justify such a trend in the model. In the two climate models (CanESM5 and MIROC6), the differences in the impact on the SOC will be in areas with more carbon,

with CanESM5 with the greatest variations in climate and in SOC demonstrating how much the SOC of Rondônia can be climate dependent. Although there is a strong relationship between climate predictors and SOC, caution is recommended when interpreting the results, especially in areas where the SOC has increased. As Rondônia is located in a region with low climatic amplitude, the adjusted model made predictions in scenarios beyond its training range.

Keywords: Rondônia, machine learning, tropical soil, recursive feature elimination

1. Introduction

Soil is the largest carbon reservoir in the terrestrial biosphere and contains approximately 1,500 Pg C up to one meter deep (Batjes, 1996; Scharlemann et al., 2014), which is twice the stored in the atmosphere or three times in vegetation (Spawn et al., 2020). Consequently, small changes in this stock can significantly alter the concentration of greenhouse gases in the atmosphere and intensify climate change (García-Palacios et al., 2021).

Understanding the relationship between soil organic carbon (SOC) and its environmental drivers contributes to reducing uncertainties in predicting the impacts of climate change and land use change for large-scale climate feedback projections on carbon (Gonçalves et al., 2021). Climate is considered one of the main drivers of SOC storage, and factors such as precipitation (Carvalhais et al., 2014), soil microbial activity (Hararuk et al., 2015), and temperature (Giardina et al., 2014) are regulators of the soil's capacity to store or release carbon. SOC is associated with the quality of water and soil, which has consequences to produce agricultural goods, including food. Therefore, in addition to the relationship with climate, it is important to understand the spatial-temporal dynamics of SOC at regional and global scales (Li et al., 2021; Ramesh et al., 2019), and mathematical models can offer great contributions.

Earth system models (Luo et al., 2016) are commonly used to project global carbon balances for the climate and simulate future scenarios for the SOC. Despite the importance of these models in understanding SOC distribution, several environmental and soil characteristics factors that are determinant in SOC dynamics are not adequately represented in the models (Todd-Brown et al., 2013). This generates great uncertainties in the future prediction of SOC stock, in addition to inadequately representing the spatial heterogeneity of this carbon, especially in smaller territorial extensions.

One way to reduce these uncertainties is with the digital soil mapping, applying machine learning techniques with field observations and environmental variables suited to the regional reality. In these techniques, a predictive model is built to represent the relationships between a

set of environmental predictors and the soil attribute such as SOC, and from there, make predictions for non-sampled locations (Fathizad et al., 2020; McBratney et al., 2003).

In the Brazilian Amazon, there are projections of SOC using machine learning. For instance, Gomes et al. (2019) and Vasques et al. (2017) performed a prediction of SOC stock for Brazil, while Hengl et al. (2017) the entire earth. All these authors used machine learning techniques, and the maps produced included the Amazon biome. However, the approach was only spatial and not temporal, with reduced point density in the Amazon area (which may be a limiting factor in this technique) and at coarse spatial resolutions.

Rondônia is one of the states inserted in the Brazilian Amazon that contains the most detailed field survey in the region, with almost three thousand profiles surveyed, including SOC contents (Cochrane and Cochrane, 2006). In 1996/1997, a soil survey was carried out in Rondônia and a detailed study of the region's geology, geomorphology, and fauna.

Rondônia contained one of the lowest population densities in Brazil and most of its native vegetation intact in the 1960s (IBGE, 1981). At that time, successive colonization programs were implemented by the Brazilian state to occupy the region and reduce the agrarian conflict that took place in southeastern Brazil (Goza, 1994; Millikan, 1992). In the mid-1980s and 1990s, the paving of BR364, the main access road to the state, was completed and the POLONOROESTE colonization program was implemented, which dramatically increased the number of immigrants and, consequently, deforestation with rampant occupation territory (Millikan, 1992).

Although the program's guidelines focused on establishing perennial crops, the state was mainly taken over by pastures in cycles followed by deforestation, the establishment of pasture and subsequent abandonment after 6 to 8 years of extensive cattle ranching. In addition, other uses such as mining and logging were also established in the region (Dale et al., 1994; Pedlowski et al., 1997).

The surveys carried out in 1996/1997 were aimed at halting the destruction of the forest and contributing to the sustainable management of land. In just over three decades, the state has lost more than 6.5 million hectares of natural vegetation (Souza et al., 2020) and future modeling indicates that by 2050 it could lose another 3.5 million hectares (Piontekowski et al., 2019). This, together with global climate change, significantly reduced rainfall in the state (Khanna et al., 2017) and this situation may worsen even further, since the region's climate is monsoon, highly dependent on convective rains (Butt et al., 2011). All these factors directly impact the SOC dynamics.

Predicting the SOC stock over time using a smaller spatial scale contributes to the assessment of the current SOC storage potential and its temporal changes considering climate change scenarios (Wang et al., 2017). The objective of this research was to model and map the SOC stock in the state of Rondônia in future scenarios up to 2100. Specifically, we aimed to: i) select the best machine learning model to predict SOC; ii) evaluate the main predictors that control SOC in Rondônia; iii) prepare average SOC stock maps, upper quartile (Q95), lower quartile (Q05) and coefficient of variation for two contrasting socioeconomic scenarios (one where the growth and equality scenario is focused on sustainability - ssp126 and the other on growth fast and unrestricted in economic production and energy use - ssp585) under two different climate models; iv) stratify SOC stock in depth, by climate model and in two shared socioeconomic scenarios (ssp126 and ssp585).

2. Material and methods

2.1. Study area and soil profiles

The state of Rondônia is located southwest of the Brazilian Amazon and has an approximate area of 225,076.83 km^2 (Pereira and Goncalves, 2021). The predominant vegetation is terra firme forest. The average annual precipitation is 2,250 mm with a well-defined dry season, between August and October, the average temperature is 26 °C and the relative humidity is approximately 85%, which makes the monthly potential evapotranspiration very constant (Alvares et al., 2013).

The geology of the state is covered by the southwestern portion of the Amazonian Craton with an approximate age of 1.78 Ga. The lithology is composed of substrates resulting from successive magmatisms, metamorphisms, sedimentations and deformations that reached different rock materials and mineral deposits reworked by more recent orogenetic events in the Amazonian Craton (Adamy, 2010). The rocks have different ages and have gone through prolonged processes of generalized flattening of the relief. The terrain is predominantly gently wavy and wavy, ranging in altitude from 33 to 1,100 m (NASA JPL, 2020). According to Ab'Saber (1969, 1967), almost all the state's territory falls within the morphoclimatic domain of the equatorial lowlands of the Amazon.

The dominant soils in the state are the Latossolos, Gleissolos and Neossolos occupying 60%, 11% and 10% of the area respectively (Cochrane and Cochrane, 2006). Latossolos occur mainly in plateaus, flat surfaces in southern Amazonia, depressions and the Parecis plateau. The Gleissolos are found in the fluvial plains of the Madeira - Mamoré and Guaporé rivers and the depressions of the Guaporé river. The Neossolos range from river plains to mountainous

areas, such as the river plains of the Madeira - Mamoré - Guaporé rivers, passing through the depressions and plateaus of the state to the Parecis plateau (Adamy, 2010; Cochrane and Cochrane, 2006).

For the implementation of the Territorial Plan for Sustainable Rural Development, the federal government divided the state into seven territories (Fig. 1) with similar economic and environmental characteristics and social, cultural, and geographic cohesion, to facilitate the planning of actions for the sustainable development of the state (BRASIL, 2006; MDA, 2005).

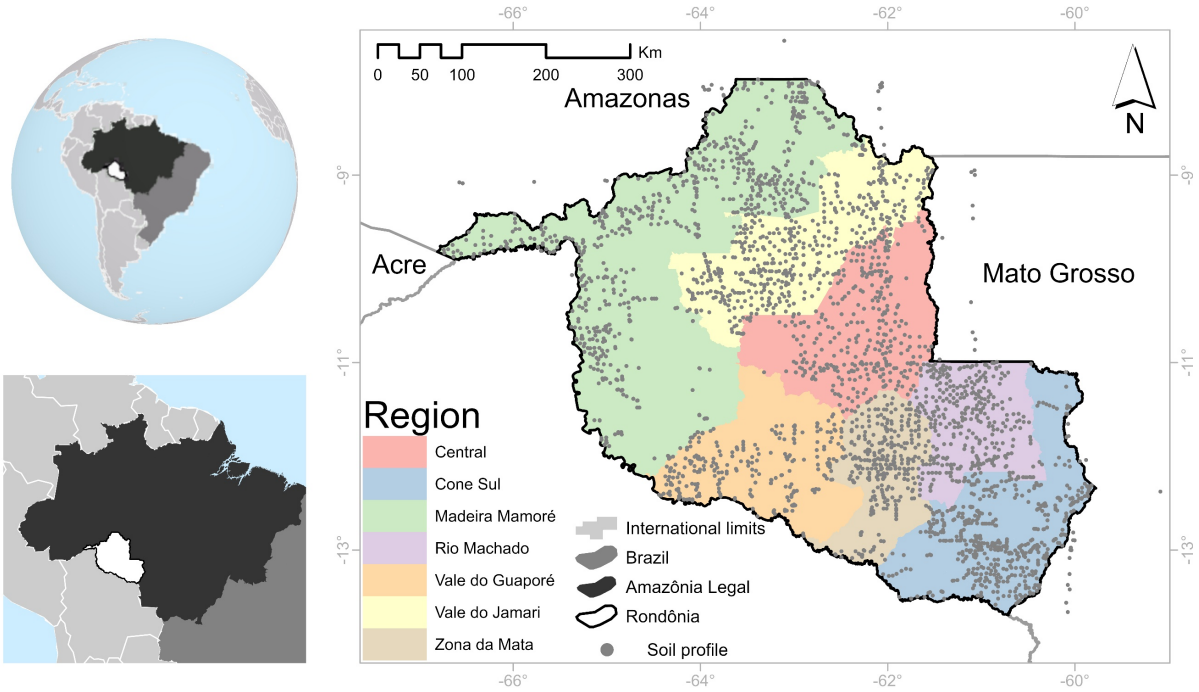

Fig. 14. Distribution of soil profiles in different regions of Rondônia State, Brazil.

The state of Rondônia has experienced a marked evolution in deforestation in the last decade. The area deforested in the state rose from 0.18% in 2010 to 4.62% in 2020, while in the other states of the Amazon deforestation ranged from 0.13% to 1.56% in the same period (Assis et al., 2019). This is a consequence of the national strategy of occupation and regional development, as the colonization projects initiated by the Brazilian State in the 1970s culminated in the unbridled occupation of the territory (Ciccantell et al., 1994; Moran, 1981) and the unsustainable use of land.

Land uses in the state mainly include extensive livestock farming with some other less representative agricultural activities, such as perennials such as coffee and annuals such as soybeans, corn, and rice (IBGE, 2019). A large part of the pastures originated from the direct conversion of primary forests (Numata et al., 2007) and currently, the state still undergoes an intense conversion of the native land cover.

For the research, data from 2,914 soil profiles distributed across the seven territories of the state of Rondônia were used (Fig. 1). These profiles were collected and analyzed between 1996 and 1997 to study the second approximation of the Socioeconomic Ecological Zoning of Rondônia (ZEERO) (Cochrane and Cochrane, 2006; SEDAM, 2021). From the physical, chemical, and biological data available in the database, the soil organic carbon content (SOC) was used, quantified by the Walkley-Black method (Nelson and Sommers, 1982) and the soil bulk density (BD) that was determined by the standard volumetric ring method. BD data were not determined for all profiles and given its importance for calculating the SOC stock (equation 1), the missing BD was estimated through machine learning, using the other physicochemical data available for all profiles (Table S10 and Table S11, Supplementary material). SOC stock were harmonized at depths 0-5, 5-15, 15-30, 30-60 and 60-100 cm using splines from the "GSIF" package (Hengl, 2020) of the R software (R Core Team, 2021).

$$\text{SOC stock} = \frac{\text{SOC} \times \text{BD} \times \text{e}}{10} \quad (1)$$

Where: SOC stock = Stock of soil organic carbon (Mg ha^{-1}); SOC = Organic carbon content in the soil (g kg^{-1}); BD = Density of the soil (g cm^{-3}); e = Thickness of the layer considered (cm).

2.2. Environmental predictors

Soil classes and lithology were derived from the ZEERO survey (Cochrane and Cochrane, 2006; SEDAM, 2021). For soil classes, up to the third categorical level was used, grouping rare soil classes for sampling. Average historical climatic variables of 30 years (1970-2000) were obtained, containing precipitation, temperature, and bioclimatic data with a resolution of approximately 1 km (Fick and Hijmans, 2017). To represent the native vegetation, satellite image data were used for the first year of imaging (1984), calculating the annual average of the enhanced vegetation index (EVI), modified soil adjusted vegetation index (MSAVI), and normalized burn ratio (NBR).), normalized burn ratio 2 (NBR2), normalized difference moisture index (NDMI), normalized difference vegetation index (NDVI) and soil adjusted vegetation index (SAVI) on the Google Earth Engine platform from images from the "USGS Landsat 5 Surface Reflectance" collection Tier 1" and with a resolution of approximately 30 m (Gorelick et al., 2017).

The digital elevation model (DEM) from NASADEM (NASA JPL, 2020) with an approximate resolution of 30 m was used, from which 48 morphometric variables were derived. All predictors are described in Table S12, Supplementary material. All data were redesigned for the South America Lambert Conformal Conic datum (ESRI:102015) through the "raster"

package (Hijmans, 2021) and harmonized for the 200 m resolution using the “cubicspline” function of the “gdalUtilities” package (O’Brien, 2021).

2.3. Data pre-processing, model selection and adjustment

The environmental predictors were submitted to the variable selection process to obtain a more parsimonious model. The process consisted of two sequential methods, in the first one, a Spearman correlation matrix was generated, and one predictor was eliminated from each pair of highly correlated predictors ($r > |0.95|$) avoiding redundancy. The remaining predictors were selected by importance, submitting them to the interactive recursive feature elimination (RFE) algorithm (Kuhn and Johnson, 2013). For the subsets applied to the RFE, a discrete range of predictors was used, which was from 2 to 40 predictors and then from 5 to 5 until the total of predictors. The quantitative predictors that achieved the best performance with the smallest subset in the RFE were adopted for each execution in the modeling process.

Given the high diversity of machine learning models, there are specific characteristics in the data analysis processes, resulting in different results depending on the study condition. Therefore, seven different machine learning algorithms were evaluated: i) cubist (Kuhn and Quinlan, 2021); ii) multivariate adaptive regression spline - earth (Milborrow, 2020); iii) stochastic gradient boosting - gbm (Greenwell et al., 2020); iv) elastic-net regularized generalized linear models - glmnet (Friedman, 2002); v) k-nearest neighbors - k-knn (Schliep and Hechenbichler, 2016); vi) random forest - rf (Liaw and Wiener, 2002); and vii) support vector machines radial kernel - svmRadial (Karatzoglou et al., 2004).

For the evaluation of the models, the database with 2,914 profiles was divided into 75% for training and cross-validation and 25% for testing. The training data was used to adjust the model with cross-validation of 10 folds and the test set was used to assess the predictive performance. The model with the best performance in the metrics of mean absolute error - MAE (1), coefficient of determination - R^2 (2) and root mean squared error - RMSE (3) in the test set was selected.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|P_i - O_i| \quad (1)$$

$$R^2 = \frac{\sum_{i=1}^{n}(P_i - \bar{O}_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O}_i)^2} \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \quad (3)$$

Where: n represents the number of samples; P_i and O_i represent the predicted and observed values at location i, respectively.

For a better understanding of the model and the main SOC proxies in Rondônia, the importance of predictors based on their contributions was evaluated. The metrics were mean min depth, node purity increase, mse increase and p-value. In addition, the interaction between predictors was evaluated by the frequency of occurrence in decision trees sequentially and the minimum average depth of the tree of these occurrences. As 100 models were built for each depth, these metrics corresponded to the average values of these models, calculated using the "randomForestExplainer" package (Paluszynska et al., 2020).

2.4. Climate changes scenarios

Climate change scenarios were evaluated over four periods, 2021 - 2040, 2041 - 2060, 2061 - 2080 and 2081 - 2100, in two contrasting scenarios, one where the scenario of growth and equality focused on sustainability (ssp126) and the other on rapid and unrestricted growth in economic production and energy use (ssp585), both from CMIP6, under two models with different climate sensitivities, the CanESM5 and the MIROC6.

The scenarios used result from the combination of "shared socioeconomic pathways (SSPs)" and representative concentration pathways (RCPs) (Fig. 15). The SSPs set the baseline on which the world will deal with climate change, such as population, technological and economic advancement. RCPs represent levels of greenhouse gases and other future radiative forcings up to 2100 (watts per square meter).

