

UNIVERSIDADE FEDERAL DE VIÇOSA

**Inteligência artificial aplicada à dimensão pedológica global para modelar,
mapear e compreender interações biota-solo-clima e dinâmicas do carbono
orgânico**

Cássio Marques Moquedace dos Santos
Doctor Scientiae

**VIÇOSA - MINAS GERAIS
2025**

CÁSSIO MARQUES MOQUEDACE DOS SANTOS

**Inteligência artificial aplicada à dimensão pedológica global para modelar,
mapear e compreender interações biota-solo-clima e dinâmicas do carbono
orgânico**

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

Orientador: Elpidio I. Fernandes Filho

Coorientadores: Lucas de Carvalho
Gomes
Rodrigo Vargas Ramos

**VIÇOSA - MINAS GERAIS
2025**

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

T

S237i
2025

Moquedace, Cássio Marques, 1991-

Inteligência artificial aplicada à dimensão pedológica global para modelar, mapear e compreender interações biota-solo-clima e dinâmicas do carbono orgânico / Cássio Marques Moquedace. – Viçosa, MG, 2025.

1 tese eletrônica (251 f.): il. (algumas color.).

Orientador: Elpídio Inácio Fernandes Filho.

Tese (doutorado) - Universidade Federal de Viçosa, Departamento de Solos, 2025.

Inclui bibliografia.

DOI: <https://doi.org/10.47328/ufvbbt.2025.262>

Modo de acesso: World Wide Web.

1. Aprendizado de máquina. 2. Fungos do solo.
3. Respiração do solo. 4. Raízes (Botânica). 5. Estrutura do solo.
I. Fernandes Filho, Elpídio Inácio, 1963-. II. Universidade Federal de Viçosa. Departamento de Solos. Programa de Pós-Graduação em Solos e Nutrição de Plantas. III. Título.

CDD 22. ed. 631.4

CÁSSIO MARQUES MOQUEDACE DOS SANTOS

Inteligência artificial aplicada à dimensão pedológica global para modelar, mapear e compreender interações biota-solo-clima e dinâmicas do carbono orgânico

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

APROVADA: 15 de maio de 2025.

Assentimento:

Cássio Marques Moquedace dos Santos
Autor

Elpidio Inacio Fernandes Filho
Orientador

Essa tese foi assinada digitalmente pelo autor em 11/08/2025 às 15:11:06 e pelo orientador em 12/08/2025 às 11:39:23. As assinaturas têm validade legal, conforme o disposto na Medida Provisória 2.200-2/2001 e na Resolução nº 37/2012 do CONARQ. Para conferir a autenticidade, acesse <https://siadoc.ufv.br/validar-documento>. No campo 'Código de registro', informe o código **P9HJ.B21R.CEMG** e clique no botão 'Validar documento'.

À minha mãe, Maria Suely dos Santos Marques, e ao meu pai, Domingos Marques Neto (*in memoriam*), que me ensinaram, com exemplo e dignidade, o valor da humildade, da honestidade e da persistência.

À minha esposa, Clara Glória Oliveira Baldi, por sua presença firme e amorosa em todos os momentos, especialmente nos mais difíceis. Sua parceria foi força, refúgio e inspiração ao longo dessa jornada.

AGRADECIMENTOS

Talvez essa parte seja a mais difícil de escrever. Apesar de toda a estatística, modelagem e inteligência artificial presentes nesta tese, a verdade é que a vida não se submete a análise de regressão. E definitivamente não é o valor de p que mede a significância das pessoas nesta trajetória. A autoria pode ser individual, mas o esforço é coletivo, direto ou indireto, visível ou silencioso. Dito isso, agradeço... À minha mãe, Maria Suely dos Santos Marques, meu alicerce. Mulher de origem simples, que não teve a oportunidade de concluir o ensino superior, mas que nunca deixou de acreditar no poder transformador da educação. Foi com sua força, coragem e sabedoria que atravessei os momentos mais difíceis. Ela não pisou numa universidade como aluna, mas hoje vê seu filho concluir o doutorado. Isso não é apenas uma conquista minha. É também a realização de um sonho dela. Obrigado por tudo, inclusive pelo que nunca saberei nomear.

À minha esposa, Clara Glória Oliveira Baldi, amor da minha vida, minha fortaleza serena e constante. Obrigado por estar ao meu lado em cada madrugada difícil, em cada ausência forçada, em cada plano que foi adiado pelo peso das demandas acadêmicas. Nada disso teria sentido sem você. Teu amor me sustentou quando a mente cansava, quando o corpo pedia trégua, quando a esperança oscilava. Com você aprendi que o amor verdadeiro não é feito de palavras grandes, mas de gestos diários, de silêncio compreendido, de presença no meio do caos. Essa conquista também é sua. É nossa.

À minha tia, Solange Mendes Moquedace dos Santos, exemplo de dedicação à educação pública. Seu apoio constante e sua trajetória sempre me inspiraram. Ao meu orientador, professor Elpídio Inácio Fernandes Filho, por sua orientação comprometida, sua escuta atenta e por confiar na minha autonomia. Agradeço pelas conversas sinceras, pela disponibilidade incansável e por me mostrar que ciência também se faz com afeto e coerência.

À professora Irene Maria Cardoso, que mesmo sem título formal nesta etapa, foi quem me trouxe até aqui. Foi por sua inspiração que fui para Viçosa atrás de um sonho. Mesmo trilhando outros caminhos, carrego com orgulho o privilégio de tê-la como referência, e, acima de tudo, como amiga. Seu exemplo de ética, humanidade e compromisso com o coletivo moldou não só meu modo de fazer ciência, mas também meu modo de estar no mundo.

Aos professores Carlos Ernesto Gonçalves Schaefer e Márcio Rocha

Francelino, pelas contribuições fundamentais à minha formação. Obrigado pela disponibilidade, pelas orientações técnicas e pelas conversas sempre diretas e produtivas. Suas posturas firmes, comprometidas e éticas foram, e seguem sendo, inspiração para o exercício da ciência com propósito.