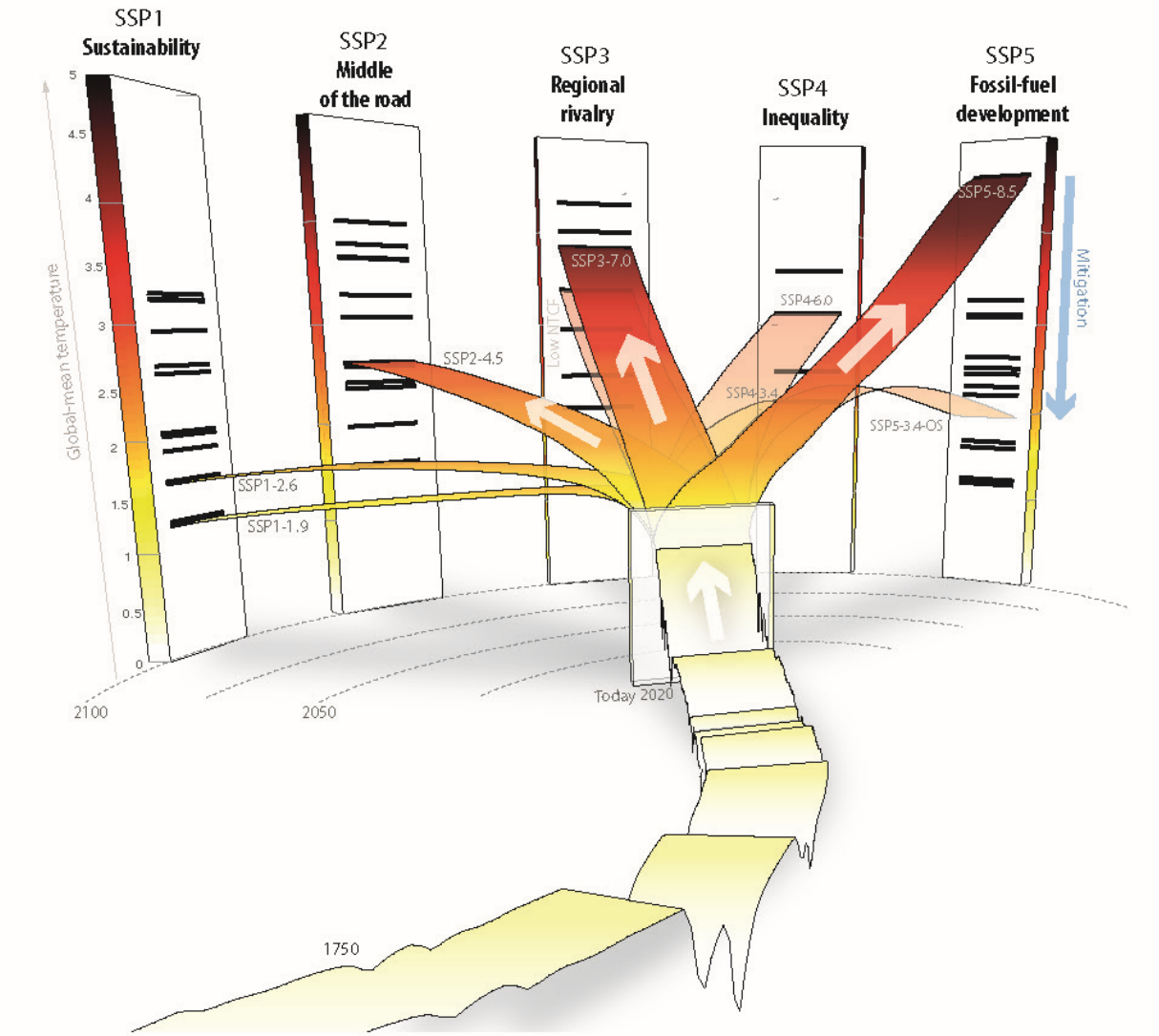

Fig. 15. All shared socioeconomic pathways (SSPs) and representative concentration pathways (RCPs) scenarios used by the IPCC in global climate change and as temperature increase trajectories to 2100 compared to a pre-industrial baseline.
Fonte: Meinshausen et al. (2020)

The ssp126 is composed of the SSP1 which represents a sustainability scenario and the 26 which is the RCP. The world gradually and in an integrated way shifts towards sustainability, with a more inclusive development that respects nature. Investments in education and health aid the demographic transition and the focus on economic expansion shifts to broad attention to human well-being. Consumption is directed towards low material growth and less intense use of natural goods and energy. The “26” of the ssp126 is the scenario CPR, simulating 2.6 watts m^{-2} (Riahi et al., 2017; Wayne, 2013).

The ssp585 is constituted by the SSP5 which represents a scenario-driven by fossil fuel-based development and the “85” represents the RCP. In SSP5 the world develops focused on competitive markets, innovation and society advancing rapidly in technology and capital.

Integration of global markets focus on investments in health and education aimed at economic and social development linked to the abundant exploitation of fossil fuels and the adoption of lifestyles with intensive use of natural goods and energy. The “85” of the scenario simulates 8.5 watts m^{-2} of CPR (Riahi et al., 2017; Wayne, 2013).

It is believed that the increase in greenhouse gases, mainly CO_2, increases the temperature and alters the precipitation (Eyring et al., 2016). In summary, the model based on the relationship between SOC stock and historical environmental predictors was adjusted, and to predict future maps, climate predictors were replaced by future data, while the other predictors remained constant.

The modeling process (training, validation, testing and prediction) was performed 100 runs, under different subsets of samples, which allowed evaluating the uncertainties in the prediction. From the 100 maps generated for each depth, climate model and scenario, an average SOC stock map was generated, an average map of the two climate models in the ssp126 scenario and another for the ssp585 scenario, all at a depth of 0-100 cm. For the same configuration, the maps of quartiles at 5% (Q05) and 95% (Q95) were generated, and the uncertainty map was generated from the coefficient of variation of each pixel.

3. Results

3.1. Performance of models and importance of predictors

The random forest (rf) model presented the best average performance and, therefore, was chosen for the prediction of maps (Fig. 16). The mean R^2 of the rf ranged from 0.17 to 0.25 between the depths, with a decreasing trend while the depth increased (0-5 cm > 5-15 cm > 15-30 cm > 60-100 cm > 30- 60 cm), except for depths of 30-60 cm and 60-100 cm, where the behavior was reversed. Analogously to R^2, the errors (MAE and RMSE) showed the same trend, ranging from 4.1 to 9.6 Mg C ha^{-1} for MAE and 6.3 to 14.2 Mg C ha^{-1} for the RMSE.

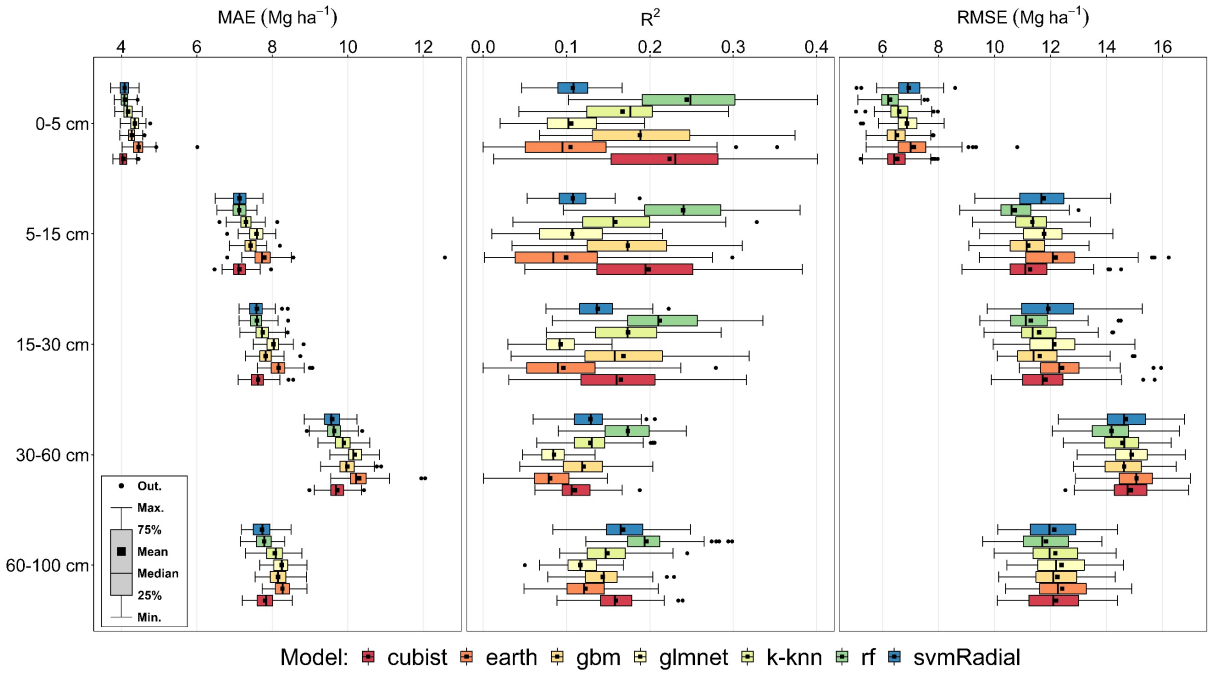

Fig. 16. Performance of different algorithms in the test set of soil organic carbon stock at different depths, in Rondônia State, Brazil. MAE = Mean absolute error; R^2 = Coefficient of determination; RMSE = Root mean squared error.

In the process of selecting predictors by correlation, 65 were selected for depths of 0-5 and 5-15 cm; 66 for depths of 15-30 and 30-60 cm; and 64 for a depth of 60-100 cm. The sets of predictors submitted to RFE resulted in different quantitative subsets, with approximate performance stabilization between 12 and 16 predictors (Figure S6, Figure S7, Figure S8, Figure S9 and Figure S10, Supplementary material).

The minimum depth of the predictors indicates the longest path from the root to the end of the tree (Fig. 17). It is observed that the soil class (Fig. 17a-d), temperature (Fig. 17a-e), precipitation (Fig. 17a-e) and relief (Fig. 17e) were the most decisive predictors for the determination of SOC stock in Rondônia, dividing the trees from its beginning (< that the third tree), until close to 500 trees.

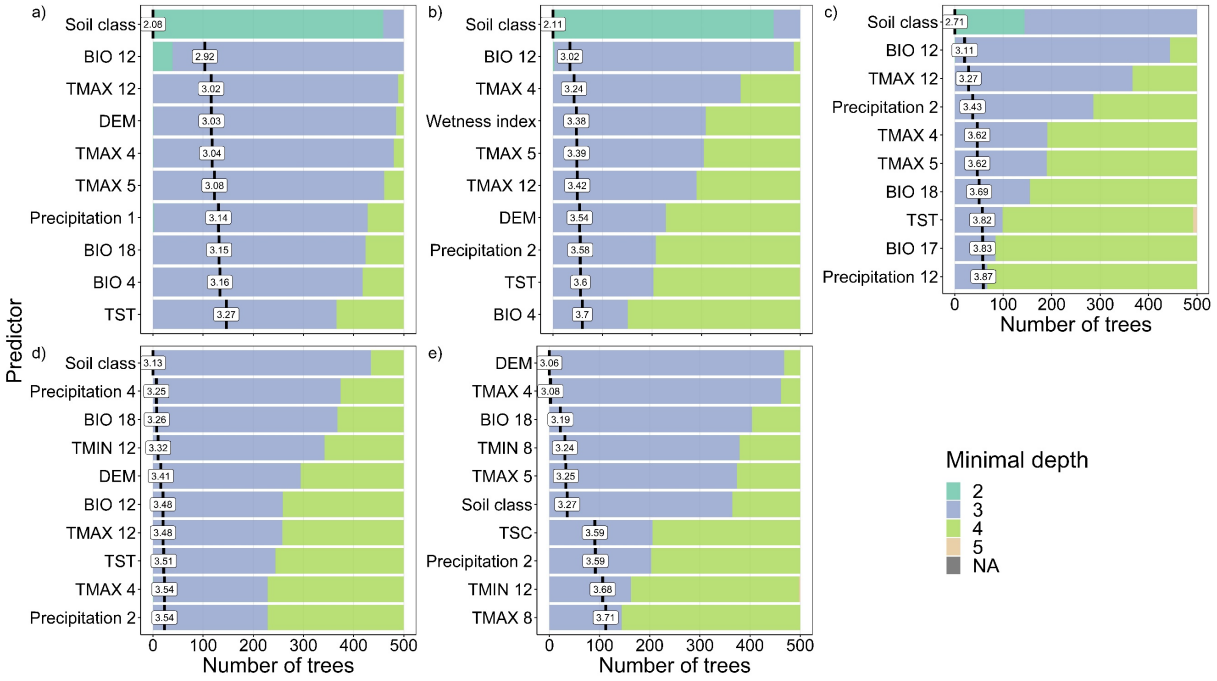

Fig. 17. The 10 most important predictors at each soil depth according to the mean of the minimum depth distribution of the random forest model decision trees adjusted for SOC stock at depths of 0-5 cm (a), 5-15 cm (b), 15-30 cm (c), 30-60 cm (d) and 60-100 cm (e) of the soil. The minimum depth indicates the first node of the random forest decision tree in which a given predictor best grouped (minimized the variance) the data set, Rondônia State, Brazil.

How the metrics that assess the importance of a predictor in rf are related helps to choose less correlated metrics, to better explore the importance of the predictor and from different perspectives. This helps to understand how the SOC behaves in face of changes in predictor values, which increases the inference capacity in the model's behavior (Figure S11, Figure S12, Figure S13, Figure S14 and Figure S15, Supplementary material).

The metrics selected for the construction of the predictors' multiple importance graph were the node purity increase and the mse increase with ranking of the most important predictors considering the p-value of the frequency test in which the predictor is used for tree division. In addition to the average minimum depth, the predictors related to precipitation (BIO 12, Precipitation 12, Precipitation 1, Precipitation 2, Precipitation 4, and BIO 18), temperature (BIO 4, TMAX 4, TMAX 5, TMIN 8 and TMIN 12) and relief (DEM and Wetness index) stood out among the five most important in all soil depths (Fig. 18). Importantly, the soil class predictor, which is one of the most important predictors of minimum root depth (Fig. 17), does not appear in this assessment due to its non-numeric nature.

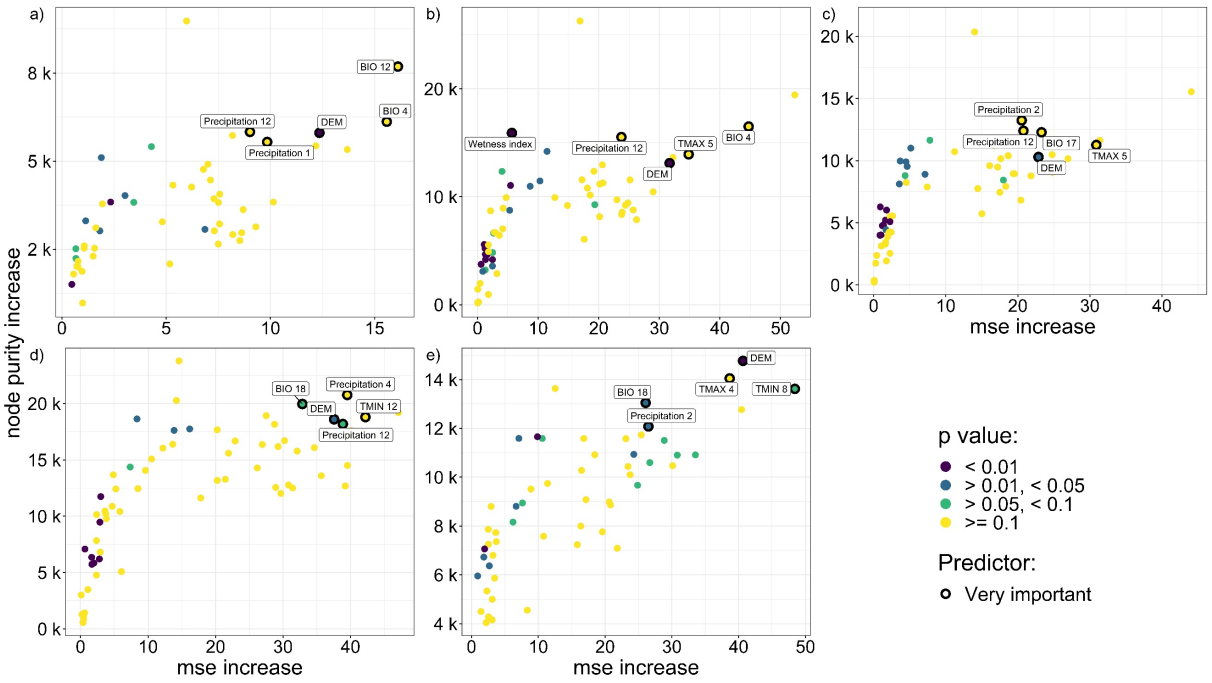

Fig. 18. Multiple importance of predictors measured by the mean increase in the mean squared error after a given predictor is swapped (mse increase) as a function of the mean increase in tree node purity by divisions in a given predictor, measured by the decrease in the sum of squares (node purity increase), with the p-value based on a binomial distribution of the number of nodes divided in the predictor assuming that the predictors are randomly distributed to form divisions (ie, if a variable is significant, it means that the predictor is used to divide with more often than the theoretical selection number if the case were only by randomness). For the random forest model adjusted for SOC stock at depths of 0-5 cm (a), 5-15 cm (b), 15-30 cm (c), 30-60 cm (d) and 60-100 cm (e) of the soil, Rondônia State, Brazil.

The importance of climate predictors increases in SOC stock when they are sequentially associated in the model's decision trees (Figure S16, Figure S17 and Figure S18, Supplementary material), mainly in the superficial layers (0-5, 5-15 and 15-30 cm). Precipitation followed by temperature, for example, reaches depths of less than 1.5, with occurrences in more than 450 trees in the model. In deeper soil layers (30-60 and 60-100 cm) predictors derived from relief, such as SSR 2 (Direct to Diffuse Ratio) (Figure S19, Supplementary material) and soil classes (Figure S20, Supplementary material) followed by climate predictors, they are important controllers of SOC stock, occurring in more than 340 of the 500 trees.

3.2. Current SOC stock and future forecasts

The current SOC map (Fig. 19a) shows that the central territory, north of the Guaporé Valley and Zona da Mata, contains the smallest SOC stock (< 100 Mg C ha^{-1}). The largest SOC stock observed range from the southwest of the Guaporé Valley to the south of the Zona da

Mata, with some patches in the southeastern Madeira Mamoré and eastern Southern Cone (> 180 Mg C ha^{-1}).

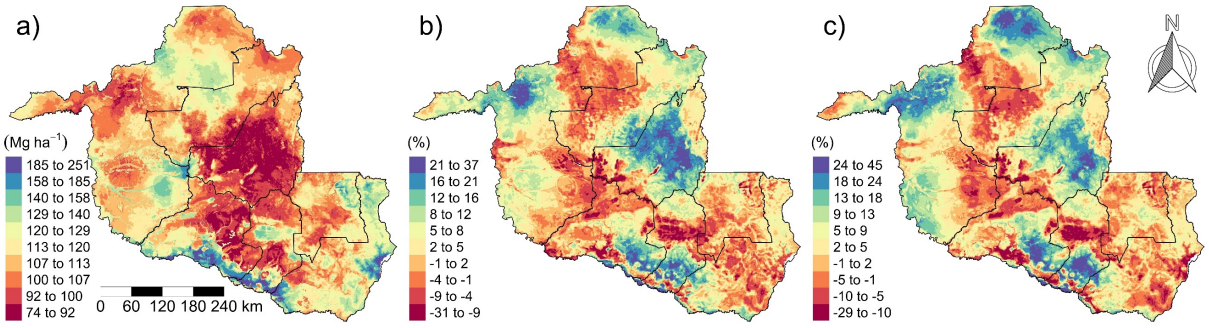

Fig. 19. Average of 100 SOC stock maps in 100 cm for the present (a) and the difference between the average prediction of the CanESM5 and MIROC6 models of the SOC stock for 2100 in scenarios ssp126 (b) and ssp585 (c) in relation to the present ([SOC 2100 - SOC present] / SOC present * 100), Rondônia State, Brazil. The black lines represent the borders of the territories described in Fig. 1.