Agradeço ao Lucas Carvalho Gomes, que conheci ainda no mestrado e que desde então tem sido presença constante nesta trajetória, mesmo agora a quilômetros de distância, na Dinamarca. Sua ajuda foi decisiva em muitas etapas deste trabalho, com generosidade, conhecimento técnico e, acima de tudo, amizade verdadeira. Lucas foi mais que coorientador. Foi companheiro de caminhada, alguém em quem confiei e continuo confiando. A geografia pode mudar, mas o vínculo permanece firme. Obrigado por seguir comigo, em tantas formas, até aqui.

Aos colegas do Laboratório de Pedometria e Geoprocessamento (LabGeo), onde cresci como pesquisador. Foi ali, entre mapas, debates e cafézinhos, que aprendi que fazer ciência exige tanto rigor quanto sensibilidade.

Aos amigos que a vida acadêmica me deu, companheiros de estrada, de café e de luta. Em especial, Rafael Gomes Siqueira, Luís Flávio Pereira, Rugana Imbana, Lucas Eduardo Realto da Silva e Arthur da Cruz Silva, por dividirem comigo o peso e a beleza dessa caminhada.

A todos os professores, técnicos e funcionários da Universidade Federal de Viçosa, pelo suporte e pela dedicação ao ensino público de qualidade, que transforma vidas como a minha.

Aos agricultores e agricultoras que contribuíram com a pesquisa e que todos os dias cuidam da terra. Vocês são parte essencial da ciência que se conecta com a realidade.

Este trabalho foi realizado com o apoio das seguintes agências de pesquisa brasileiras: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Por fim, agradeço à ciência. Àquela com os pés no chão, feita com propósito, ética e compromisso. A ciência que se faz com gente e para gente.

"Olhe novamente para esse ponto. É aqui. É o lar. Somos nós." (Carl Sagan, em
Pale Blue Dot, 1994)

RESUMO

SANTOS, Cássio Marques Moquedace dos, D.Sc., Universidade Federal de Viçosa, maio de 2025. **Inteligência artificial aplicada à dimensão pedológica global para modelar, mapear e compreender interações biota-solo-clima e dinâmicas do carbono orgânico.** Orientador: Elpidio Inacio Fernandes Filho. Coorientadores: Lucas de Carvalho Gomes e Rodrigo Vargas Ramos.

O solo é um componente central da biosfera terrestre, sustentando funções ecológicas críticas como a produção de alimentos, a regulação de ciclos biogeoquímicos e o sequestro de carbono. Entretanto, sua complexidade funcional, especialmente nos compartimentos subterrâneos, tem sido historicamente negligenciada em modelos globais de carbono. Essa lacuna é particularmente crítica diante das mudanças climáticas, que podem converter o solo de sumidouro a fonte de carbono. Neste contexto, objetivou-se com esta pesquisa desenvolver uma abordagem funcional e explicável, baseada em inteligência artificial, para modelar e compreender a dinâmica do carbono orgânico do solo (SOC) em escala global, com foco nos preditores biológicos: fungos, raízes e respiração. A pesquisa foi estruturada em quatro artigos, todos ancorados no uso do algoritmo quantile random forest (QRF) como estratégia de modelagem preditiva. Essa abordagem permitiu não apenas estimar valores medianos com robustez, mas também quantis e intervalos de incerteza. O primeiro investigou a distribuição global da biodiversidade fúngica do solo, utilizando dados da base GSM e algoritmos de machine learning; o segundo mapeou traços funcionais de raízes a partir de bancos de dados globais integrados; o terceiro quantificou as taxas atuais e futuras de respiração do solo (R_s e R_h) e seus preditores edafoclimáticos; e o quarto integrou variáveis edáficas, biológicas e climáticas em um modelo explicável de SOC, utilizando valores de Shapley, curvas de dependência parcial e mapas de dominância funcional para a interpretação dos resultados. Os resultados revelaram que a distribuição de fungos mostrou-se altamente heterogênea, com baixa representatividade em regiões tropicais e áridas, onde sua diversidade ou especialização funcional é mais crítica. As raízes apresentaram forte controle sobre a estruturação da rizosfera e a ciclagem de nutrientes, com destaque para a colonização micorrízica como preditor-chave em regiões tropicais. A modelagem da respiração evidenciou o papel moderador da textura do solo: argilas protegeram o SOC e amorteceram as respostas ao aquecimento, enquanto solos arenosos acentuaram a vulnerabilidade climática. Finalmente, os modelos de SOC revelaram interações não lineares e contextuais entre biota, clima e solo, com padrões espaciais distintos de dominância

preditiva. Os achados evidenciaram a fragilidade de modelos tradicionais que ignoram a biota do solo e dependem unicamente de proxies climáticos ou vegetacionais baseados em sensoriamento remoto. A introdução de variáveis funcionais e abordagens explicáveis permitiu revelar hierarquias e sinergias entre os fatores que controlam o carbono, apontando para um novo paradigma metodológico: o *Soil-Informed Machine Learning*. Os resultados mostraram que as condições que promovem o acúmulo de SOC em certos contextos (e.g., alta colonização micorrízica em solos argilosos úmidos) podem ter efeitos opostos em outros (e.g., solos arenosos secos). Evidenciou-se também que as incertezas da modelagem estão fortemente associadas à sub-representação de dados em regiões críticas, como zonas tropicais, boreais e semiáridas. Esta pesquisa amplia o entendimento sobre a dimensão pedoecológica do carbono no solo, ao integrar processos subterrâneos, interações biológicas e conhecimento edafológico em modelos globalmente consistentes e interpretáveis. A inteligência artificial explicável mostrou-se uma ferramenta promissora para decifrar relações complexas e orientar políticas mais robustas de conservação e mitigação. No entanto, avanços duradouros dependerão da consolidação de bancos de dados harmonizados, de protocolos de ciência aberta e da valorização institucional do solo como componente central das estratégias climáticas globais.

Palavras-chave: aprendizado de máquina; pedometria; fungos do solo; traços radiculares; respiração do solo

ABSTRACT

SANTOS, Cássio Marques Moquedace dos, D.Sc., Universidade Federal de Viçosa, May, 2025. **Artificial intelligence applied to the global pedological dimension to model, map and understand biota-soil-climate interactions and organic carbon dynamics.** Adviser: Elpidio Inacio Fernandes Filho. Co-advisers: Lucas de Carvalho Gomes and Rodrigo Vargas Ramos.