The predictions for the ssp126 and ssp585 scenarios pointed to different intensities of changes in SOC stock, both for C increase and for reduction (Fig. 19). In most of the territories that contain the largest stock, these will be reduced in the order of 4 to 30% in both scenarios, and in the northwest of the Madeira Mamoré territory the reductions may be more severe in the ssp585 scenario compared to the ssp126. These areas will experience significant temperature rise (Figure S21, Supplementary material), however, the precipitation reduction will be less pronounced (Figure S22, Supplementary material). The north, central and northwest regions of the state indicate SOC gains in both scenarios. Gain magnitudes can range from 37% in ssp126 scenario to 45% in ssp585, and in ssp126 scenario the changes may be in larger areas compared to ssp585. This contrast is especially pronounced in the central territory of the state. According to climate models, these areas of SOC increase will suffer a marked reduction in precipitation - BIO 12 (Figure S17, Supplementary material), a significant increase in both maximum temperature - TMAX 4 (Figure S16, Supplementary material) and in temperature seasonality - BIO 4 (Figure S23, Supplementary material). These three predictors were considered very important for rf (Fig. 18).

In the quartiles (Q05 and Q95) the spatial behavior was similar to the mean maps, varying in magnitude for the two evaluated scenarios (Fig. 20 and Fig. 21). Q05 expressed with greater intensity the SOC gains when compared to Q95. A considerable extension of Q05 indicated a gain in SOC, in Q95 they indicated no change in inventories or less expressive gains, especially in the ssp585 scenario.

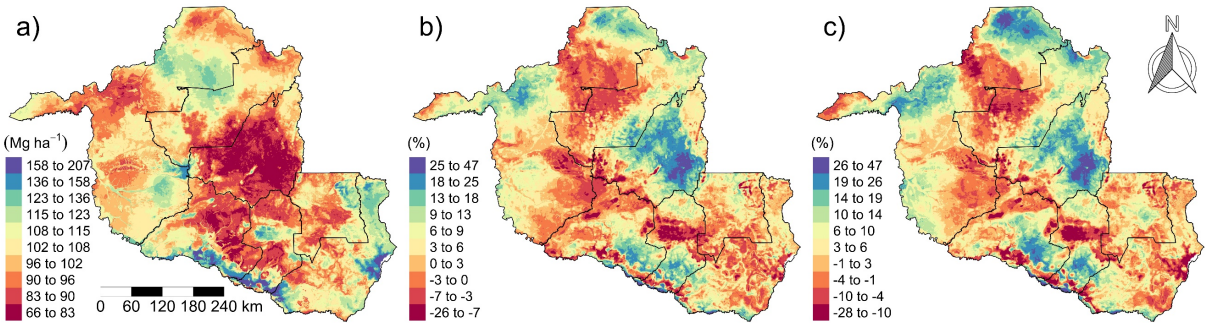

Fig. 20. Lower range (Q05) of 100 SOC stock maps in 100 cm for the present (a) and the difference between the average Q05 prediction of the CanESM5 and MIROC6 models of the SOC stock for 2100 in scenarios ssp126 (b) and ssp585 (c) in relation to the present ([SOC 2100 - SOC present] / SOC present * 100), Rondônia State, Brazil. The black lines represent the borders of the territories described in Fig. 1.

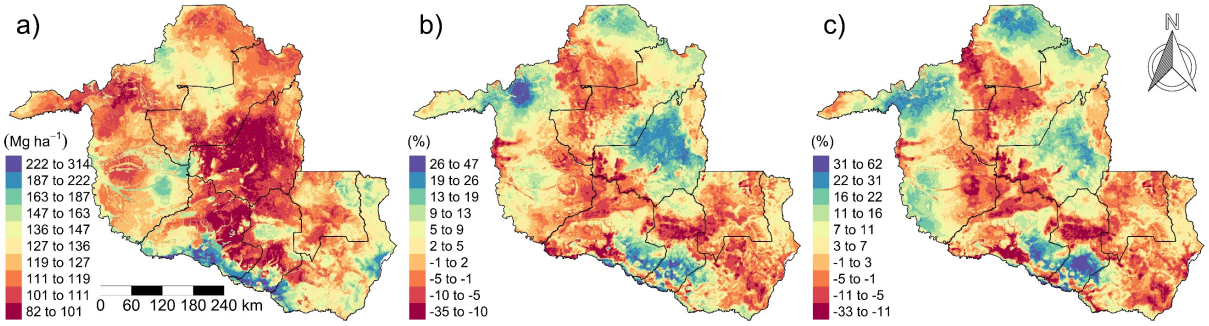

Fig. 21. Upper range (Q95) of 100 SOC stock maps in 100 cm for the present (a) and the difference between the average Q95 prediction of the CanESM5 and MIROC6 models of the SOC stock for 2100 in scenarios ssp126 (b) and ssp585 (c) in relation to the present ([SOC 2100 - SOC present] / SOC present * 100), Rondônia State, Brazil. The black lines represent the borders of the territories described in Fig. 1.

Forecast uncertainties ranged between 3 and 22% (Fig. 22) for both present SOC stock and future forecasts. The spatial behavior of uncertainties was also similar between the scenarios, with the greatest uncertainties in the southern regions of the state (> 14%), mainly south of the Zona da Mata. Most uncertainties were below 10% variation in predictions.

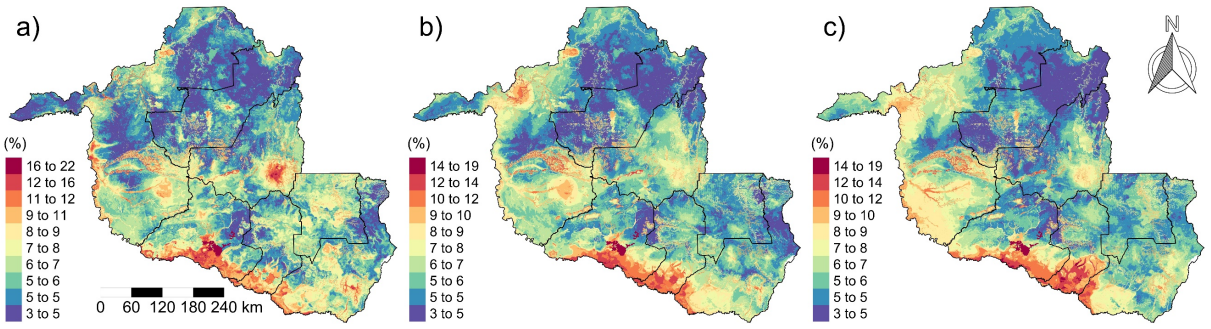

Fig. 22. Coefficient of variation of the 100 maps in 100 cm depth for the present (a) and the difference between the average prediction of the models CanESM5 and MIROC6 of the SOC stock for 2100 in the scenarios ssp126 (b) and ssp585 (c) in relation to the present ([SOC 2100 - SOC present] / SOC present * 100), Rondônia State, Brazil. The black lines represent the borders of the territories described in Fig. 1.

In general, mean, and median values varied little between periods and between models, except for the difference between the present and future projections, where the behavior changed in the distribution of SOC stock at all depths. The two climate models in the ssp126 scenario expressed the greatest variations in the highest SOC values (Fig. 23). At 100 cm depth, the CanESM5 model shows a more pronounced reduction in the largest SOC stock, when compared to the MIROC6 with advancing periods (2021 to 2100). Except for the 5-15 cm depth for the period 2021-2040, the CanESM5 remains below the MIROC6 at the highest SOC values at all depths and periods.

In the ssp585 scenario (Fig. 24), the behavior between the models was analogous to ssp126, the current maximum SOC values were higher than the maximum values of future forecasts, with CanESM5 with lower estimates than MIROC6. However, it was observed that the maximum values of the present in relation to the future periods significantly reduced in the deeper layers (30-60 cm and 60-100 cm) than in the superficial layers (0-5 cm, 5-15 cm, and 15-30 cm).

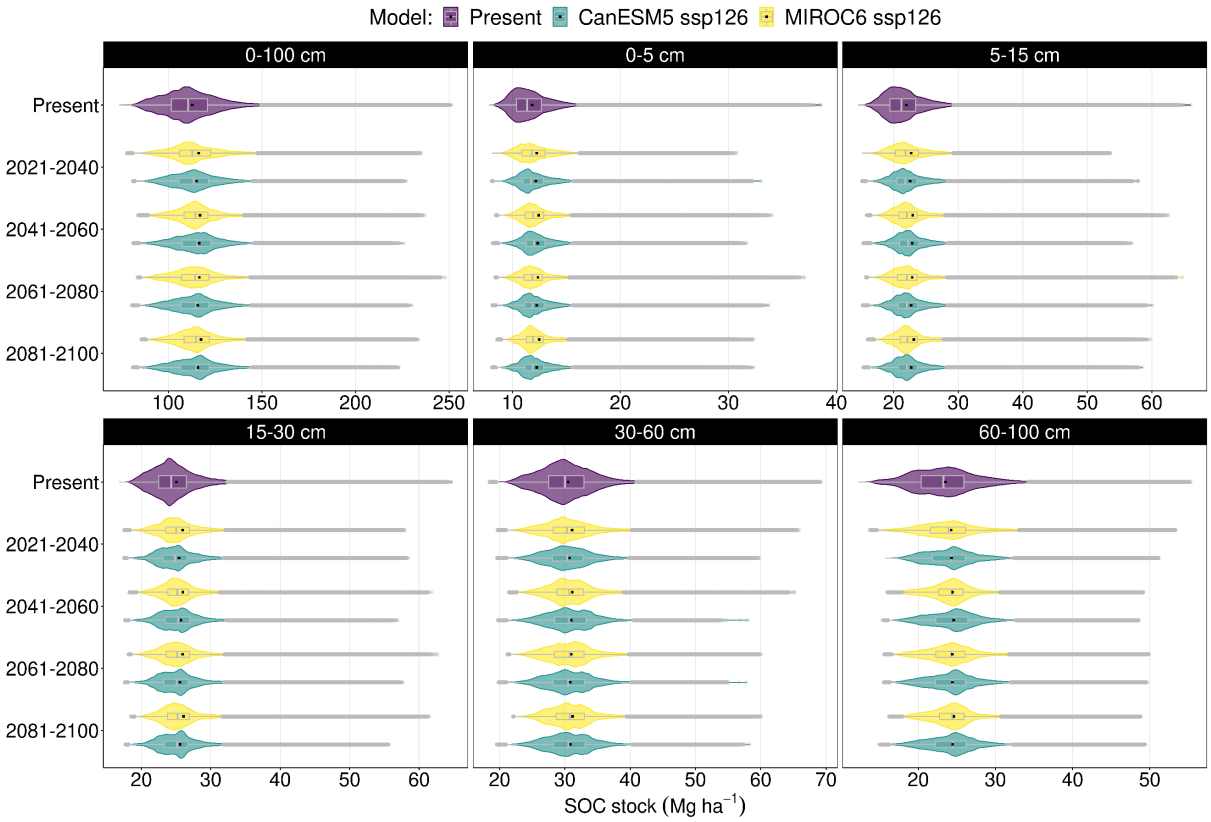

Fig. 23. Soil organic carbon (SOC) stock Rondônia State, Brazil, for the present and forecasts for four different periods between 2020 and 2100 under two climate models in the CMIP6 ssp126 scenario at different depths.

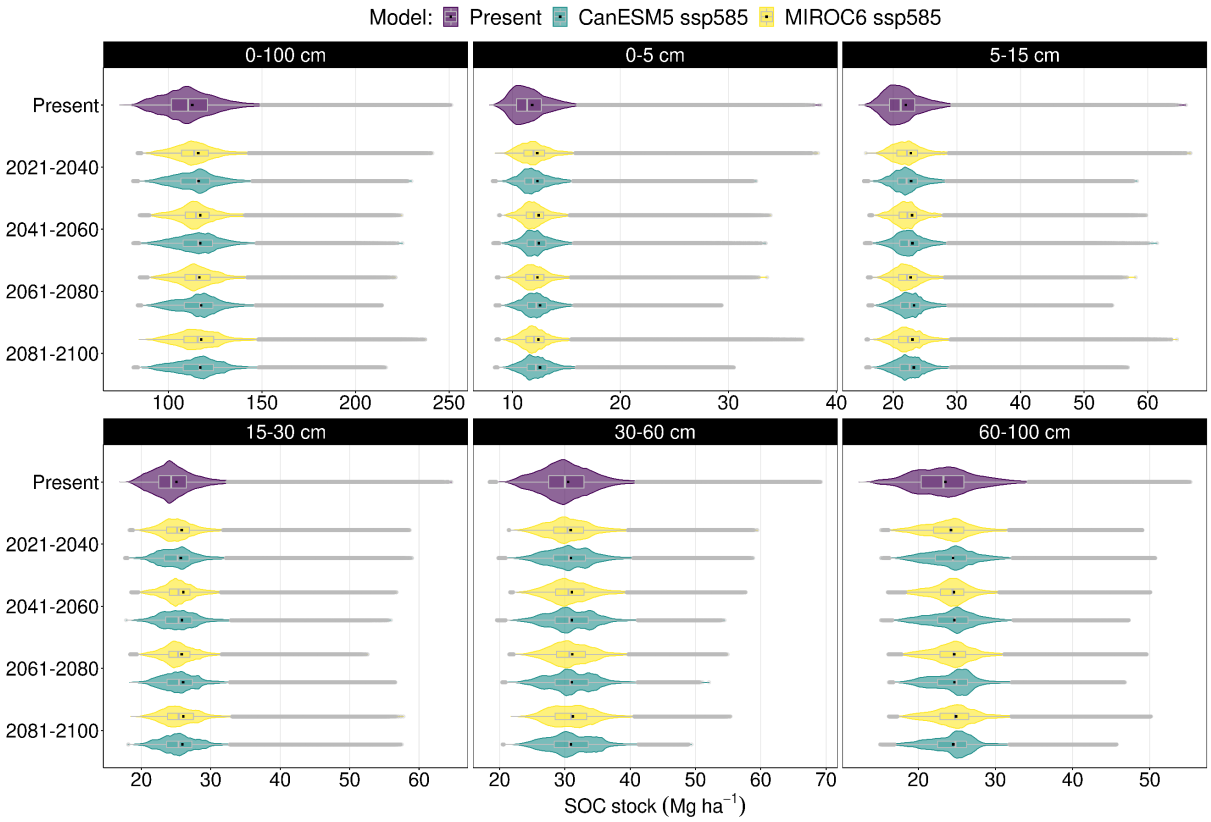

Fig. 24. Soil organic carbon (SOC) stock Rondônia State, Brazil, for the present and forecasts for four different periods between 2020 and 2100 under two climate models in the CMIP6 ssp585 scenario at different depths.

4. Discussion

4.1. Performance of models and importance of predictors

The superiority of the performance observed in the rf compared to the other models is explained by the characteristics of this model, which randomly selects predictors to split the nodes. In the case of frequently used predictors, randomness can minimize the model's bias and increase its generalizability (Breiman, 2001). Therefore, the rf may present results more consistent with the SOC surveyed in the field, as the relationship between predictors and the SOC is complex (Heung et al., 2016). This superior generalization of rf compared to other machine learning models is confirmed by other authors who studied SOC stock from environmental predictors (Sodango et al., 2021; Wang et al., 2021). In contrast to this, other studies have found better performance in other decision tree-based models compared to rf (Ahirwal et al., 2021; Goydaragh et al., 2021), this shows how models can vary depending on the analysis situation, demonstrating the importance in the selection of models.

Soil class, relief and climate were important predictors of SOC stock in Rondônia. The soil class mainly describes texture and structure, important properties in soil carbon dynamics. Soils with larger SOC stock, for example, tend to have a fine texture and good structure

(Soucémarianadin et al., 2018), which may justify the importance of the predictor for deeper division of decision trees in predicting SOC stock in Rondônia.

The soil class was the precursor predictor of tree divisions in the superficial layers of the soil, while in the deeper layers, and in a less pronounced way, the predictor was the relief, in both cases the climate was the sequential definer and with high occurrence within the decision trees. This suggests that in deeper layers the position in the landscape may influence SOC stock more than soil classes, but in all cases, climate is the main SOC driver.

Climate and relief are closely related, the milder climate for example tends to be found higher up in the landscape (Klinges and Scheffers, 2021). These combinations are capable of intensifying or attenuating the SOC dynamics, changing the quantitative ones and explaining the importance of these predictors in the model. The strong influence of climate on SOC stock in Rondônia suggests the sensitivity of this SOC to future climate changes (Gonçalves et al., 2021).

4.2. Current SOC stock and future forecasts

In the regions with the lowest SOC stock, Latossolos that suffered strong anthropogenic action predominated, with change in land cover and conversion from natural forest to extensive cattle raising. Such changes reduce SOC stock (Durrer et al., 2021). The areas that contain the largest stock are located in places with a predominance of Gleissolos and Organossolos, with seasonal flooding characteristics. The environments in which such soils occur favor the accumulation of SOC due to the absence of O_2, which limits the decomposition of carbon by the soil biota (Hinson et al., 2017).

The regions with higher altitudes (southeast of Madeira Mamoré and Southern Cone) will suffer a significant reduction in SOC stock in future scenarios, mainly in ssp585. This prediction is due to the rise in temperature (Figure S21, Supplementary material) and water availability (Figure S22, Supplementary material), which consequently increase soil biota respiration and transform soil carbon into CO_2 (Giardina et al., 2014). According to the report of the Intergovernmental Panel on Climate Change (IPCC), the tropics will suffer intense heat waves, mainly in the Amazon, intensifying forest fires and aggravating CO_2 emissions by reducing SOC stock (Jia et al., 2019).

Climate estimates indicate that in the tropics there will be continuous warming and the result will be new climates. This new configuration will place ecosystems exposed to extreme temperatures and rainfall that will change their structure and functioning (Olsson et al., 2019). With this, even the wetlands of Gleissolos and Organossolos that contain the largest SOC stock

may change due to the increase in the concentration of CO_2, changes in precipitation, forest fires, increase in temperature and methane emission. These direct and indirect interactions are complex and little known, which makes projections difficult (Gonçalves et al., 2021).

Projections for the northern, central and northwestern regions of the state, where SOC stock is limited (< 90 Mg ha^{-1}), indicate an increase in the SOC for both scenarios (ssp126 and ssp585). In these areas, despite climate projections pointing to an increase in temperature, the reduction in precipitation will be more intense (-200 mm in the ssp126 scenario and reaching -300 mm in the ssp585 scenario). The water limitation can reduce the microbial activity and justify the model's tendency to present SOC accumulation even with the temperature increase.

There are discussions about the neutralization of carbon emission due to the fertilization effect of CO_2 (eCO_2) on plants. The increase in atmospheric carbon dioxide can have a positive effect on plants by increasing their carbon uptake and boosting greater photosynthetic activity, which can consequently increase the primary productivity of vegetation (Jiang et al., 2020; Walker et al., 2019). This phenomenon can generate negative feedback for climate changes, reducing atmospheric CO_2 concentration (Arora et al., 2013) and producing more organic carbon for soil storage.

Estimates show that in some tropical forest places primary vegetation production can increase and attenuate the depletion of SOC stock (Jia et al., 2019; Terrer et al., 2019). For this to occur, ecosystems must adapt and promote a balance between emission and storage. However, the balance between increased respiration in tropical climate and carbon absorption from increased primary production is one of the fundamental uncertainties for the size of carbon stock across the planet (Jia et al., 2019). Although eCO_2 can increase the productivity of vegetation in dry lands, water limitation may have a more pronounced effect, canceling out the fertilization effect (Mirzabaev et al., 2019).

The climate models used differ mainly in areas with the highest SOC stock. The biggest changes of the CanESM5 model in SOC stock compared to the MIROC6 model in all evaluated periods demonstrates how much climate changes can alter the SOC stock of Rondônia. The effect of these changes on SOC stock can take time, as the SOC stabilization process is dynamic and can take decades after ecosystem disturbances (Varney et al., 2020). The projections made are dependent on changes in the microbial community in response to disturbances in ecosystems, such as increased aridity and/or changes in agricultural practices, which alter the dynamics of SOC stock and consequently impact future projections of carbon inventories (Ramírez et al., 2020).

The reduction of the largest SOC stock at all depths and periods, especially in the ssp585 scenario, points to the need for mitigation actions to mitigate the loss of this carbon. The adoption of agroforestry systems, for example, can increase the incorporation of above- and below-ground biomass, such as woody material, root biomass and litter, in addition to providing a wide range of uses for farmers (Rodríguez et al., 2021), which is especially important in the state of Rondônia, as the state is mostly made up of family farmers. Gomes et al. (2020) for example, showed that the conversion of conventional coffee plantations to agroforestry systems can mitigate the effects of climate change and allow coffee production even in the face of increased temperature in future scenarios. The improvement in the microclimate, the incorporation of biomass, plant diversity and, consequently, microbial diversity, among other beneficial factors offered by agroforestry systems, conditions the increase in SOC stock (Chatterjee et al., 2018).

Although the model has established a strong relationship with climate predictors, caution is required in interpreting the results, especially in areas that indicate an increase in SOC. The model does not directly consider the anthropogenic perturbation of future land use and occupation change, which is an important SOC proxy. According to the IPCC, ecosystem degradation and agrarian conflicts will increase pressure on forests, resulting in deforestation and loss of SOC (Smith et al., 2019). This may indicate an overestimation of the future values presented here. In addition, Rondônia is in an Am-type climate region throughout its territory, with low amplitude in the seasonality of temperature (BIO 4), precipitation and temperature, therefore, the adjusted model made predictions in future scenarios with values outside its range training which can impact the magnitude of the prediction of SOC stock.

Data limitation is one of the main constraints for inventorying global SOC inventories (Weintraub et al., 2019). Developing observation systems can help to understand the local inventories of SOC stock and, at the same time, contribute to the validation of models with larger territorial extensions.

5. Conclusion

Random forest was the model with the best performance for predicting soil organic carbon (SOC) in Rondônia. Soil classes, precipitation and maximum temperature were the most important SOC proxies in Rondônia. The largest stock of SOC are in the south and southeast of the state and the smallest in the central territory. For 2100, estimates indicate that for the scenario of growth and equality focused on sustainability (ssp126), and for the scenario of rapid

and unrestricted growth in economic production and energy use (ssp585) there will be losses and gains in SOC stock.

Areas with the highest SOC stock may have SOC reduction and areas where there is little carbon may have gains, this dynamic will be mainly dependent on precipitation. The SOC gain and loss dynamics will be more pronounced in the ssp585 scenario. In the two climate models used (CanESM5 and MIROC6) the differences in the impact on the SOC will be in areas with more carbon, with CanESM5 with the greatest variations. Rondônia's low climatic range may have restricted the model's explanatory capacity for future scenarios with ranges beyond its training range.

References

Ab'Saber, A.N., 1969. Domínios morfoclimáticos e províncias fitogeográficas do Brasil. Orientação 3.

Ab'Saber, A.N., 1967. Problemas geomorfológicos da Amazônia brasileira, in: Atas Do Simpósio Sobre a Biota Amazônica. Rio de Janeiro, pp. 35–67.

Adamy, A., 2010. Geodiversidade do estado de Rondônia, CPRM. Porto Velho.

Ahirwal, J., Nath, A., Brahma, B., Deb, S., Sahoo, U.K., Nath, A.J., 2021. Patterns and driving factors of biomass carbon and soil organic carbon stock in the Indian Himalayan region. Science of the Total Environment 770, 145292. https://doi.org/10.1016/j.scitotenv.2021.145292

Alvares, C.A., Stape, J.L., Sentelhas, P.C., Moraes Gonçalves, J.L., Sparovek, G., 2013. Köppen's climate classification map for Brazil. Meteorologische Zeitschrift 22, 711–728. https://doi.org/10.1127/0941-2948/2013/0507

Arora, V.K., Boer, G.J., Friedlingstein, P., Eby, M., Jones, C.D., Christian, J.R., Bonan, G., Bopp, L., Brovkin, V., Cadule, P., Hajima, T., Ilyina, T., Lindsay, K., Tjiputra, J.F., Wu, T., 2013. Carbon-concentration and carbon-climate feedbacks in CMIP5 earth system models. Journal of Climate 26, 5289–5314. https://doi.org/10.1175/JCLI-D-12-00494.1

Assis, L.F.F.G., Ferreira, K.R., Vinhas, L., Maurano, L., Almeida, C., Carvalho, A., Rodrigues, J., Maciel, A., Camargo, C., 2019. TerraBrasilis: A Spatial Data Analytics Infrastructure for Large-Scale Thematic Mapping. ISPRS International Journal of Geo-Information 8, 513. https://doi.org/10.3390/ijgi8110513

Batjes, N.H., 1996. Total carbon and nitrogen in the soils of the world. European Journal of Soil Science 47, 151–163. https://doi.org/10.1111/j.1365-2389.1996.tb01386.x

BRASIL, 2006. Lei complementar N.º 233, de 06 de junho de 2000. Governador do estado, Brasília.

Breiman, L., 2001. Random forests, Machine Learning. https://doi.org/10.1023/A:1010933404324

Butt, N., Oliveira, P.A., Costa, M.H., 2011. Evidence that deforestation affects the onset of the rainy season in Rondonia, Brazil. Journal of Geophysical Research Atmospheres 116, D11120. https://doi.org/10.1029/2010JD015174

Carvalhais, N., Forkel, M., Khomik, M., Bellarby, J., Jung, M., Migliavacca, M., Mu, M., Saatchi, S., Santoro, M., Thurner, M., Weber, U., Ahrens, B., Beer, C., Cescatti, A., Randerson, J.T., Reichstein, M., 2014. Global covariation of carbon turnover times with climate in terrestrial ecosystems. Nature 514, 213–217. https://doi.org/10.1038/nature13731

Chatterjee, N., Nair, P.K.R., Chakraborty, S., Nair, V.D., 2018. Changes in soil carbon stocks across the Forest-Agroforest-Agriculture/Pasture continuum in various agroecological regions: A meta-analysis. Agriculture, Ecosystems and Environment 266, 55–67. https://doi.org/10.1016/j.agee.2018.07.014

Ciccantell, P.S., Schmink, M., Wood, C.H., 1994. Contested Frontiers in Amazonia., Contemporary Sociology. Columbia University Press, New York. https://doi.org/10.2307/2074888

Cochrane, T.T., Cochrane, T.A., 2006. Diversity of the land resources in the Amazonian state of Rondônia, Brazil. Acta Amazonica 36, 91–102. https://doi.org/10.1590/s0044-59672006000100011

Dale, V.H., O'Neill, R. v., Southworth, F., Pedlowski, M., 1994. Modelando los efectos del manejo de tierras en los asentamientos de la Amazonia Brasilera de Rondônia. Conservation Biology 8, 196–206. https://doi.org/10.1046/j.1523-1739.1994.08010196.x

Durrer, A., Margenot, A.J., Silva, L.C.R., Bohannan, B.J.M., Nusslein, K., van Haren, J., Andreote, F.D., Parikh, S.J., Rodrigues, J.L.M., 2021. Beyond total carbon: conversion of amazon forest to pasture alters indicators of soil C cycling. Biogeochemistry 152, 179–194. https://doi.org/10.1007/s10533-020-00743-x

Eyring, V., Bony, S., Meehl, G.A., Senior, C.A., Stevens, B., Stouffer, R.J., Taylor, K.E., 2016. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. Geoscientific Model Development 9, 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Fathizad, H., Ardakani, M.A.H., Heung, B., Sodaiezadeh, H., Rahmani, A., Fathabadi, A., Scholten, T., Taghizadeh-Mehrjardi, R., 2020. Spatio-temporal dynamic of soil quality in the central Iranian desert modeled with machine learning and digital soil assessment techniques. Ecological Indicators 118, 106736. https://doi.org/10.1016/j.ecolind.2020.106736

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology 37, 4302–4315. https://doi.org/10.1002/joc.5086

Friedman, J.H., 2002. Stochastic gradient boosting. Computational Statistics and Data Analysis 38, 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2

García-Palacios, P., Crowther, T.W., Dacal, M., Hartley, I.P., Reinsch, S., Rinnan, R., Rousk, J., van den Hoogen, J., Ye, J.-S., Bradford, M.A., 2021. Evidence for large microbial-mediated losses of soil carbon under anthropogenic warming. Nature Reviews Earth & Environment 1–11. https://doi.org/10.1038/s43017-021-00178-4

Giardina, C.P., Litton, C.M., Crow, S.E., Asner, G.P., 2014. Warming-related increases in soil CO2 efflux are explained by increased below-ground carbon flux. Nature Climate Change 4, 822–827. https://doi.org/10.1038/nclimate2322

Gomes, L.C., Bianchi, F.J.J.A., Cardoso, I.M., Fernandes, R.B.A., Filho, E.I.F., Schulte, R.P.O., 2020. Agroforestry systems can mitigate the impacts of climate change on coffee production: A spatially explicit assessment in Brazil. Agriculture, Ecosystems and Environment 294, 106858. https://doi.org/10.1016/j.agee.2020.106858

Gomes, L.C., Faria, R.M., Souza, E., Veloso, G.V., Schaefer, C.E.G.R., Fernandes Filho, E.I., 2019. Modelling and mapping soil organic carbon stocks in Brazil. Geoderma 340, 337–350. https://doi.org/10.1016/j.geoderma.2019.01.007

Gonçalves, D.R.P., Mishra, U., Wills, S., Gautam, S., 2021. Regional environmental controllers influence continental scale soil carbon stocks and future carbon dynamics. Scientific Reports 11, 6474. https://doi.org/10.1038/s41598-021-85992-y

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Goydaragh, M.G., Taghizadeh-Mehrjardi, R., Jafarzadeh, A.A., Triantafilis, J., Lado, M., 2021. Using environmental variables and Fourier Transform Infrared Spectroscopy to predict soil organic carbon. Catena 202, 105280. https://doi.org/10.1016/j.catena.2021.105280

Goza, F., 1994. Brazilian frontier settlement: The case of Rondônia. Population and Environment 16, 37–60. https://doi.org/10.1007/BF02208002

Greenwell, B., Boehmke, B., Cunningham, J., GBM Developers, 2020. gbm: Generalized Boosted Regression Models.

Hararuk, O., Smith, M.J., Luo, Y., 2015. Microbial models with data-driven parameters predict stronger soil carbon responses to climate change. Global Change Biology 21, 2439–2453. https://doi.org/10.1111/gcb.12827

Hengl, T., 2020. GSIF: Global Soil Information Facilities.

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., 2017. SoilGrids250m: Global gridded soil information based on machine learning. PLoS ONE 12, e0169748. https://doi.org/10.1371/journal.pone.0169748

Heung, B., Ho, H.C., Zhang, J., Knudby, A., Bulmer, C.E., Schmidt, M.G., 2016. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. Geoderma 265, 62–77. https://doi.org/10.1016/j.geoderma.2015.11.014

Hijmans, R.J., 2021. raster: Geographic Data Analysis and Modeling.

Hinson, A.L., Feagin, R.A., Eriksson, M., Najjar, R.G., Herrmann, M., Bianchi, T.S., Kemp, M., Hutchings, J.A., Crooks, S., Boutton, T., 2017. The spatial distribution of soil organic carbon in tidal wetland soils of the continental United States. Global Change Biology 23, 5468–5480. https://doi.org/10.1111/gcb.13811

IBGE, 2019. Censo Agropecuário 2017: resultados definitivos. Rio de Janeiro.

IBGE, 1981. Censo Demográfico 1940-2010.

Jia, G., Shevliakova, Elena, Artaxo, Paulo, Noblet-Ducoudré, Nathalie, Houghton, Richard, Anderegg, W., Bernier, P., Carlo Espinoza, J., Semenov, S., Xu, X., Shevliakova, E, Artaxo, P, de Noblet-Ducoudré, N, Houghton, R, House, J., Kitajima, K., Lennard, C., Popp, A., Sirin, A., Sukumar, R., Shukla, P., Skea, J., Calvo Buendia, E., Masson-Delmotte, V., Pörtner, H., Roberts, D., Zhai, P., Slade, R., Connors, S., van Diemen, R., Ferrat, M., Haughey, E., Luz, S., Neogi, S., Pathak, M., Petzold, J., Portugal Pereira, J., Vyas, P., Huntley, E., Kissick, K., Malley, J., 2019. 2 Land-climate interactions. Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems.

Jiang, M., Medlyn, B.E., Drake, J.E., Duursma, R.A., Anderson, I.C., Barton, C.V.M., Boer, M.M., Carrillo, Y., Castañeda-Gómez, L., Collins, L., Crous, K.Y., de Kauwe, M.G., dos Santos, B.M., Emmerson, K.M., Facey, S.L., Gherlenda, A.N., Gimeno, T.E., Hasegawa, S., Johnson, S.N., Kännaste, A., Macdonald, C.A., Mahmud, K., Moore, B.D., Nazaries, L., Neilson, E.H.J., Nielsen, U.N., Niinemets, Ü., Noh, N.J., Ochoa-Hueso, R., Pathare, V.S., Pendall, E., Pihlblad, J., Piñeiro, J., Powell, J.R., Power, S.A., Reich, P.B., Renchon, A.A., Riegler, M., Rinnan, R., Rymer, P.D., Salomón, R.L., Singh, B.K., Smith, B., Tjoelker, M.G., Walker, J.K.M., Wujeska-Klause, A., Yang, J., Zaehle, S., Ellsworth, D.S., 2020. The fate of carbon in a mature forest under carbon dioxide enrichment. Nature 580, 227–231. https://doi.org/10.1038/s41586-020-2128-9

Karatzoglou, A., Hornik, K., Smola, A., Zeileis, A., 2004. kernlab - An S4 package for kernel methods in R. Journal of Statistical Software 11, 1–20. https://doi.org/10.18637/jss.v011.i09

Khanna, J., Medvigy, D., Fueglistaler, S., Walko, R., 2017. Regional dry-season climate changes due to three decades of Amazonian deforestation. Nature Climate Change 7, 200–204. https://doi.org/10.1038/nclimate3226

Klinges, D.H., Scheffers, B.R., 2021. Microgeography, Not Just Latitude, Drives Climate Overlap on Mountains from Tropical to Polar Ecosystems. The American Naturalist 197, 75–92. https://doi.org/10.1086/711873

Kuhn, M., Johnson, K., 2013. Applied predictive modeling, Applied Predictive Modeling. Springer New York, New York, NY. https://doi.org/10.1007/978-1-4614-6849-3

Kuhn, M., Quinlan, R., 2021. Cubist: Rule - And Instance-Based Regression Modeling.

Li, H., Wu, Y., Chen, J., Zhao, F., Wang, F., Sun, Y., Zhang, G., Qiu, L., 2021. Responses of soil organic carbon to climate change in the Qilian Mountains and its future projection. Journal of Hydrology 596, 126110. https://doi.org/10.1016/j.jhydrol.2021.126110

Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest, R News.

Luo, Y., Ahlström, A., Allison, S.D., Batjes, N.H., Brovkin, V., Carvalhais, N., Chappell, A., Ciais, P., Davidson, E.A., Finzi, A., Georgiou, K., Guenet, B., Hararuk, O., Harden, J.W., He, Y., Hopkins, F., Jiang, L., Koven, C., Jackson, R.B., Jones, C.D., Lara, M.J., Liang, J., McGuire, A.D., Parton, W., Peng, C., Randerson, J.T., Salazar, A., Sierra, C.A., Smith, M.J., Tian, H., Todd-Brown, K.E.O., Torn, M., van Groenigen, K.J., Wang, Y.P., West, T.O., Wei, Y., Wieder, W.R., Xia, J., Xu, Xia, Xu, Xiaofeng, Zhou, T., 2016. Toward more realistic projections of soil carbon dynamics by Earth system models. Global Biogeochemical Cycles 30, 40–56. https://doi.org/10.1002/2015GB005239

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4

MDA - Ministério do Desenvolvimento Agrário, 2005. Plano Territorial de Desenvolvimento Rural Sustentável: guia para o planejamento. Documento de Apoio Nº 02. Secretaria de Desenvolvimento Territorial - SDT, Brasília.

Meinshausen, M., Nicholls, Z.R.J., Lewis, J., Gidden, M.J., Vogel, E., Freund, M., Beyerle, U., Gessner, C., Nauels, A., Bauer, N., Canadell, J.G., Daniel, J.S., John, A., Krummel, P.B., Luderer, G., Meinshausen, N., Montzka, S.A., Rayner, P.J., Reimann, S., Smith, S.J., van den Berg, M., Velders, G.J.M., Vollmer, M.K., Wang, R.H.J., 2020. The shared socio-economic pathway (SSP) greenhouse gas concentrations and their extensions to 2500. Geoscientific Model Development 13, 3571–3605. https://doi.org/10.5194/gmd-13-3571-2020

Milborrow, S., 2020. earth: Multivariate Adaptive Regression Splines.

Millikan, B.H., 1992. Tropical deforestation, land degradation, and society: lessons from Rondonia, Brazil. Latin American Perspectives 19, 45–72. https://doi.org/10.1177/0094582X9201900103

Mirzabaev, A., Wu, J., Evans, J., García-Oliva, F., Hussein, I.A.G., Iqbal, M.H., Kimutai, J., Knowles, T., Meza, F., Nedjraoui, D., Tena, F., Türkeş, M., Vázquez, R.J., Weltz, M., 2019. 3 Desertification. Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems.