Soil is a central component of the terrestrial biosphere, supporting critical ecological functions such as food production, regulation of biogeochemical cycles, and carbon sequestration. However, its functional complexity—particularly in the subsurface compartments—has historically been neglected in global carbon models. This gap is especially critical in the face of climate change, which may shift soils from carbon sinks to sources. In this context, the objective of this research was to develop a functional and explainable approach, based on artificial intelligence, to model and understand the dynamics of soil organic carbon (SOC) at a global scale, with a focus on biological predictors: fungi, roots, and respiration. The research was structured into four complementary studies, all anchored in the use of the quantile random forest (QRF) algorithm as a predictive modeling strategy. This approach enabled robust estimation not only of median values but also of quantiles and uncertainty intervals. The first study investigated the global distribution of soil fungal biodiversity using GSM database records and machine learning algorithms; the second mapped functional root traits using integrated global datasets; the third quantified current and future rates of soil respiration (R_s and R_h) and their edaphoclimatic predictors; and the fourth integrated soil, biological, and climatic variables into an explainable SOC model, using Shapley values, partial dependence plots, and functional dominance maps to interpret the results. The findings revealed that fungal distribution was highly heterogeneous, with poor representation in tropical and arid regions—precisely where their diversity and functional specialization are most critical. Roots exerted strong control over rhizosphere structure and nutrient cycling, with mycorrhizal colonization emerging as a key predictor in tropical environments. Soil respiration modeling highlighted the moderating role of soil texture: clay soils protected SOC and buffered warming responses, whereas sandy soils amplified climatic vulnerability. Finally, SOC models uncovered nonlinear and context-dependent interactions between biota, climate, and soil, with distinct spatial patterns of predictive dominance. These results exposed the fragility of traditional models that ignore soil biota and rely solely on climatic or vegetation proxies derived from remote sensing.

The inclusion of functional variables

and explainable modeling approaches revealed hierarchies and synergies among carbon-controlling factors, pointing to a new methodological paradigm: Soil-Informed Machine Learning. The findings also showed that conditions promoting SOC accumulation in certain contexts (e.g., high mycorrhizal colonization in moist clay-rich soils) may have opposite effects in others (e.g., dry sandy soils). Moreover, modeling uncertainties were strongly associated with data underrepresentation in critical regions, such as tropical, boreal, and semi-arid zones. This research advances our understanding of the pedoecological dimension of soil carbon by integrating subsurface processes, biological interactions, and edaphological knowledge into globally consistent and interpretable models. Explainable artificial intelligence proved to be a promising tool to disentangle complex relationships and guide more robust conservation and mitigation policies. Nonetheless, long-term progress will depend on the consolidation of harmonized databases, open science protocols, and the institutional recognition of soil as a central component of global climate strategies

Keywords: machine learning; pedometry; soil fungi; root traits; soil respiration

SUMÁRIO

Introdução geral.....	14
Referências	18
Artigo 1: Fungos do solo em escala global: implicações ecológicas e funcionais para a saúde do solo	24
1. Introdução.....	25
2. Material e métodos	28
2.1. Base de dados	28
2.2. Camadas de preditores ambientais	30
2.3. Modelagem geoespacial	34
2.3.1. Processos de modelagem e mapeamento.....	34
2.3.2. Seleção de variáveis.....	34
2.3.3. Ajuste do modelo e predição dos mapas	35
2.3.4. Importância dos preditores	38
3. Resultados.....	39
3.1. Banco de dados e modelagem	39
3.2. Distribuição espacial da abundância e riqueza de fungos no solo.....	41
3.3. Moduladores de fungos no solo e a interação entre riqueza e abundância.....	46
4. Discussão.....	48
4.1. Distribuição espacial de fungos no solo	48
4.2. Incertezas espaciais na distribuição de fungos no solo.....	50
4.3. Estratificação dos fungos no solo e seus principais controladores.....	51
4.4. Interação entre abundância e riqueza de fungos no solo	54
4.5. Limitação e futuras direções sobre mapeamento da biodiversidade do solo em escala global	57
5. Conclusão	58
Referências	59
Artigo 2: Mapeamento global de traços funcionais de raízes: insights sobre dinâmicas edáficas e climáticas	74
1. Introdução.....	75
2. Material e métodos	78
2.1. Banco de dados.....	78
2.2. Preditores ambientais.....	80
2.3. Modelagem geoespacial	82
2.3.1. Seleção de preditores	83
2.3.2. Ajuste do modelo e predição dos mapas	84
2.3.3. Importância dos preditores	86
3. Resultados.....	86
3.1. Distribuição dos dados e performance do modelo.....	86
3.2. Distribuição espacial de diferentes traços funcionais de raízes.....	89

3.3.	Importância dos preditores e dependência parcial.....	95
4.	Discussão.....	97
4.1.	Dados globais de alta resolução sobre traços funcionais de raízes.....	97
4.2.	Relações ambientais e traços funcionais de raízes	99
4.3.	Governança biótica e edáfica dos traços funcionais de raízes.....	103
4.4.	Limitações e perspectivas futuras.....	106
5.	Conclusão	107
	Referências	108
	Material suplementar	127
	Artigo 3: Aprendizado de máquina e dados pedológicos: perspectivas mais realistas para a modelagem da respiração do solo e feedbacks carbono-clima.....	133
1.	Introdução.....	134
2.	Material e métodos	138
2.1.	Dados de respiração do solo e variáveis explicativas.....	138
2.2.	Predição e avaliação do modelo	142
3.	Resultados.....	145
3.1.	Efeitos de overfitting em projeções de R_s e R_h	145
3.2.	R_s e R_h presentes e principais fatores	148
3.3.	Taxas futuras de respiração do solo	155
3.4.	Principais fatores que impulsionam as mudanças futuras em R_s e R_h	160
4.	Discussão.....	163
4.1.	Restrições e fatores desencadeantes da respiração do solo em ecorregiões no futuro 163	
4.2.	Importância das propriedades do solo para a respiração do solo	167
4.3.	Desafios atuais e futuras direções na modelagem da respiração do solo	170
5.	Conclusão	171
	Referências	172
	Material Suplementar	188
	Artigo 4: Interações solo–biota–clima e explicabilidade em modelos globais de SOC.....	191
1.	Introdução.....	192
2.	Material e métodos	196
2.1.	Base de dados	196
2.2.	Preditores explicativos do estoque de SOC	198
2.3.	Modelagem espacial	199
2.4.	Contribuição dos preditores ambientais para o estoque de SOC.....	201
3.	Resultados.....	203
3.1.	Viés amostral e implicações para a modelagem global	203
3.2.	Acurácia do modelo e limitações na predição de extremos	204