Moran, E.F., 1981. Developing the Amazon. Indiana University Press.

NASA JPL, 2020. NASADEM Merged DEM Global 1 arc second V001. NASA EOSDIS Land Processes DAAC. https://doi.org/https://doi.org/10.5067/MEaSUREs/NASADEM/NASADEM_HGT.001

Nelson, D.W., Sommers, L.E., 1982. Methods of Soil Analysis. Part 2. Chemical and Microbiological Properties Methods of Soil Analysis. Part 2. Chemical and Microbiological Properties., Madison, WI: American Society of Agronomy, Soil Science Society of America. Madison.

Numata, I., Chadwick, O.A., Roberts, D.A., Schimel, J.P., Sampaio, F.F., Leonidas, F.C., Soares, J. v., 2007. Temporal nutrient variation in soil and vegetation of post-forest pastures as a function of soil order, pasture age, and management, Rondônia, Brazil. Agriculture, Ecosystems and Environment 118, 159–172. https://doi.org/10.1016/j.agee.2006.05.019

O'Brien, J., 2021. gdalUtilities: Wrappers for "GDAL" Utilities Executables.

Olsson, L., Barbosa, Humberto, Bhadwal, Suruchi, Manuel Moreno, J., Vera, C., Salisu Barau, A., Barbosa, H, Bhadwal, S, Cowie, A., Delusca, K., Flores-Renteria, D., Hermans, K., Jobbagy, E., Kurz, W., Li, D., Shukla, P., Skea, J., Calvo Buendia, E., Masson-Delmotte, V., Pörtner, H., Roberts, D.C., Zhai, P., Slade, R., Connors, S., van Diemen, R., Ferrat, M., Haughey, E., Luz, S., Neogi, S., Pathak, M., Petzold, J., Portugal Pereira, J., Vyas, P., Huntley, E., Kissick, K., Belkacemi, M., Malley, J., 2019. 4 Land degradation. Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems.

Paluszynska, A., Biecek, P., Jiang, Y., 2020. randomForestExplainer: Explaining and Visualizing Random Forests in Terms of Variable Importance.

Pedlowski, M.A., Dale, V.H., Matricardi, E.A.T., da Silva Filho, E.P., 1997. Patterns and impacts of deforestation in Rondonia, Brazil. Landscape and Urban Planning 38, 149–157. https://doi.org/10.1016/S0169-2046(97)00030-3

Pereira, R.H.M., Goncalves, C.N., 2021. geobr: Download Official Spatial Data Sets of Brazil.

Piontekowski, V.J., Ribeiro, F.P., Matricardi, E.A.T., Lustosa, I.M., Bussinguer, A.P., Gatto, A., 2019. Modeling deforestation in the state of Rondonia. Floresta e Ambiente 26, 20180441. https://doi.org/10.1590/2179-8087.044118

R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

Ramesh, T., Bolan, N.S., Kirkham, M.B., Wijesekara, H., Kanchikerimath, M., Srinivasa Rao, C., Sandeep, S., Rinklebe, J., Ok, Y.S., Choudhury, B.U., Wang, H., Tang, C., Wang, X., Song, Z., Freeman, O.W., 2019. Soil organic carbon dynamics: Impact of land use changes and management practices: A review, in: Advances in Agronomy. Academic Press Inc., pp. 1–107. https://doi.org/10.1016/bs.agron.2019.02.001

Ramírez, P.B., Fuentes-Alburquenque, S., Díez, B., Vargas, I., Bonilla, C.A., 2020. Soil microbial community responses to labile organic carbon fractions in relation to soil type and land use along a climate gradient. Soil Biology and Biochemistry 141, 107692. https://doi.org/10.1016/j.soilbio.2019.107692

Riahi, K., van Vuuren, D.P., Kriegler, E., Edmonds, J., O'Neill, B.C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., Lutz, W., Popp, A., Cuaresma, J.C., KC, S., Leimbach, M., Jiang, L., Kram, T., Rao, S., Emmerling, J., Ebi, K., Hasegawa, T., Havlik, P., Humpenöder, F., da Silva, L.A., Smith, S., Stehfest, E., Bosetti, V., Eom, J., Gernaat, D.,

Masui, T., Rogelj, J., Strefler, J., Drouet, L., Krey, V., Luderer, G., Harmsen, M., Takahashi, K., Baumstark, L., Doelman, J.C., Kainuma, M., Klimont, Z., Marangoni, G., Lotze-Campen, H., Obersteiner, M., Tabeau, A., Tavoni, M., 2017. The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. Global Environmental Change 42, 153–168. https://doi.org/10.1016/j.gloenvcha.2016.05.009

Rodríguez, L., Suárez, J.C., Rodriguez, W., Artunduaga, K.J., Lavelle, P., 2021. Agroforestry systems impact soil macroaggregation and enhance carbon storage in Colombian deforested Amazonia. Geoderma 384, 114810. https://doi.org/10.1016/j.geoderma.2020.114810

Scharlemann, J.P.W., Tanner, E.V.J., Hiederer, R., Kapos, V., 2014. Global soil carbon: Understanding and managing the largest terrestrial carbon pool. Carbon Management 5, 81–91. https://doi.org/10.4155/cmt.13.77

Schliep, K., Hechenbichler, K., 2016. kknn: Weighted k-Nearest Neighbors.

SEDAM, 2021. Coordenadoria de Geociências [WWW Document]. SEDAM. URL http://cogeo.sedam.ro.gov.br/ (accessed 7.18.20).

Smith, P., Nkem, J., Calvin, K., Campbell, D., Cherubini, F., Grassi, G., Korotkov, V., Hoang, A.L., Lwasa, S., McElwee, P., Nkonya, E., Saigusa, N., Soussana, J.F., Taboada, M.A., 2019. 6 Interlinkages Between Desertification, Land Degradation, Food Security and Greenhouse Gas Fluxes: Synergies, Trade-offs and Integrated Response Options. Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems.

Sodango, T.H., Sha, J., Li, X., Noszczyk, T., Shang, J., Aneseyee, A.B., Bao, Z., 2021. Modeling the spatial dynamics of soil organic carbon using remotely-sensed predictors in Fuzhou City, China. Remote Sensing 13, 1682. https://doi.org/10.3390/rs13091682

Soucémarianadin, L.N., Cécillon, L., Guenet, B., Chenu, C., Baudin, F., Nicolas, M., Girardin, C., Barré, P., 2018. Environmental factors controlling soil organic carbon stability in French forest soils. Plant and Soil 426, 267–286. https://doi.org/10.1007/s11104-018-3613-x

Souza, C.M., Shimbo, J.Z., Rosa, M.R., Parente, L.L., Alencar, A.A., Rudorff, B.F.T., Hasenack, H., Matsumoto, M., Ferreira, L.G., Souza-Filho, P.W.M., de Oliveira, S.W., Rocha, W.F., Fonseca, A. v., Marques, C.B., Diniz, C.G., Costa, D., Monteiro, D., Rosa, E.R., Vélez-Martin, E., Weber, E.J., Lenti, F.E.B., Paternost, F.F., Pareyn, F.G.C.,

Siqueira, J. v., Viera, J.L., Neto, L.C.F., Saraiva, M.M., Sales, M.H., Salgado, M.P.G., Vasconcelos, R., Galano, S., Mesquita, V. v., Azevedo, T., 2020. Reconstructing three decades of land use and land cover changes in brazilian biomes with landsat archive and earth engine. Remote Sensing 12, 2735. https://doi.org/10.3390/RS12172735

Spawn, S.A., Sullivan, C.C., Lark, T.J., Gibbs, H.K., 2020. Harmonized global maps of above and belowground biomass carbon density in the year 2010. Scientific Data 7, 112. https://doi.org/10.1038/s41597-020-0444-4

Terrer, C., Jackson, R.B., Prentice, I.C., Keenan, T.F., Kaiser, C., Vicca, S., Fisher, J.B., Reich, P.B., Stocker, B.D., Hungate, B.A., Peñuelas, J., McCallum, I., Soudzilovskaia, N.A., Cernusak, L.A., Talhelm, A.F., van Sundert, K., Piao, S., Newton, P.C.D., Hovenden, M.J., Blumenthal, D.M., Liu, Y.Y., Müller, C., Winter, K., Field, C.B., Viechtbauer, W., van Lissa, C.J., Hoosbeek, M.R., Watanabe, M., Koike, T., Leshyk, V.O., Polley, H.W., Franklin, O., 2019. Nitrogen and phosphorus constrain the CO2 fertilization of global plant biomass. Nature Climate Change. https://doi.org/10.1038/s41558-019-0545-2

Todd-Brown, K.E.O., Randerson, J.T., Post, W.M., Hoffman, F.M., Tarnocai, C., Schuur, E.A.G., Allison, S.D., 2013. Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. Biogeosciences 10, 1717–1736. https://doi.org/10.5194/bg-10-1717-2013

Varney, R.M., Chadburn, S.E., Friedlingstein, P., Burke, E.J., Koven, C.D., Hugelius, G., Cox, P.M., 2020. A spatial emergent constraint on the sensitivity of soil carbon turnover to global warming. Nature Communications 11, 5544. https://doi.org/10.1038/s41467-020-19208-8

Vasques, G.M., Dart, R.O., Baca, J.F.M., Ceddia, M.B., Santos, M. de L.M., 2017. Soil Organic Carbon Stock at 0-30 cm Map for Brazil, Embrapa. Rio de Janeiro.

Walker, A.P., de Kauwe, M.G., Medlyn, B.E., Zaehle, S., Iversen, C.M., Asao, S., Guenet, B., Harper, A., Hickler, T., Hungate, B.A., Jain, A.K., Luo, Y., Lu, X., Lu, M., Luus, K., Megonigal, J.P., Oren, R., Ryan, E., Shu, S., Talhelm, A., Wang, Y.P., Warren, J.M., Werner, C., Xia, J., Yang, B., Zak, D.R., Norby, R.J., 2019. Decadal biomass increment in early secondary succession woody ecosystems is increased by CO 2 enrichment. Nature Communications 10, 454. https://doi.org/10.1038/s41467-019-08348-1

Wang, K., Qi, Y., Guo, W., Zhang, J., Chang, Q., 2021. Retrieval and mapping of soil organic carbon using sentinel-2A spectral images from bare cropland in autumn. Remote Sensing 13, 1072. https://doi.org/10.3390/rs13061072

Wang, S., Zhuang, Q., Wang, Q., Jin, X., Han, C., 2017. Mapping stocks of soil organic carbon and soil total nitrogen in Liaoning Province of China. Geoderma 305, 250–263. https://doi.org/10.1016/j.geoderma.2017.05.048

Wayne, G.P., 2013. The Beginner's Guide to Representative Concentration Pathways (RCPs). Skeptical Sciece 1.0, 1–24.

Weintraub, S.R., Flores, A.N., Wieder, W.R., Sihi, D., Cagnarini, C., Gonçalves, D.R.P., Young, M.H., Li, L., Olshansky, Y., Baatz, R., Sullivan, P.L., Groffman, P.M., 2019. Leveraging Environmental Research and Observation Networks to Advance Soil Carbon Science. Journal of Geophysical Research: Biogeosciences 124, 1047–1055. https://doi.org/10.1029/2018JG004956

Supplementary material

Table S10. Summary of the performance metrics of the models evaluated to build the soil bulk density pedotransfer function, Rondônia State, Brazil.

Model	Tranning			Test		
	RMSE (g cm^{-3})	MAE (g cm^{-3})	R^2	RMSE (g cm^{-3})	MAE (g cm^{-3})	R^2
avNNet	0,323	0,292	0,395	0,329	0,299	0,260
brnn	0,083	0,064	0,703	0,107	0,083	0,557
cubist	0,000	0,000	1,000	0,086	0,053	0,715
dnn	0,142	0,114	0,278	0,155	0,126	0,219
earth	0,663	0,389	0,005	1,102	0,480	0,000
extraTrees	0,012	0,008	0,995	0,084	0,061	0,735
gbm	0,176	0,140	0,000	0,176	0,141	0,011
gcvEarth	0,088	0,068	0,667	0,107	0,085	0,549
pcaNNet	0,323	0,292	0,380	0,329	0,299	0,389
rf	0,036	0,026	0,961	0,095	0,073	0,658
svmLinear	0,095	0,073	0,608	0,106	0,080	0,562
svmPoly	0,090	0,068	0,645	0,111	0,083	0,515
svmRadial	0,032	0,021	0,959	0,080	0,052	0,751
xgbDART	0,012	0,009	0,994	0,101	0,075	0,602
xgbTree	0,001	0,001	1,000	0,101	0,073	0,603

Where: MAE = Mean absolute error; R2 = Coefficient of determination; RMSE = Root mean squared error.

Table S11. Relative importance and description of the ten main predictors used in the construction of the soil density pedotransfer function with the svmRadial model, Rondônia State, Brazil.

Predictor	Importance	Description
SAND	100,00 %	Sand
SOC	93,86 %	Organic carbon content, Walkley and Black method
CLAY	82,02 %	Clay
HEX	61,44 %	Hydrogen extracted $cmol_c\ kg^{-1}$ (extracted with calcium acetate solution)
N	55,07 %	Total nitrogen
CTC	51,35 %	Exchangeable cation capacity (Ca + Mg + K + Na) $cmol_c\ kg^{-1}$
ACEX	50,02 %	H + Al cmol kg^{-1}
Mg	41,74 %	Mg^{2+} exchangeable in $cmol_c\ kg^{-1}$
SILT	33,79 %	Silt
CTCC	24,56 %	CTC x 100 / % clay

Table S12. Environmental variables used as predictors of SOC stock for Rondônia State, Brazil.

Morphometric	T	Bioclimatic	Precipitation	Soil	Vegetation
Accumulation	Tmin 1	Bio 1	Precip. 1	Geology	EVI
Aspect	Tmin 2	Bio 2	Precip. 2	Soil class	MSAVI
Convergence Index	Tmin 3	Bio 3	Precip. 3	-	NBR
Cross-Sectional Curvature	Tmin 4	Bio 4	Precip. 4	-	NBR2
Curvature Classification	Tmin 5	Bio 5	Precip. 5	-	NDVI
Difference	Tmin 6	Bio 6	Precip. 6	-	NMDI
Diffuse Insolation	Tmin 7	Bio 7	Precip. 7	-	SAVI
Direct Insolation	Tmin 8	Bio 8	Precip. 8	-	-
Direct to Diffuse Ratio	Tmin 9	Bio 9	Precip. 9	-	-

Diurnal Anisotropic Heating	Tmin 10	Bio 10	Precip. 10	-	-
Duration of Insolation	Tmin 11	Bio 11	Precip. 11	-	-
Easterness	Tmin 12	Bio 12	Precip. 12	-	-
Flow Line Curvature	Tmax 1	Bio 13	-	-	-
General Curvature	Tmax 2	Bio 14	-	-	-
Hill Height	Tmax 3	Bio 15	-	-	-
Hill Index	Tmax 4	Bio 16	-	-	-
Hillslope Index	Tmax 5	Bio 17	-	-	-
Landforms	Tmax 6	Bio 18	-	-	-
Longitudinal Curvature	Tmax 7	Bio 19	-	-	-
Mass Balance Index	Tmax 8	-	-	-	-
Maximal Curvature	Tmax 9	-	-	-	-
DEM	Tmax 10	-	-	-	-
Mid-Slope Positon	Tmax 11	-	-	-	-
Minimal Curvature	Tmax 12	-	-	-	-
Multiresolution Index of Valley Bottom Flatness (MRVBF)	-	-	-	-	-
Normalized Height	-	-	-	-	-
Northerness	-	-	-	-	-
Plan Curvature	-	-	-	-	-
Profile Curvature	-	-	-	-	-
Protection Index	-	-	-	-	-
Real Surface Area	-	-	-	-	-
Slope	-	-	-	-	-
Slope Height	-	-	-	-	-
Standardized Height	-	-	-	-	-
Sunrise	-	-	-	-	-
Sunset	-	-	-	-	-
Surface Specific Points	-	-	-	-	-
Tangential Curvature	-	-	-	-	-
Terrain Ruggedness Index (TRI)	-	-	-	-	-
Terrain Surface Convexity	-	-	-	-	-
Terrain Surface Texture	-	-	-	-	-
Topographic Position Index	-	-	-	-	-
Topographic Wetness Index	-	-	-	-	-
Total Curvature	-	-	-	-	-
Total Insolation	-	-	-	-	-
Valley Depth	-	-	-	-	-
Valley Index	-	-	-	-	-
Vector Terrain Ruggedness (VRM)	-	-	-	-	-

Where: DEM = Digital elevation model.

T = Temperature; Tmin = Minimum temperature, Tmax = Maximum temperature (1, 2, 3, ..., 12 = Jan., Feb., Mar., ..., Dec.).

Bio 1 = Annual mean temperature; Bio 2 = Mean diurnal range (Mean of monthly (max temp - min temp)); Bio 3 = Isothermality (Bio 2 / Bio 7) (×100); Bio 4 = Temperature seasonality (standard deviation × 100); Bio 5 = Max temperature of warmest month; Bio 6 = Min temperature of coldest month; Bio 7 = Temperature annual range (Bio 5 – Bio 6); Bio 8 = Mean temperature of wettest quarter; Bio 9 = Mean temperature of driest quarter; Bio 10 = Mean temperature of warmest quarter; Bio 11 = Mean temperature of coldest quarter; Bio 12 = Annual precipitation; Bio 13 = Precipitation of wettest month; Bio 14 = Precipitation of driest month; Bio 15 = Precipitation seasonality (coefficient of variation); Bio 16 = Precipitation of wettest quarter; Bio 17 = Precipitation of Driest Quarter; Bio 18 = Precipitation of Warmest Quarter; Bio 19 = Precipitation of coldest quarter.

Precip. = Precipitation (1, 2, 3, ..., 12 = Jan, Feb., Mar., ..., Dec.).

EVI = Enhanced vegetation index; MSAVI = Modified soil adjusted vegetation index; NBR = Normalized burn ratio; NBR2 = Normalized burn ratio 2; NDMI = Normalized difference moisture index; NDVI = Normalized difference vegetation index; SAVI = Soil adjusted vegetation index.