3.3.	Padrões geográficos e fatores ambientais controladores do SOC	205
3.4.	Interações solo–biota–clima e a importância funcional dos preditores	207
3.4.1.	Importância global dos preditores ambientais	207
3.4.2.	Respostas funcionais às variáveis ambientais	208
3.4.3.	Padrões espaciais de dominância de importância com base nos valores de Shapley	212
4.	Discussão	217
4.1.	Viés amostral e implicações para a modelagem global	217
4.2.	Acurácia do modelo e limitações na predição de extremos	218
4.3.	Padrões geográficos e fatores ambientais que controlam o SOC	219
4.4.	Interações solo–biota–clima e a importância funcional dos preditores	220
4.4.1.	Importância global dos preditores ambientais	220
4.4.2.	Respostas funcionais às variáveis ambientais	222
4.4.3.	Padrões espaciais de dominância de importância com base nos valores de Shapley	224
4.5.	Inteligência artificial explicável aplicada à ciência do solo	228
4.6.	Limitações e perspectivas para o avanço da modelagem global de carbono no solo	230
5.	Conclusão	233
	Referências	234
	Conclusão geral	250

1 **Introdução geral**

2 O solo constitui a base funcional da estabilidade ecológica terrestre e socioeconômica
3 global, sustentando aproximadamente 95% da produção alimentar (FAO, 2015) e
4 desempenhando papel vital na regulação dos ciclos da água, do carbono e dos nutrientes
5 (Delgado-Baquerizo et al., 2025; Zomer et al., 2017). Longe de ser apenas um suporte inerte
6 para o crescimento vegetal, é um ecossistema dinâmico, habitado por uma biodiversidade
7 significativa que impulsiona processos ecológicos fundamentais e assegura serviços ambientais
8 essenciais para a sobrevivência humana (Kopittke et al., 2022). Essa multifuncionalidade torna
9 o solo um ativo estratégico não apenas para a agricultura, mas também para a conservação da
10 biodiversidade e para a mitigação das mudanças climáticas. Contudo, apesar de sua centralidade
11 no funcionamento do sistema terrestre, o solo ainda é subvalorizado nas agendas ambientais e
12 políticas, o que contribui para sua contínua degradação em escala planetária (Pozza & Field,
13 2020).

14 Entre os múltiplos serviços ecossistêmicos que o solo oferece, destaca-se a sua
15 capacidade de armazenamento de carbono, abrigando entre 1.460 e 1.550 Pg de carbono
16 orgânico, quantidade superior à do carbono presente na atmosfera ou na biomassa viva (Batjes,
17 2006; Lal, 2004). A manutenção desse reservatório subterrâneo é crítica para a regulação
18 climática, pois perturbações podem converter o solo de sumidouro em fonte expressiva de
19 carbono, retroalimentando o aquecimento global (Lal et al., 2007).

20 No entanto, aproximadamente 33% dos solos globais já apresentam algum grau de
21 degradação significativa (Delgado-Baquerizo et al., 2025), impulsionada por práticas de manejo
22 insustentáveis, desmatamento, intensificação agrícola e eventos climáticos extremos (Lal,
23 2014; Pozza & Field, 2020). A degradação compromete a capacidade de fixação de carbono,
24 reduz a produtividade e intensifica processos de desertificação. Dada a lentidão da formação de
25 novos solos (Whenua, 2025), a sua perda representa uma ameaça quase irreversível para a

26 estabilidade ecológica e climática (Delgado-Baquerizo et al., 2021; Eldridge et al., 2017; Qu et
27 al., 2024). Preservar o solo é, portanto, uma estratégia de longo prazo.

28 Compreender o solo como um ecossistema vivo exige adentrar à escala de seus
29 principais motores biológicos (Tedersoo et al., 2021; Yamauchi et al., 2021). Fungos e raízes
30 desempenham funções centrais na dinâmica do carbono do solo (Emilia Hannula & Morriën,
31 2022; G. Wang et al., 2023; X. Wang et al., 2024), enquanto a respiração do solo atua como um
32 indicador sensível da atividade biológica e da permanência dos estoques subterrâneos (Bond-
33 Lamberty et al., 2004; Bond-Lamberty & Thomson, 2010). Fungos filamentosos por exemplo,
34 promovem a formação e estabilização de agregados, protegendo fisicamente o carbono contra
35 a decomposição microbiana (Goede et al., 2025). Além disso, associações simbióticas, como as
36 micorrízicas, ampliam a eficiência da aquisição de nutrientes, especialmente em ambientes de
37 baixa fertilidade química (Freschet, Roumet, et al., 2021; Van Der Heijden et al., 2008).

38 As raízes, por sua vez, exibem elevada plasticidade funcional, modulando a aquisição
39 de recursos, a estabilidade física do solo e a dinâmica do carbono em resposta a variações
40 ambientais (Yaffar et al., 2024). Traços funcionais como densidade radicular, comprimento
41 específico e capacidade de formar simbioses refletem estratégias adaptativas que moldam o
42 funcionamento dos ecossistemas (Freschet, Pagès, et al., 2021). Complementando essa
43 atividade subterrânea, a respiração do solo integra a contribuição de raízes, microrganismos e
44 fungos simbióticos, liberando anualmente entre 75 e 100 Pg de carbono para a atmosfera (Bond-
45 Lamberty & Thomson, 2010). Esse fluxo representa o maior movimento natural de carbono da
46 biosfera terrestre para a atmosfera, funcionando como termômetro da vitalidade dos processos
47 subterrâneos (Elberling et al., 2013; Ren et al., 2024; von Lützow & Kögel-Knabner, 2009).