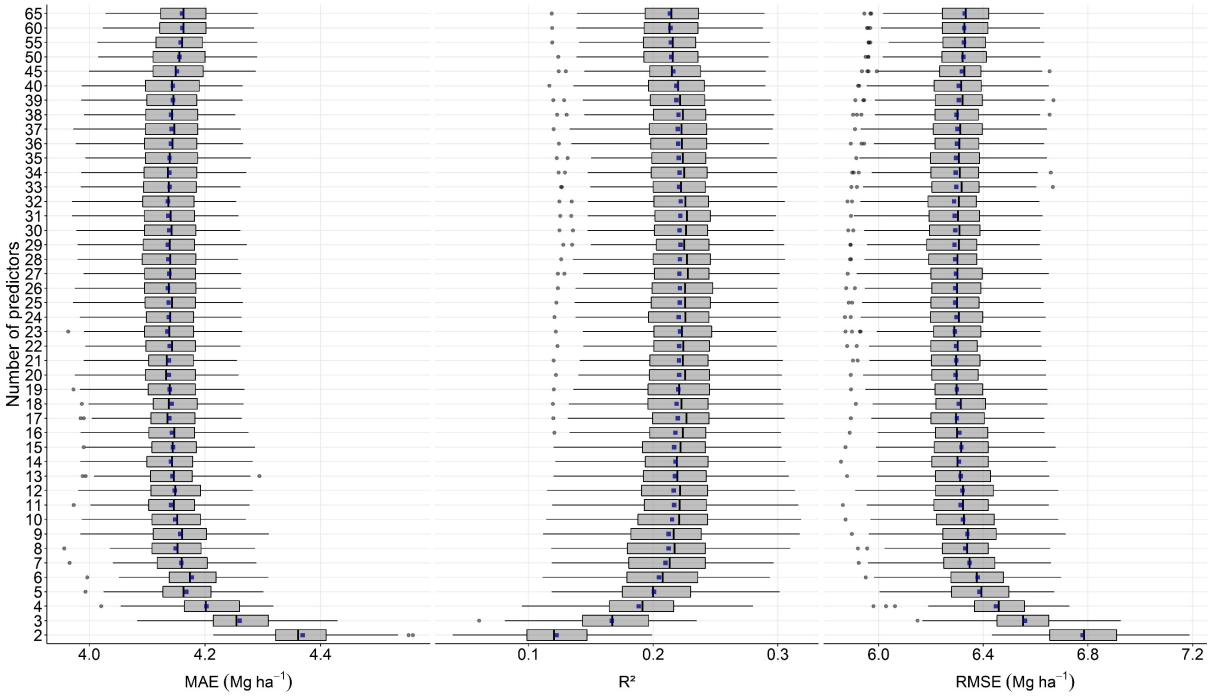

Figure S6. Performance of recursive feature elimination for random forest on different subsets of predictors of the SOC stock in the 0-5 cm layer for the 100 runs, Rondônia State, Brazil. Blue squares in boxplots mean the arithmetic mean. MAE = Mean absolute error; R2 = Coefficient of determination; RMSE = Root mean squared error.

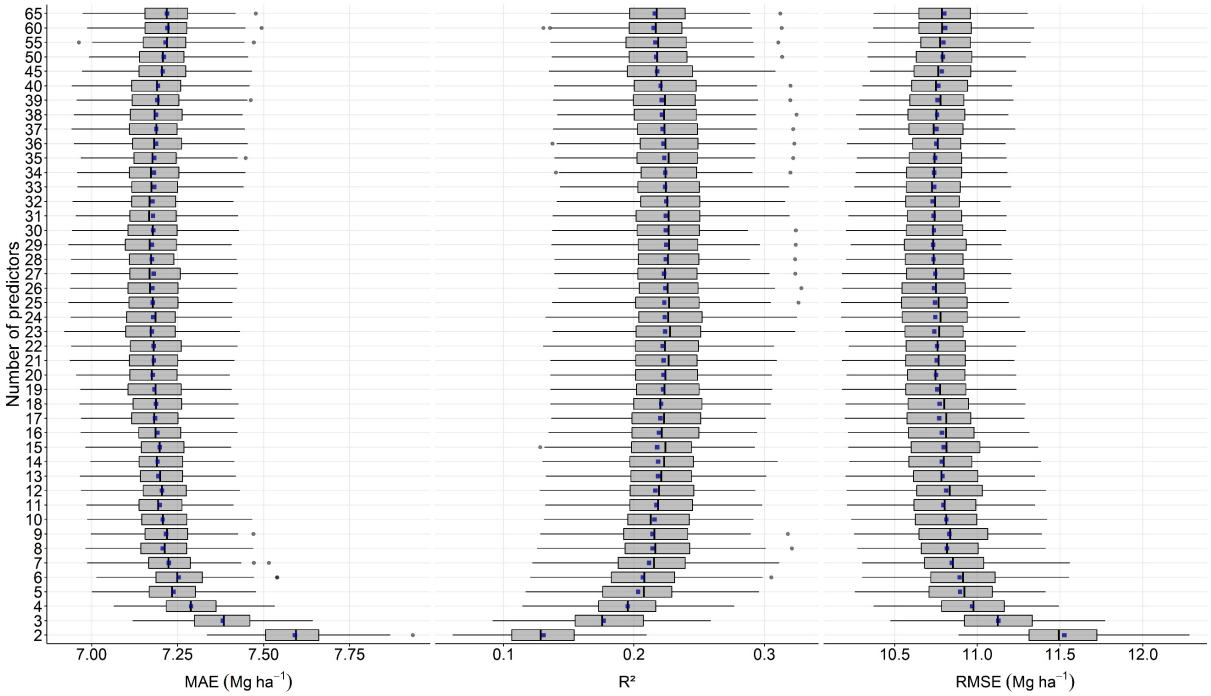

Figure S7. Performance of recursive feature elimination for random forest on different subsets of predictors of the SOC stock in the 5-15 cm layer for the 100 runs, Rondônia State, Brazil. Blue squares in boxplots mean the arithmetic mean. MAE = Mean absolute error; R2 = Coefficient of determination; RMSE = Root mean squared error.

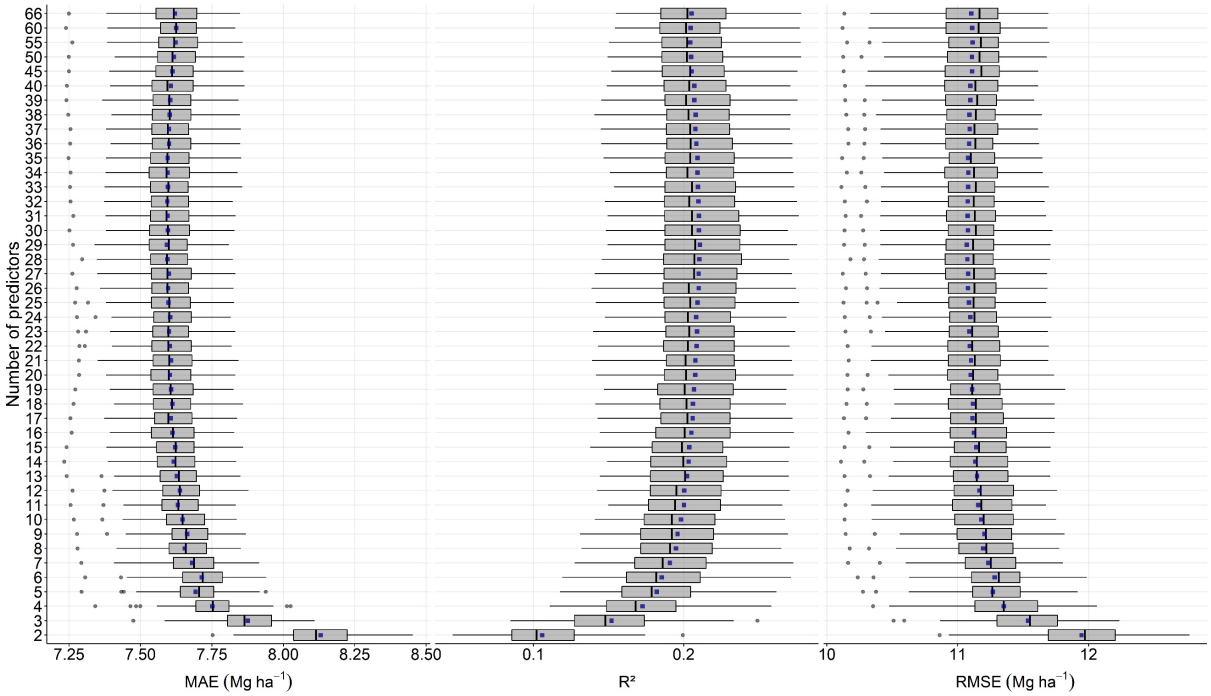

Figure S8. Performance of recursive feature elimination for random forest on different subsets of predictors of the SOC stock in the 15-30 cm layer for the 100 runs, Rondônia State, Brazil. Blue squares in boxplots mean the arithmetic mean. MAE = Mean absolute error; R2 = Coefficient of determination; RMSE = Root mean squared error.

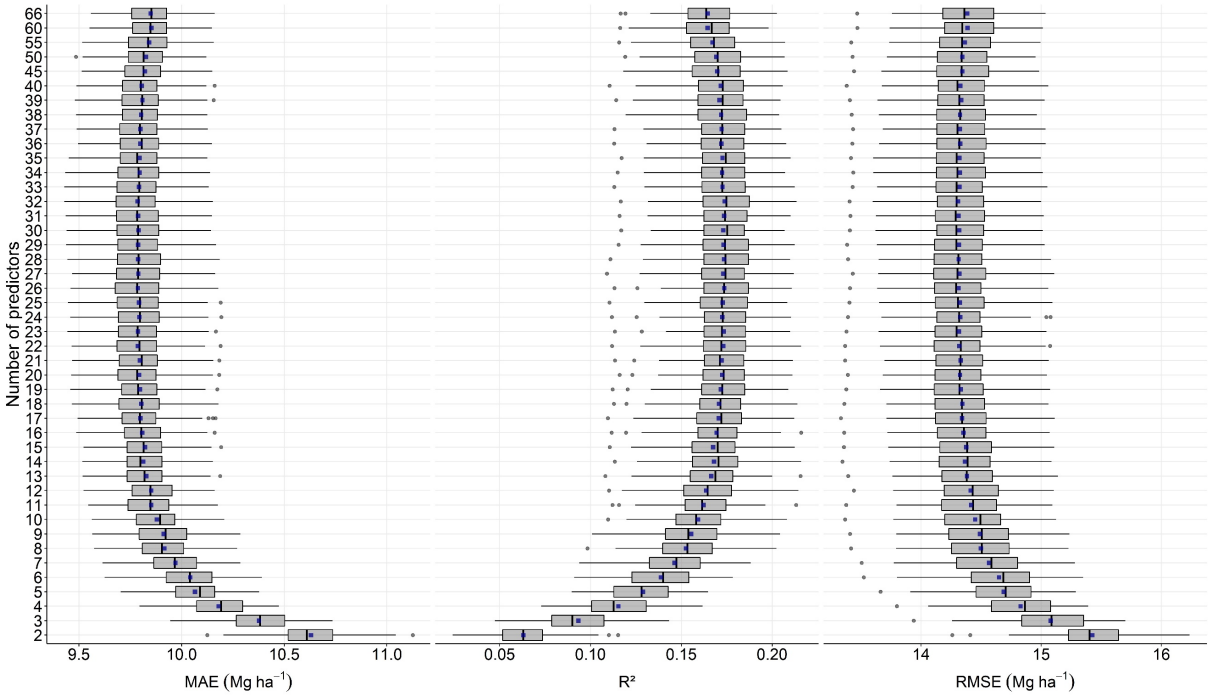

Figure S9. Performance of recursive feature elimination for random forest on different subsets of predictors of the SOC stock in the 30-60 cm layer for the 100 runs, Rondônia State, Brazil. Blue squares in boxplots mean the arithmetic mean. MAE = Mean absolute error; R2 = Coefficient of determination; RMSE = Root mean squared error.

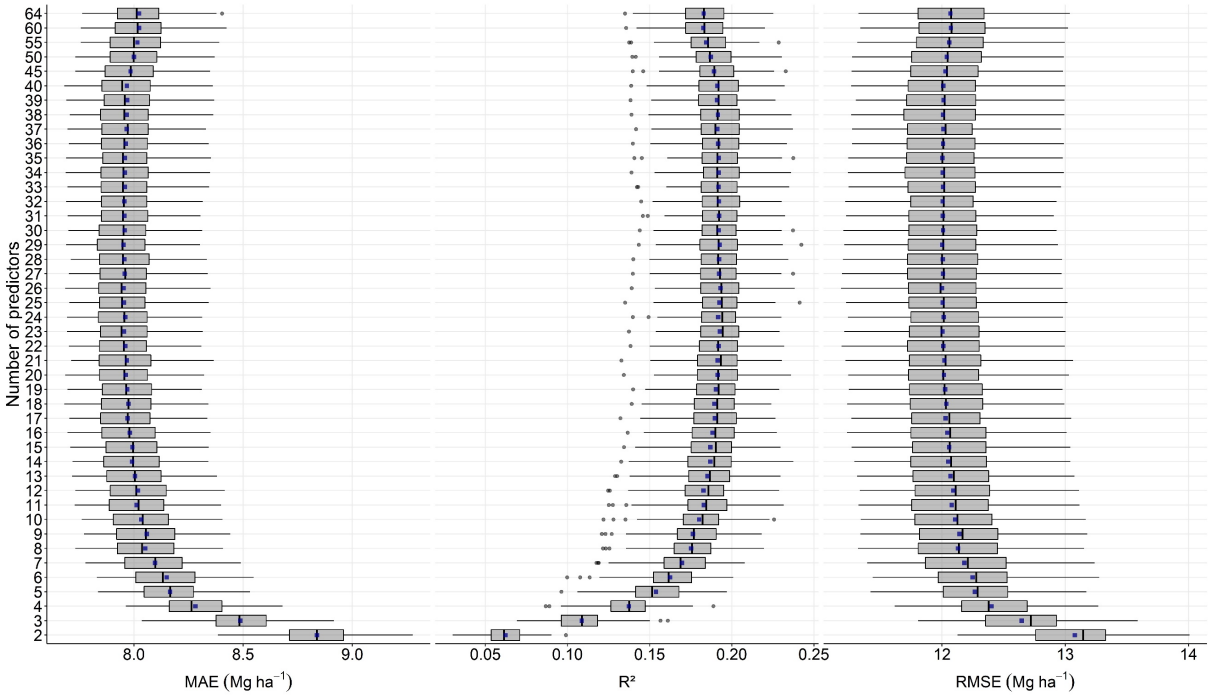

Figure S10. Performance of recursive feature elimination for random forest on different subsets of predictors of the SOC stock in the 60-100 cm layer for the 100 runs, Rondônia State, Brazil. Blue squares in boxplots mean the arithmetic mean. MAE = Mean absolute error; R2 = Coefficient of determination; RMSE = Root mean squared error.

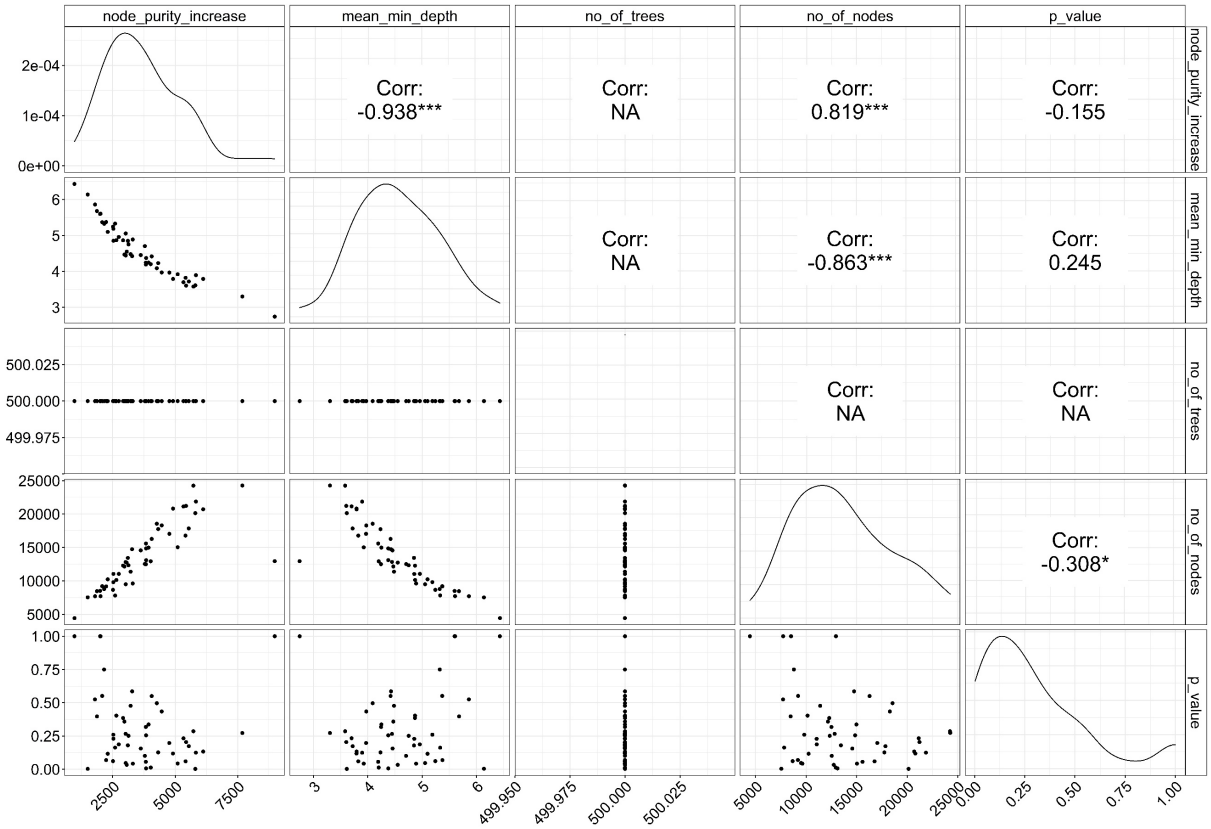

Figure S11. Bilateral relationships between measures of importance of predictors used in random forest to predict SOC stock in the 0-5 cm layer, Rondônia State, Brazil.