48 Cada um desses componentes — fungos, raízes e respiração — pode ser abordado como
49 variável funcional chave para explicar a dinâmica do carbono do solo, refletindo diferentes
50 aspectos da permanência e da vulnerabilidade desse reservatório. Embora haja compreensão

51 consolidada em nível ecológico, a sua incorporação explícita nos modelos globais de predição
52 do carbono do solo ainda é incipiente (Scharlemann et al., 2014; Thorsøe et al., 2023).

53 Apesar da relevância dos processos biológicos subterrâneos, os modelos tradicionais de
54 carbono do solo falham em representá-los adequadamente (Hengl et al., 2017; Poggio et al.,
55 2021). Persistem limitações históricas graves, como o viés amostral que privilegia dados
56 provenientes de regiões temperadas, em detrimento das zonas tropicais, boreais e semiáridas,
57 justamente onde se encontram grandes estoques de carbono (Scharlemann et al., 2014). Além
58 da desigualdade na cobertura amostral, as estruturas preditivas convencionais são
59 frequentemente simplificadas, baseando-se apenas em covariáveis climáticas e proxies de
60 vegetação advindos de sensoriamento remoto, enquanto negligenciam a diversidade funcional
61 de fungos e raízes e outros processos que governam o armazenamento de carbono (Amelung et
62 al., 2020; Padarian et al., 2020; Thorsøe et al., 2023).

63 O recente avanço das técnicas de machine learning na ciência do solo abriu novas
64 possibilidades para superar essas limitações (Padarian et al., 2020; Wadoux, 2025). A
65 combinação entre grandes volumes de dados ambientais, algoritmos sofisticados e maior
66 capacidade computacional permitiu ganhos notáveis em acurácia preditiva, sobretudo na
67 modelagem espacial de estoques de carbono (Heuvelink et al., 2021). No entanto, o uso
68 indiscriminado de modelos como “caixa-preta”, que priorizam a performance preditiva em
69 detrimento da interpretabilidade, gerou preocupações legítimas sobre a confiabilidade
70 ecológica das projeções (Minasny et al., 2024; Padarian et al., 2020). Modelos utilizados como
71 “caixa-preta”, que oferecem alta acurácia preditiva, mas pouca transparência, correm o risco de
72 perpetuar relações espúrias ou de obscurecer os mecanismos funcionais que governam a
73 dinâmica do carbono no solo (Minasny & McBratney, 2025).

74 Diante dessas limitações, emerge a necessidade de um novo paradigma de modelagem
75 pedológica, que integre inteligência artificial explicável com conhecimento pedoecológico. A

76 abordagem conhecida como *Soil Science-Informed Machine Learning* (SoilML) propõe a
77 combinação explícita de dados observacionais, princípios edafológicos e processos biológicos
78 no treinamento dos algoritmos (Minasny et al., 2024). Essa integração busca não apenas
79 aprimorar a precisão das projeções, mas reconectar a modelagem à compreensão dos
80 mecanismos que sustentam a dinâmica do carbono no solo.

81 A aplicação de inteligência artificial explicável, nesse contexto, permite interpretar a
82 importância relativa e espacial de variáveis como traços funcionais de fungos e raízes, dinâmica
83 da respiração, variabilidade textural e mineralógica (Heuvelink et al., 2021; Wadoux, 2025;
84 Wadoux et al., 2023). Modelos orientados por esse novo paradigma têm potencial para gerar
85 projeções mais transparentes, robustas e pedoecologicamente fundamentadas, apoiando
86 estratégias de conservação e mitigação mais eficazes (Minasny et al., 2024).

87 Construir esse novo paradigma implica abandonar a dependência exclusiva de proxies
88 simplistas e reconhecer a complexidade dos sistemas edáficos, aliando ciência orientada por
89 dados ao conhecimento pedológico consolidado (Wadoux, 2025). Nesse cenário, a inteligência
90 artificial transcende sua função de mera ferramenta preditiva, tornando-se instrumento de
91 descoberta científica capaz de fundamentar hipóteses, identificar padrões emergentes e
92 consolidar uma visão da dimensão pedoecológica mais precisa do papel do solo no sistema
93 terrestre (Minasny & McBratney, 2025; Padarian et al., 2020).

94 A fim de atingir esse objetivo, a tese foi organizada em quatro artigos: (1) Fungos do
95 solo em escala global: implicações ecológicas e funcionais para a saúde do solo; (2)
96 Mapeamento global de traços funcionais de raízes: insights sobre dinâmicas edáficas e
97 climáticas; (3) Aprendizado de máquina e dados pedológicos: perspectivas mais realistas para
98 a modelagem da respiração do solo e feedbacks carbono-clima; e (4) Interações solo-biota-
99 clima e explicabilidade em modelos globais de SOC.

101 **Referências**

- 102 Amelung, W., Bossio, D., de Vries, W., Kögel-Knabner, I., Lehmann, J., Amundson, R., Bol,
103 R., Collins, C., Lal, R., Leifeld, J., Minasny, B., Pan, G., Paustian, K., Rumpel, C.,
104 Sanderman, J., van Groenigen, J. W., Mooney, S., van Wesemael, B., Wander, M., &
105 Chabbi, A. (2020). Towards a global-scale soil climate mitigation strategy. *Nature*
106 *Communications*, 11(1), 1–10. <https://doi.org/10.1038/s41467-020-18887-7>
- 107 Batjes, N. H. (2006). Organic carbon stocks in the soils of Brazil. *Soil Use and Management*,
108 21(1), 22–24. <https://doi.org/10.1111/j.1475-2743.2005.tb00102.x>
- 109 Bond-Lamberty, B., & Thomson, A. (2010). Temperature-associated increases in the global soil
110 respiration record. *Nature*, 464(7288), 579–582. <https://doi.org/10.1038/nature08930>
- 111 Bond-Lamberty, B., Wang, C., & Gower, S. T. (2004). A global relationship between the
112 heterotrophic and autotrophic components of soil respiration? *Global Change Biology*,
113 10(10), 1756–1766. <https://doi.org/10.1111/j.1365-2486.2004.00816.x>
- 114 Delgado-Baquerizo, M., Eldridge, D. J., Liu, Y. R., Liu, Z. W., Coleine, C., & Trivedi, P. (2025).
115 Soil biodiversity and function under global change. *PLoS Biology*, 23(3 March),
116 e3003093. <https://doi.org/10.1371/journal.pbio.3003093>
- 117 Delgado-Baquerizo, M., Eldridge, D. J., Liu, Y. R., Sokoya, B., Wang, J. T., Hu, H. W., He, J.
118 Z., Bastida, F., Moreno, J. L., Bamigboye, A. R., Blanco-Pastor, J. L., Cano-Díaz, C., Illán,
119 J. G., Makhalanyane, T. P., Siebe, C., Trivedi, P., Zaady, E., Verma, J. P., Wang, L., ...
120 Fierer, N. (2021). Global homogenization of the structure and function in the soil
121 microbiome of urban greenspaces. *Science Advances*, 7(28).
122 <https://doi.org/10.1126/sciadv.abg5809>
- 123 Elberling, B., Michelsen, A., Schädel, C., Schuur, E. A. G., Christiansen, H. H., Berg, L.,
124 Tamstorf, M. P., & Sigsgaard, C. (2013). Long-term CO₂ production following permafrost
125 thaw. *Nature Climate Change*, 3(10), 890–894. <https://doi.org/10.1038/nclimate1955>

- 126 Eldridge, D. J., Delgado-Baquerizo, M., Travers, S. K., Val, J., Oliver, I., Hamonts, K., & Singh,
127 B. K. (2017). Competition drives the response of soil microbial diversity to increased
128 grazing by vertebrate herbivores. *Ecology*, 98(7), 1922–1931.
129 <https://doi.org/10.1002/ecy.1879>
- 130 Emilia Hannula, S., & Morriën, E. (2022). Will fungi solve the carbon dilemma? *Geoderma*,
131 413, 115767. <https://doi.org/10.1016/j.geoderma.2022.115767>
- 132 FAO. (2015). Healthy soils are the basis for healthy food production. Em *Fao*. FAO.
133 <http://www.fao.org/3/a-i4405e.pdf>
- 134 Freschet, G. T., Pagès, L., Iversen, C. M., Comas, L. H., Rewald, B., Roumet, C., Klimešová,
135 J., Zadworny, M., Poorter, H., Postma, J. A., Adams, T. S., Bagniewska-Zadworna, A.,
136 Bengough, A. G., Blancaflor, E. B., Brunner, I., Cornelissen, J. H. C., Garnier, E., Gessler,
137 A., Hobbie, S. E., ... McCormack, M. L. (2021). A starting guide to root ecology:
138 strengthening ecological concepts and standardising root classification, sampling,
139 processing and trait measurements. *New Phytologist*, 232(3), 973–1122.
140 <https://doi.org/10.1111/nph.17572>
- 141 Freschet, G. T., Roumet, C., Comas, L. H., Weemstra, M., Bengough, A. G., Rewald, B.,
142 Bardgett, R. D., De Deyn, G. B., Johnson, D., Klimešová, J., Lukac, M., McCormack, M.
143 L., Meier, I. C., Pagès, L., Poorter, H., Prieto, I., Wurzbürger, N., Zadworny, M.,
144 Bagniewska-Zadworna, A., ... Stokes, A. (2021). Root traits as drivers of plant and
145 ecosystem functioning: current understanding, pitfalls and future research needs. *New*
146 *Phytologist*, 232(3), 1123–1158. <https://doi.org/10.1111/nph.17072>
- 147 Goede, S. P. C., Hannula, S. E., Jansen, B., & Morriën, E. (2025). Fungal-mediated soil
148 aggregation as a mechanism for carbon stabilization. *The ISME Journal*.
149 <https://doi.org/10.1093/ismejo/wraf074>

- 150 Hengl, T., De Jesus, J. M., Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A.,
151 Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A.,
152 Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I.,
153 Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based
154 on machine learning. *PLoS ONE*, *12*(2), e0169748.
155 <https://doi.org/10.1371/journal.pone.0169748>
- 156 Heuvelink, G. B. M., Angelini, M. E., Poggio, L., Bai, Z., Batjes, N. H., van den Bosch, R.,
157 Bossio, D., Estella, S., Lehmann, J., Olmedo, G. F., & Sanderman, J. (2021). Machine
158 learning in space and time for modelling soil organic carbon change. *European Journal of*
159 *Soil Science*, *72*(4), 1607–1623. <https://doi.org/10.1111/ejss.12998>
- 160 Kopittke, P. M., Berhe, A. A., Carrillo, Y., Cavagnaro, T. R., Chen, D., Chen, Q. L., Román
161 Dobarco, M., Dijkstra, F. A., Field, D. J., Grundy, M. J., He, J. Z., Hoyle, F. C., Kögel-
162 Knabner, I., Lam, S. K., Marschner, P., Martinez, C., McBratney, A. B., McDonald-
163 Madden, E., Menzies, N. W., ... Minasny, B. (2022). Ensuring planetary survival: the
164 centrality of organic carbon in balancing the multifunctional nature of soils. *Critical*
165 *Reviews in Environmental Science and Technology*, *52*(23), 4308–4324.
166 <https://doi.org/10.1080/10643389.2021.2024484>
- 167 Lal, R. (2004). Soil carbon sequestration impacts on global climate change and food security.
168 *Science*, *304*(5677), 1623–1627. <https://doi.org/10.1126/science.1097396>
- 169 Lal, R. (2014). Soil conservation and ecosystem services. *International Soil and Water*
170 *Conservation Research*, *2*(3), 36–47. [https://doi.org/10.1016/S2095-6339\(15\)30021-6](https://doi.org/10.1016/S2095-6339(15)30021-6)
- 171 Lal, R., Follett, R. F., Stewart, B. A., & Kimble, J. M. (2007). Soil carbon sequestration to
172 mitigate climate change and advance food security. *Soil Science*, *172*(12), 943–956.
173 <https://doi.org/10.1097/ss.0b013e31815cc498>