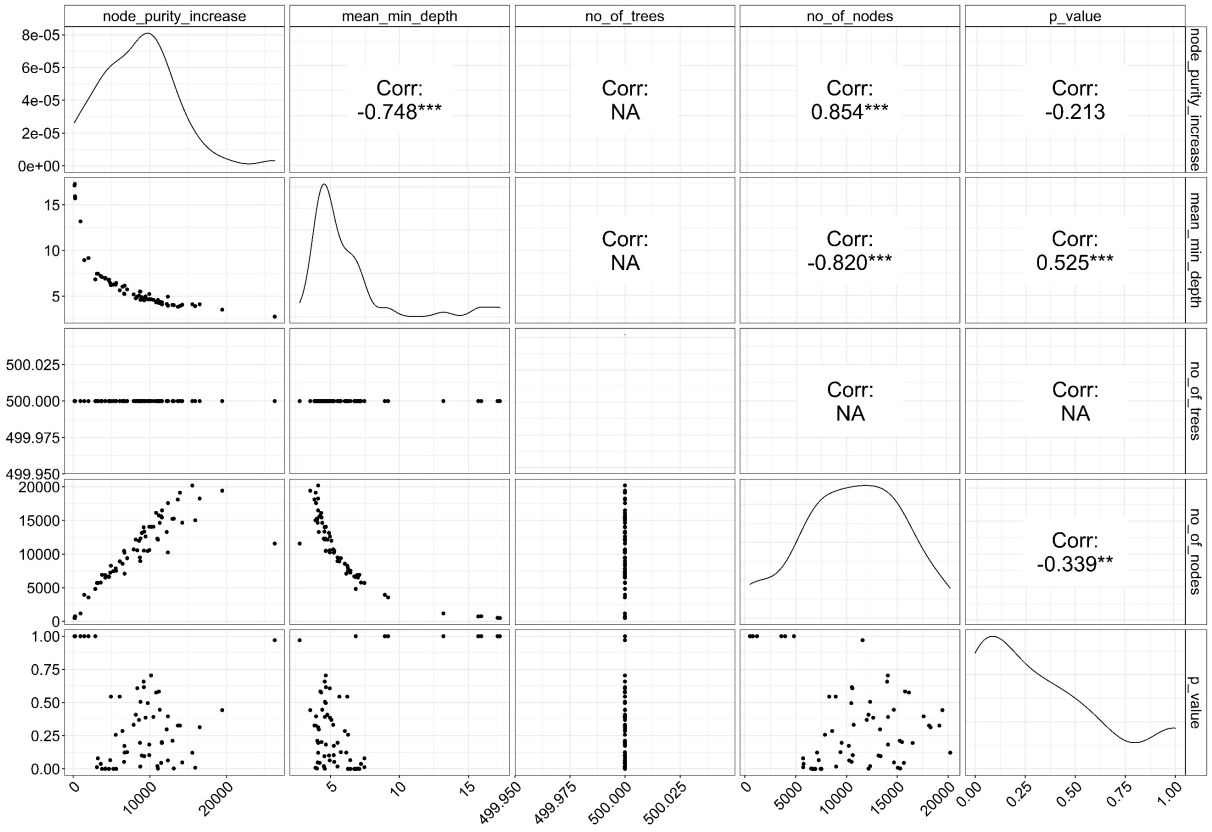

Figure S12. Bilateral relationships between measures of importance of predictors used in random forest to predict SOC stock in the 0-5 cm layer, Rondônia State, Brazil.

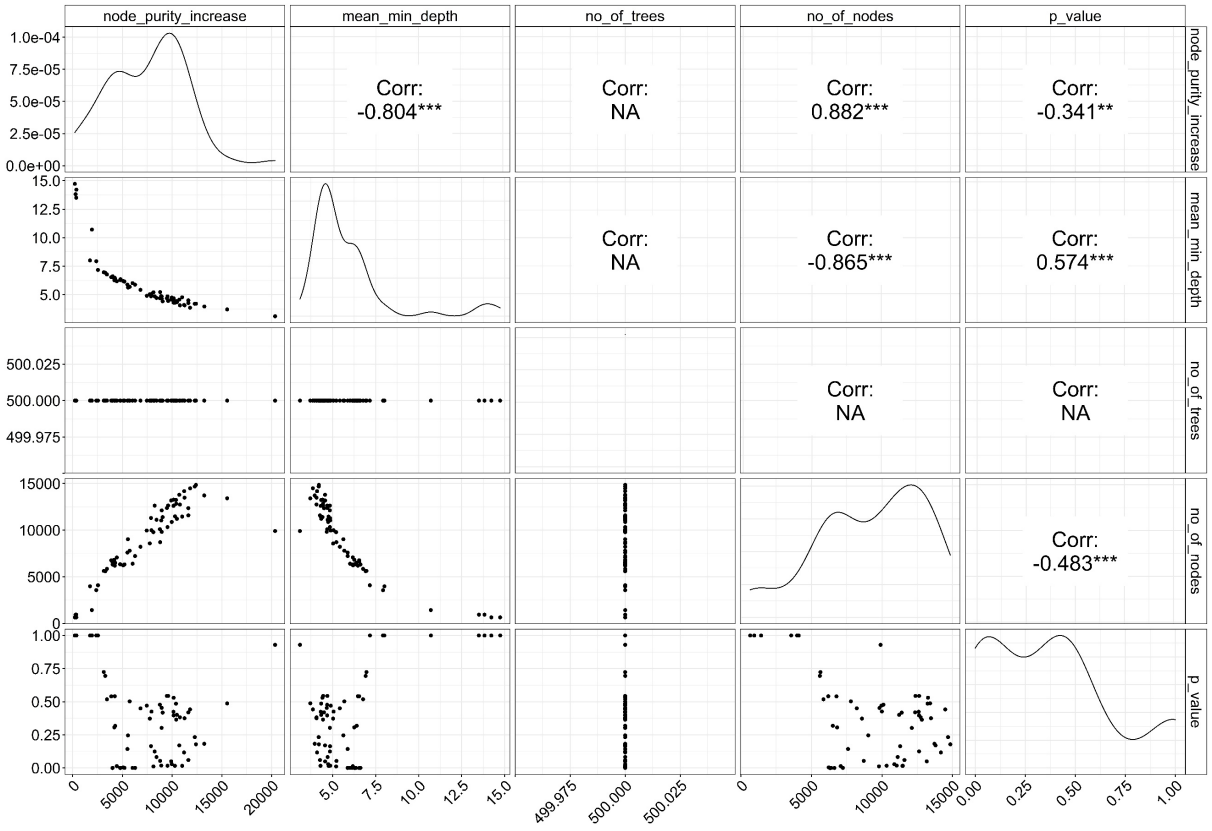

Figure S13. Bilateral relationships between measures of importance of predictors used in random forest to predict SOC stock in the 5-15 cm layer, Rondônia State, Brazil.

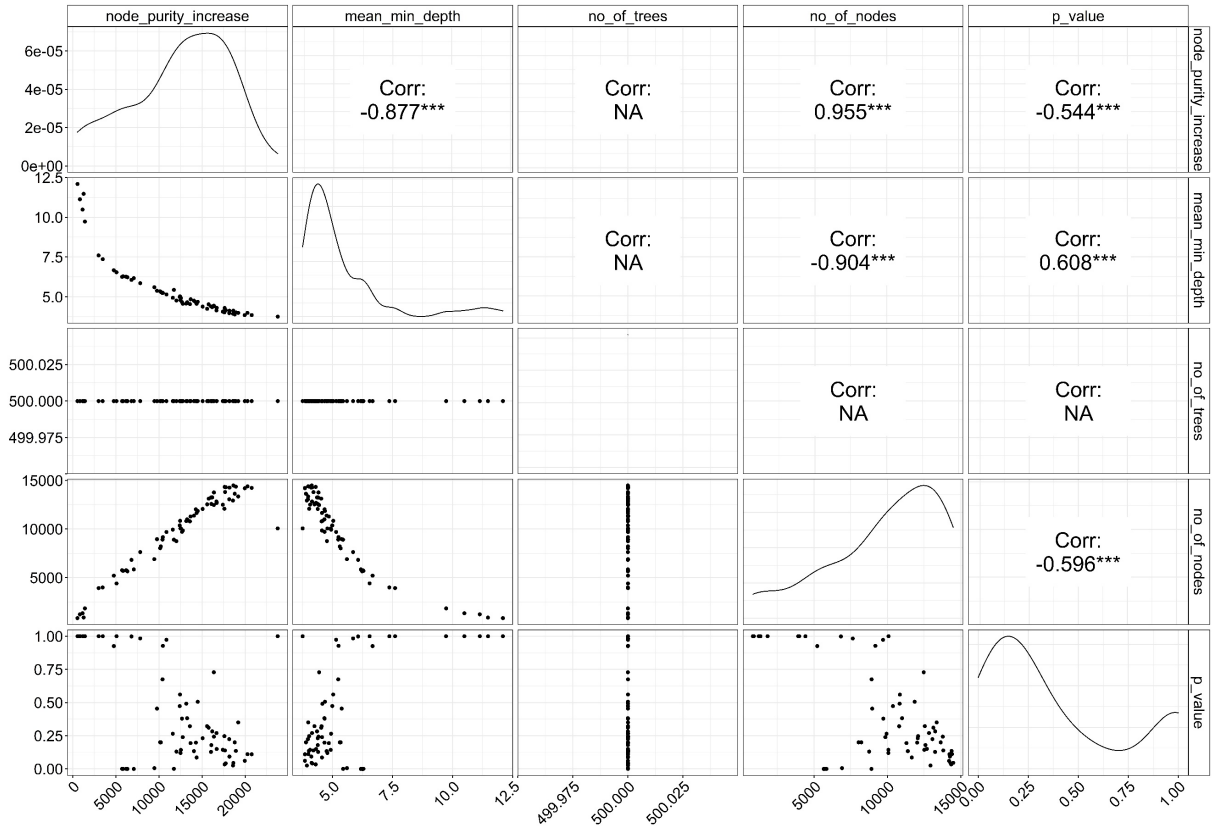

Figure S14. Bilateral relationships between measures of importance of predictors used in random forest to predict SOC stock in the 30-60 cm layer, Rondônia State, Brazil.

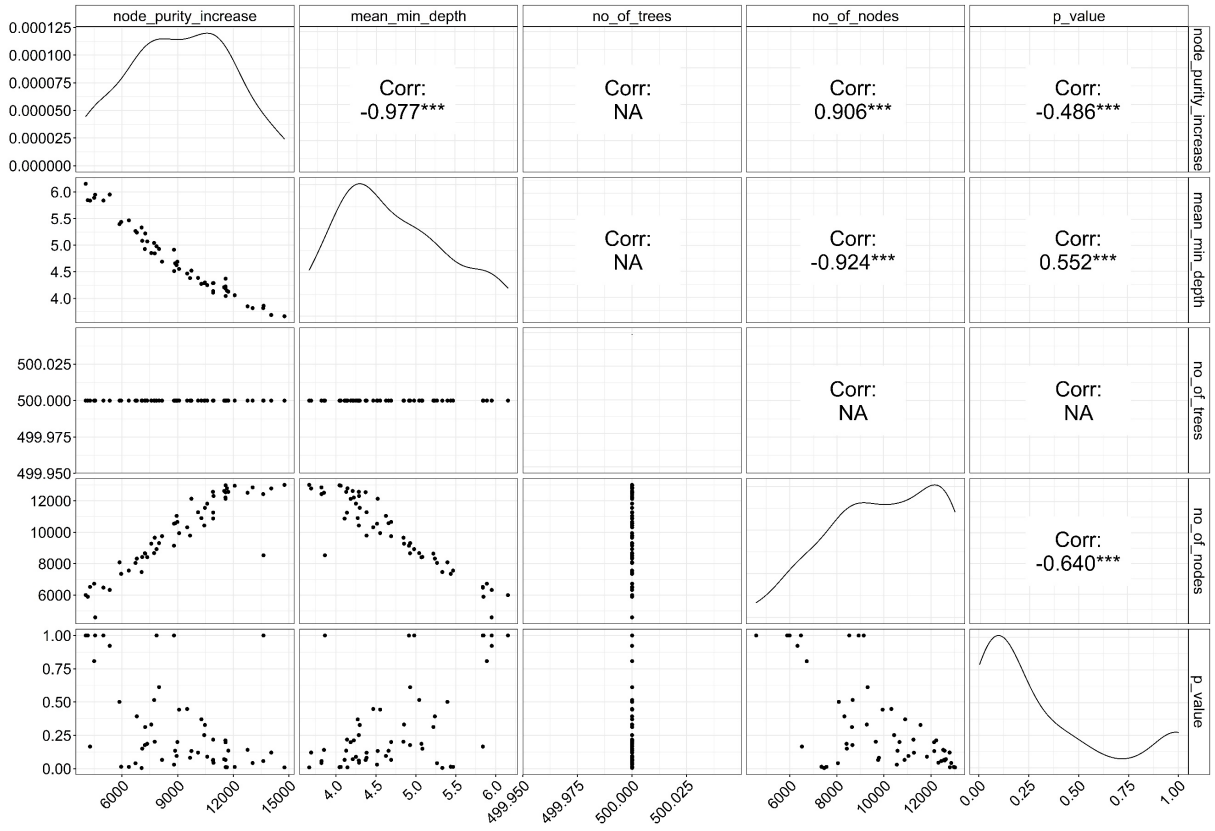

Figure S15. Bilateral relationships between measures of importance of predictors used in random forest to predict SOC stock in the 0-5 cm layer, Rondônia State, Brazil.

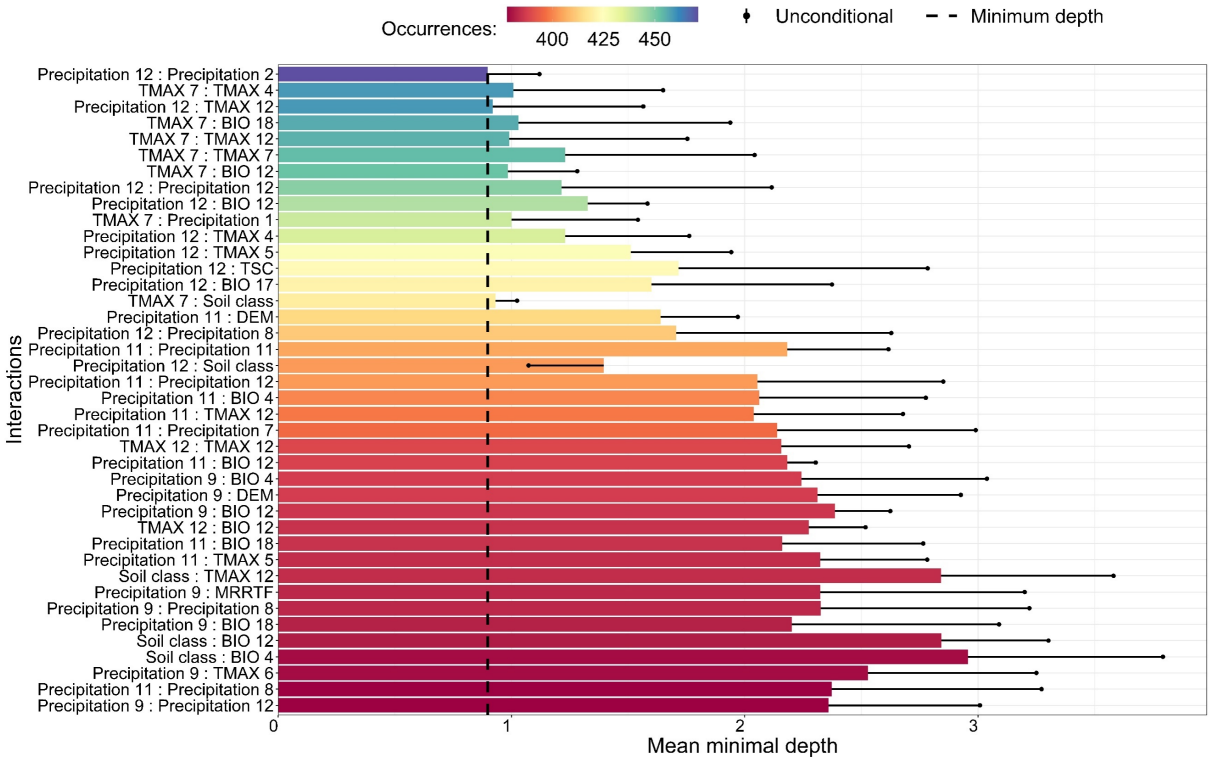

Figure S16. Importance of predictors measured by mean minimum depth for 40 most frequent interactions found in random forest relative to prediction of SOC stock at depth of 0-5 cm, Rondônia State, Brazil.

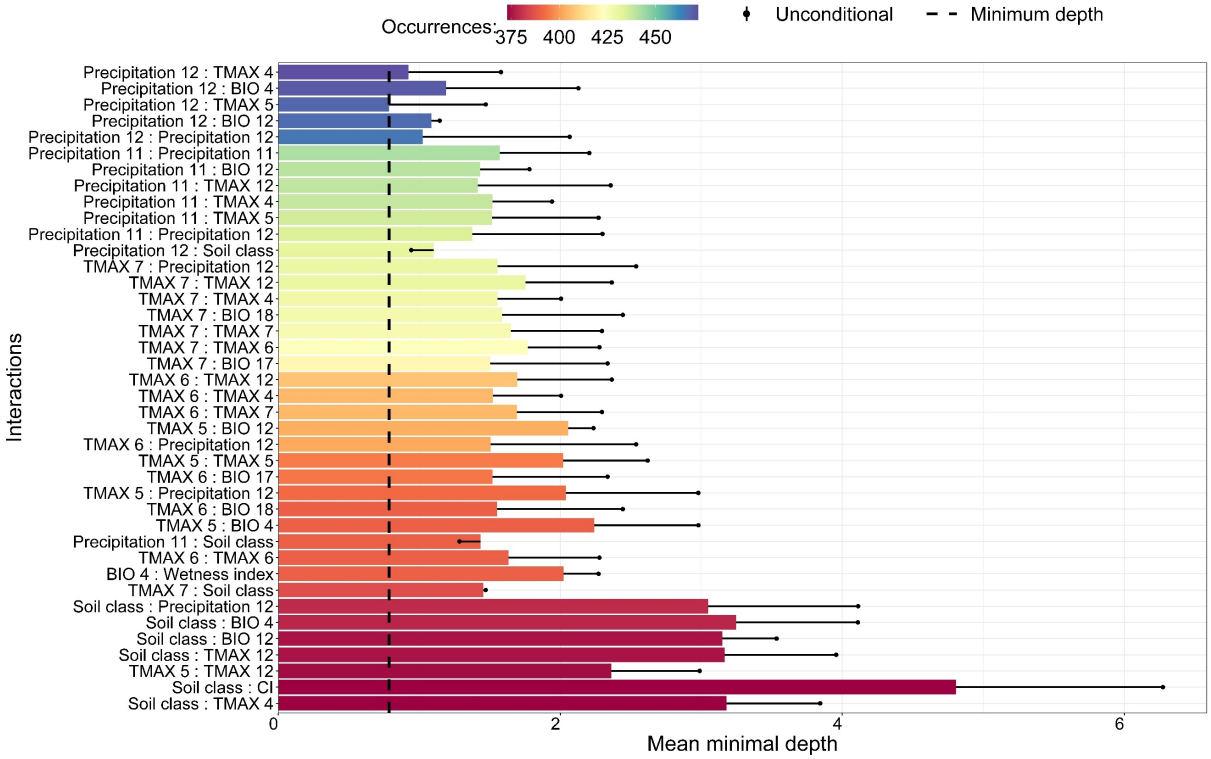

Figure S17. Importance of predictors measured by mean minimum depth for 40 most frequent interactions found in random forest relative to prediction of SOC stock at depth of 5-15 cm, Rondônia State, Brazil.