- 174 Minasny, B., Bandai, T., Ghezzehei, T. A., Huang, Y. C., Ma, Y., McBratney, A. B., Ng, W.,
175 Norouzi, S., Padarian, J., Rudiyanto, Sharififar, A., Styc, Q., & Widyastuti, M. (2024). Soil
176 Science-Informed Machine Learning. *Geoderma*, 452, 117094.
177 <https://doi.org/10.1016/j.geoderma.2024.117094>
- 178 Minasny, B., & McBratney, A. B. (2025). Machine Learning and Artificial Intelligence
179 Applications in Soil Science. *European Journal of Soil Science*, 76(2), e70093.
180 <https://doi.org/10.1111/ejss.70093>
- 181 Padarian, J., Minasny, B., & McBratney, A. B. (2020). Machine learning and soil sciences: A
182 review aided by machine learning tools. *Soil*, 6(1), 35–52. [https://doi.org/10.5194/soil-6-](https://doi.org/10.5194/soil-6-35-2020)
183 [35-2020](https://doi.org/10.5194/soil-6-35-2020)
- 184 Poggio, L., De Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., &
185 Rossiter, D. (2021). SoilGrids 2.0: Producing soil information for the globe with quantified
186 spatial uncertainty. *Soil*, 7(1), 217–240. <https://doi.org/10.5194/soil-7-217-2021>
- 187 Pozza, L. E., & Field, D. J. (2020). The science of Soil Security and Food Security. *Soil Security*,
188 1, 100002. <https://doi.org/10.1016/j.soisec.2020.100002>
- 189 Qu, X., Li, X., Bardgett, R. D., Kuzyakov, Y., Revillini, D., Sonne, C., Xia, C., Ruan, H., Liu,
190 Y., Cao, F., Reich, P. B., & Delgado-Baquerizo, M. (2024). Deforestation impacts soil
191 biodiversity and ecosystem services worldwide. *Proceedings of the National Academy of*
192 *Sciences of the United States of America*, 121(13), e2318475121.
193 <https://doi.org/10.1073/pnas.2318475121>
- 194 Ren, S., Wang, T., Guenet, B., Liu, D., Cao, Y., Ding, J., Smith, P., & Piao, S. (2024). Projected
195 soil carbon loss with warming in constrained Earth system models. *Nature*
196 *Communications*, 15(1), 1–10. <https://doi.org/10.1038/s41467-023-44433-2>

- 197 Scharlemann, J. P. W., Tanner, E. V. J., Hiederer, R., & Kapos, V. (2014). Global soil carbon:
198 Understanding and managing the largest terrestrial carbon pool. *Carbon Management*,
199 5(1), 81–91. <https://doi.org/10.4155/cmt.13.77>
- 200 Tedersoo, L., Mikryukov, V., Anslan, S., Bahram, M., Khalid, A. N., Corrales, A., Agan, A.,
201 Vasco-Palacios, A. M., Saitta, A., Antonelli, A., Rinaldi, A. C., Verbeken, A., Sulistyono, B.
202 P., Tamgnoue, B., Furneaux, B., Ritter, C. D., Nyamukondiwa, C., Sharp, C., Marín, C.,
203 ... Abarenkov, K. (2021). The Global Soil Mycobiome consortium dataset for boosting
204 fungal diversity research. *Fungal Diversity*, 111(1), 573–588.
205 <https://doi.org/10.1007/s13225-021-00493-7>
- 206 Thorsøe, M. H., Keesstra, S., De Boever, M., Buchová, K., Bøe, F., Castanheira, N. L., Chenu,
207 C., Cornu, S., Don, A., Fohrafellner, J., Farina, R., Fornara, D., da Conceição Gonçalves,
208 M., Graversgaard, M., Heller, O., Inselsbacher, E., Jacobs, A., Mavsar, S., Meurer, K. H.
209 E., ... Munkholm, L. J. (2023). Sustainable soil management: Soil knowledge use and
210 gaps in Europe. *European Journal of Soil Science*, 74(6), e13439.
211 <https://doi.org/10.1111/ejss.13439>
- 212 Van Der Heijden, M. G. A., Bardgett, R. D., & Van Straalen, N. M. (2008). The unseen majority:
213 Soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems.
214 *Ecology Letters*, 11(3), 296–310. <https://doi.org/10.1111/j.1461-0248.2007.01139.x>
- 215 von Lützow, M., & Kögel-Knabner, I. (2009). Temperature sensitivity of soil organic matter
216 decomposition-what do we know? *Biology and Fertility of Soils*, 46(1), 1–15.
217 <https://doi.org/10.1007/s00374-009-0413-8>
- 218 Wadoux, A. M. J. C. (2025). Artificial intelligence in soil science. *European Journal of Soil*
219 *Science*, 76(2), e70080. <https://doi.org/10.1111/ejss.70080>