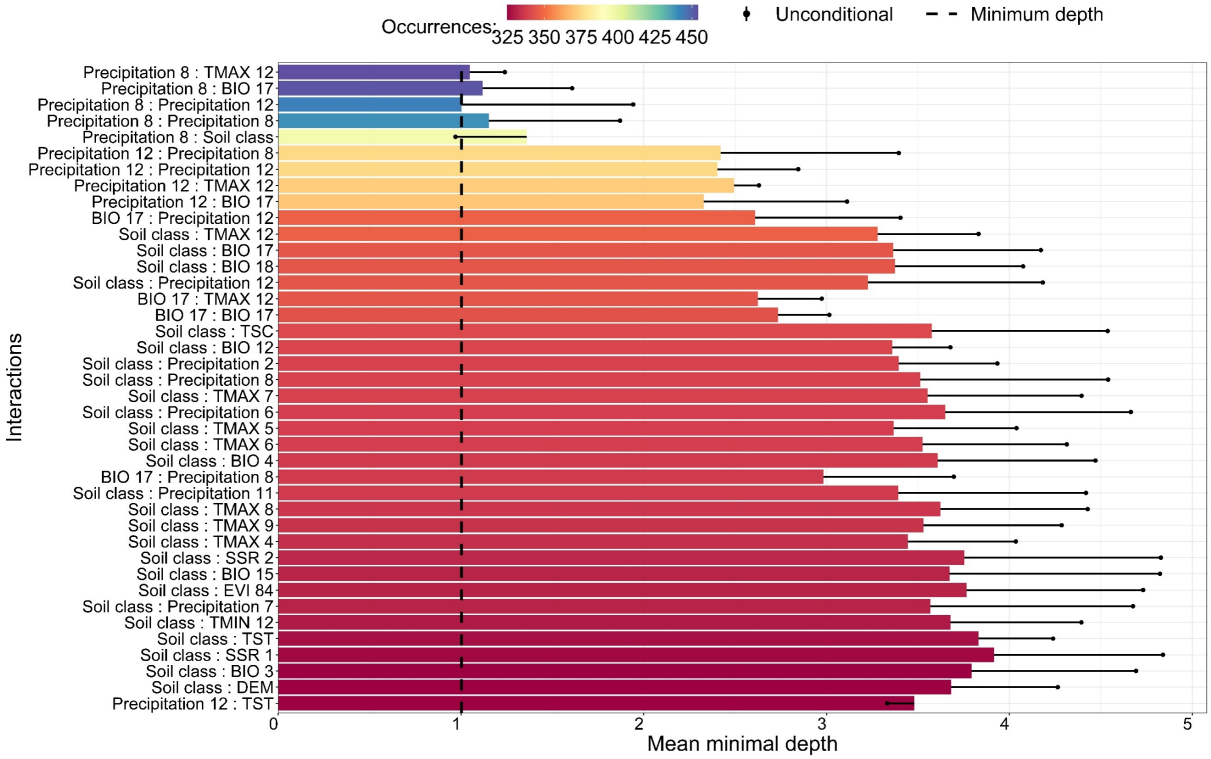

Figure S18. Importance of predictors measured by mean minimum depth for 40 most frequent interactions found in random forest relative to prediction of SOC stock at depth of 15-30 cm, Rondônia State, Brazil.

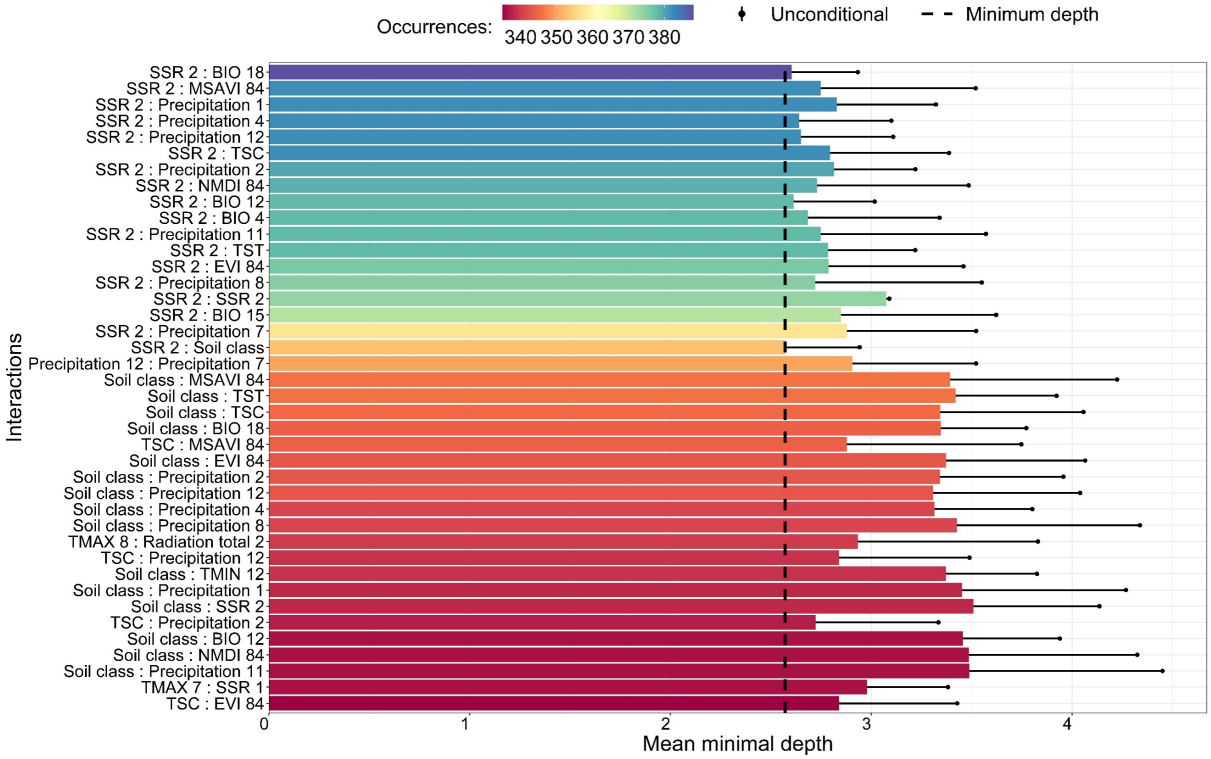

Figure S19. Importance of predictors measured by mean minimum depth for 40 most frequent interactions found in random forest relative to prediction of SOC stock at depth of 30-60 cm, Rondônia State, Brazil.

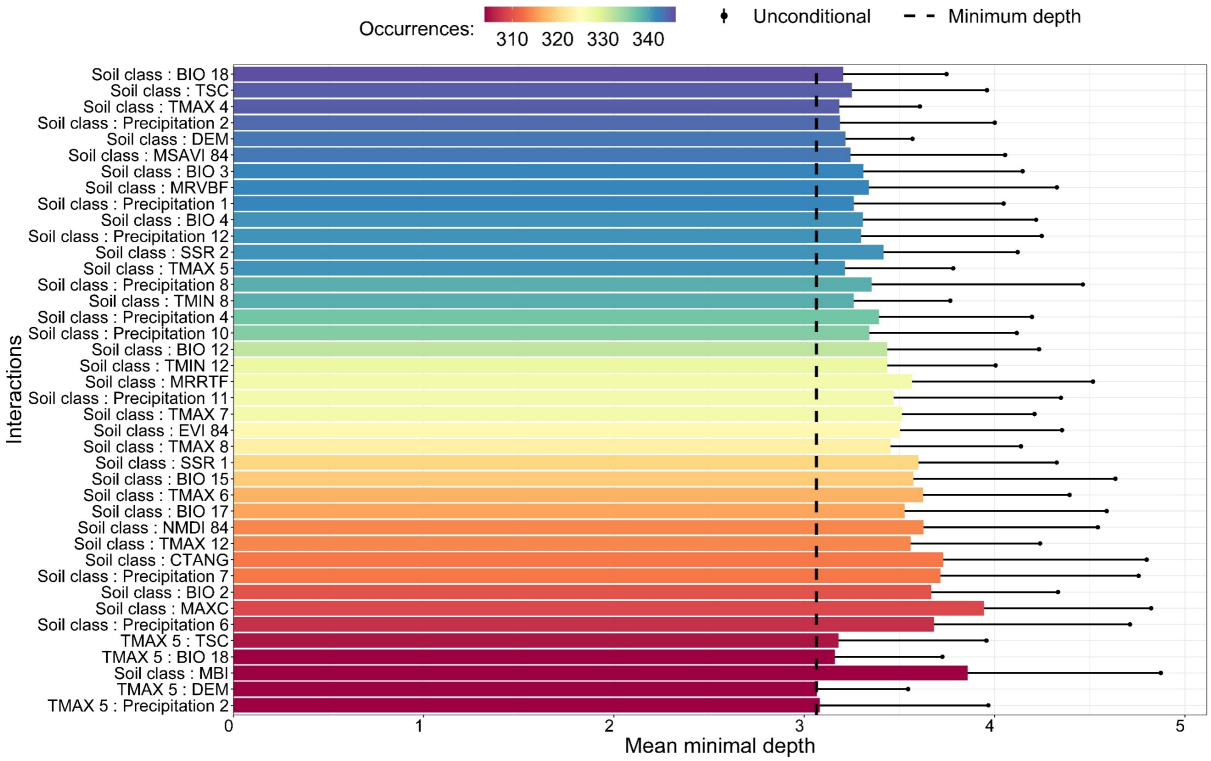

Figure S20. Importance of predictors measured by mean minimum depth for 40 most frequent interactions found in random forest relative to prediction of SOC stock at depth of 5-15 cm, Rondônia State, Brazil.

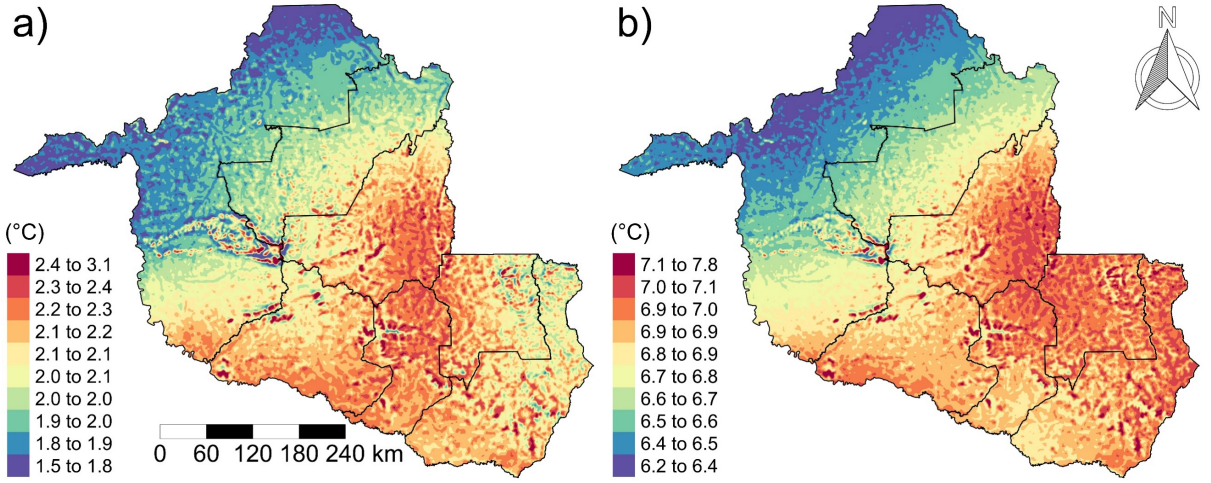

Figure S21. Difference in April maximum temperature - TMAX 4 between the average of the CanESM5 and MIROC6 models for 2100 in the ssp126 (a) and ssp585 (b) scenarios in relation to the present (TMAX 4 future - TMAX 4 present), Rondônia State, Brazil. The black lines represent the borders of the territories described in Fig. 1.

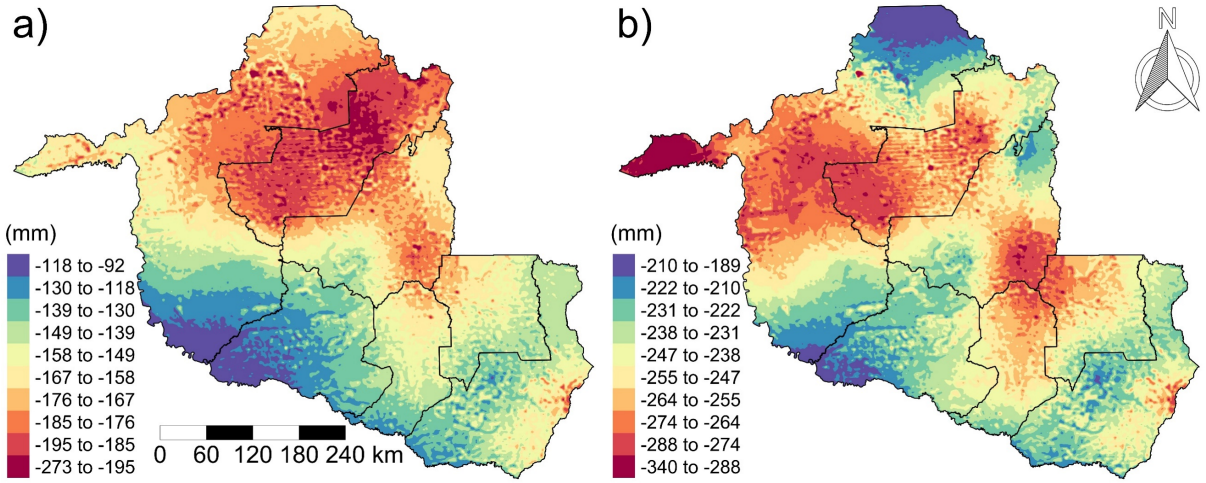

Figure S22. Difference in annual precipitation - BIO 12 between the average of the CanESM5 and MIROC6 models for 2100 in the ssp126 (a) and ssp585 (b) scenarios in relation to the present (BIO 12 future - BIO 12 present), Rondônia State, Brazil. The black lines represent the borders of the territories described in Fig. 1.

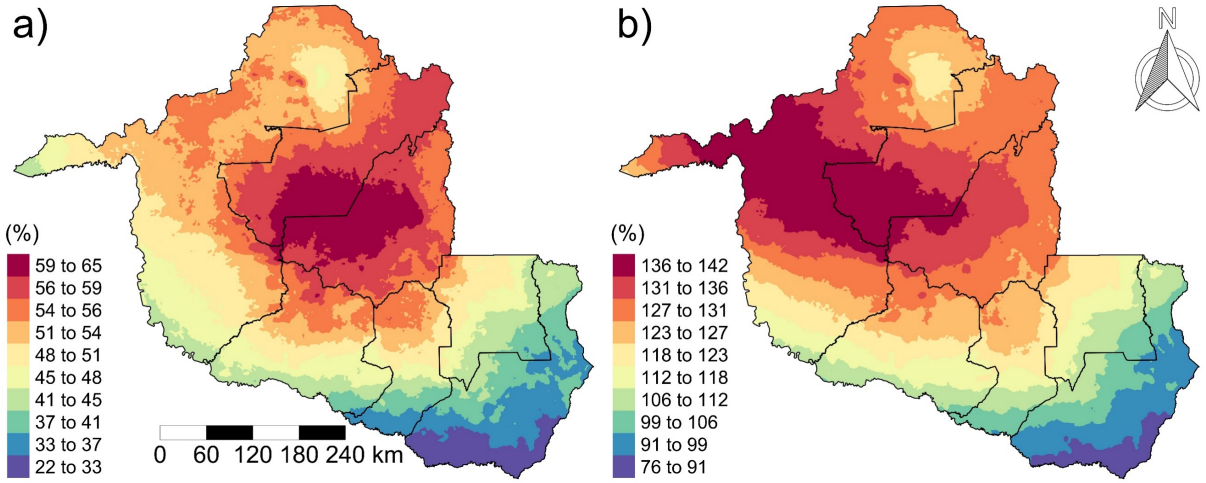

Figure S23. Difference in temperature seasonality (standard deviation × 100) - BIO 4 between the average of the CanESM5 and MIROC6 models for 2100 in the ssp126 (a) and ssp585 (b) scenarios in relation to the present (BIO 4 future - BIO 4 present), Rondônia State, Brazil. The black lines represent the borders of the territories described in Fig. 1.

Considerações finais

As técnicas de machine learning permitiram mapear o estoque de SOC de Rondônia e avaliar as incertezas. O random forest obteve a melhor performance (R^2 = 0,17 a 0,25; MAE = 4,1 a 9,6 Mg C ha^{-1}; e RMSE = 6,3 a 14,2 Mg C ha^{-1}) e foi o selecionado para as predições.

As classes de solo, temperatura e precipitação foram os principais controladores de SOC em Rondônia. O estoque de SOC no estado de Rondônia em um metro de profundidade é de aproximadamente 2.530,91 Tg. Os Latossolos e Gleissolos somados contém 70,80% do estoque total de SOC do estado. As áreas protegidas armazenam a maior densidade de SOC do estado, no entanto, o maior estoque total de SOC se encontra em áreas não protegidas. Isto reafirma a importância no fomento a sistemas de produção mais sustentáveis que sejam potencialmente capazes de conservar e/ou aumentar o estoque de SOC e mitigar as mudanças no clima.

Para 2100 as estimativas apontam que tanto para o cenário de crescimento e igualdade focado na sustentabilidade (ssp126), quanto o cenário de crescimento rápido e irrestrito na produção econômica e no uso de energia (ssp585) experimentarão perdas e ganhos no estoque de SOC. Áreas com o maior estoque podem ter redução do SOC e áreas onde existe pouco carbono poderão ter ganhos, essa dinâmica será dependente principalmente da precipitação.

A dinâmica de perdas e ganhos de SOC será mais pronunciada no cenário ssp585. Nos dois modelos climáticos utilizados (CanESM5 e MIROC6) as diferenças de impacto no SOC serão nas áreas com mais carbono, sendo o CanESM5 com as maiores variações. A comparação do modelo ajustado no trabalho e outros modelos com os dados observados conferiram uma validação adicional ao modelo e aos mapas de estoque de SOC gerados para Rondônia.