- 220 Wadoux, A. M. J. C., Saby, N. P. A., & Martin, M. P. (2023). Shapley values reveal the drivers
221 of soil organic carbon stock prediction. *Soil*, 9(1), 21–38. [https://doi.org/10.5194/soil-9-](https://doi.org/10.5194/soil-9-21-2023)
222 21-2023
- 223 Wang, G., Xiao, L., Lin, Z., Zhang, Q., Guo, X., Cowie, A., Zhang, S., Wang, M., Chen, S.,
224 Zhang, G., Shi, Z., Sun, W., & Luo, Z. (2023). Most root-derived carbon inputs do not
225 contribute to long-term global soil carbon storage. *Science China Earth Sciences*, 66(5),
226 1072–1086. <https://doi.org/10.1007/s11430-022-1031-5>
- 227 Wang, X., Liang, C., Dini-Andreote, F., Zhou, S., & Jiang, Y. (2024). Impacts of trophic
228 interactions on carbon accrual in soils. *Trends in Microbiology*, 33(3), 277–284.
229 <https://doi.org/10.1016/j.tim.2024.10.009>
- 230 Whenua, M. (2025). *How do soils form?*
231 <https://soils.landcareresearch.co.nz/topics/understanding-soils/how-do-soils-form>
- 232 Yaffar, D., Lugli, L. F., Wong, M. Y., Norby, R. J., Addo-Danso, S. D., Arnaud, M., Cordeiro,
233 A. L., Dietterich, L. H., Diaz-Toribio, M. H., Lee, M. Y., Ghimire, O. P., Smith-Martin, C.
234 M., Toro, L., Andersen, K., McCulloch, L. A., Meier, I. C., Powers, J. S., Sanchez-Julia,
235 M., Soper, F. M., & Cusack, D. F. (2024). Tropical root responses to global changes: A
236 synthesis. *Global Change Biology*, 30(7), e17420. <https://doi.org/10.1111/gcb.17420>
- 237 Yamauchi, D. H., Garces, H. G., Teixeira, M. D. M., Rodrigues, G. F. B., Ullmann, L. S., Garces,
238 A. G., Hebelbarbosa, F., & Bagagli, E. (2021). Soil mycobiome is shaped by vegetation
239 and microhabitats: A regional-scale study in Southeastern Brazil. *Journal of Fungi*, 7(8),
240 587. <https://doi.org/10.3390/jof7080587>
- 241 Zomer, R. J., Bossio, D. A., Sommer, R., & Verchot, L. V. (2017). Global Sequestration
242 Potential of Increased Organic Carbon in Cropland Soils. *Scientific Reports*, 7(1), 1–8.
243 <https://doi.org/10.1038/s41598-017-15794-8>

244 **Artigo 1: Fungos do solo em escala global: implicações ecológicas e funcionais para a**
245 **saúde do solo**
246

247 1. Introdução

248 O solo é vivo e influencia a saúde das plantas, animais e pessoas (Almario et al., 2014;
249 Larson, 1991; Warkentin, 1995). Além disso, a intrincada comunidade biótica existente no solo
250 por muito tempo foi negligenciada e/ou simplificada. Apenas recentemente descobriu-se o quão
251 enorme é a diversidade e a rede de interações dos microrganismos no solo (Bahram et al., 2018).
252 Estimativas apontam que apenas 0,01% do mundo microbiano já foi descrito (Locey & Lennon,
253 2016), e pouco se conhece sobre sua ecologia (Baldrian, 2019). Este crescente conhecimento
254 sobre a diversidade microbiana destaca a importância de uma abordagem mais holística no
255 estudo dos solos e a carência de informações sobre a temática.

256 Estudos recentes têm buscado compreender o solo como um organismo vivo. Nesse
257 contexto, o termo saúde do solo (*soil health*), definido como “a capacidade contínua do solo de
258 funcionar como um ecossistema vivo essencial que sustenta plantas, animais e seres humanos”
259 (J. Lehmann et al., 2020), passou a ser amplamente utilizado na última década (Bünemann et
260 al., 2018). Em uma busca filtrada por datas na base Scopus (acesso em julho de 2024),
261 observou-se que, até 2015, havia pouco mais de mil publicações — entre artigos e revisões —
262 contendo esse termo. Por outro lado, entre 2015 e o presente, já são mais de 9 mil trabalhos
263 científicos publicados.

264 O crescente interesse da comunidade científica, especialmente entre os cientistas do
265 solo, decorre da compreensão de que o solo não deve ser visto apenas como um reservatório
266 químico de nutrientes para as plantas, mas sim como um ecossistema dinâmico, cuja
267 funcionalidade depende da atividade de microrganismos na regulação dos ciclos
268 biogeoquímicos e, conseqüentemente, na provisão de múltiplos serviços ecossistêmicos.

269 Dentre os organismos do solo, os fungos destacam-se como um dos grupos mais
270 diversos do planeta, sendo superado apenas pelos insetos (Blackwell, 2011). Eles estão
271 distribuídos globalmente, em todos os tipos de ecossistemas, desde os terrestres até os

272 aquáticos, e mesmo em condições extremas, como na Antártica (Gessner et al., 2007; Rosa et
273 al., 2009, 2025). Os fungos desempenham papéis fundamentais no funcionamento dos
274 ecossistemas, modulando a ciclagem de nutrientes, a decomposição da matéria orgânica, o
275 armazenamento e a persistência do carbono, e a nutrição das plantas por meio da formação de
276 simbioses mutualísticas (Delgado-Baquerizo et al., 2016; J. Li et al., 2019).

277 Nesse contexto, estudos em ambientes controlados têm avançado na identificação de
278 atributos microbianos do solo, como a relação fungo:bactéria, que podem regular a resistência
279 de funções ecossistêmicas específicas, como a mineralização de nitrogênio ou a respiração do
280 solo, frente a fatores de mudança global, como a intensificação do uso da terra e a seca (De
281 Vries & Shade, 2013; Downing & Leibold, 2010).

282 No entanto, apesar do reconhecimento da microbiota como pilar central da
283 biodiversidade e do funcionamento dos ecossistemas, ainda há uma notável escassez de
284 informações sobre a distribuição espacial de sua abundância, riqueza e composição (Fitter,
285 2005). Em contraste, a variação em larga escala da diversidade de outros grupos taxonômicos
286 já é amplamente documentada (Gaston, 2000; Rosenzweig, 1995).

287 Sabe-se que microrganismos podem seguir uma relação táxon-área, com aumento da
288 diversidade conforme a área amostrada (Horner-Devine et al., 2004), mas ainda faltam
289 evidências empíricas robustas que revelem os padrões globais da diversidade microbiana e seus
290 fatores moduladores, especialmente no caso dos fungos do solo. Alguns atributos microbianos
291 do solo, como a abundância e a riqueza da comunidade, têm se mostrado fundamentais para
292 sustentar a resistência da multifuncionalidade do solo frente a distúrbios climáticos (Emilia
293 Hannula & Morriën, 2022).

294 Nesse contexto, modelar espacialmente esses atributos é fundamental para compreender
295 como a funcionalidade ecológica do solo pode variar entre ecorregiões com condições
296 ambientais e históricas contrastantes. Ecorregiões frias e úmidas, por exemplo, tendem a abrigar

297 comunidades microbianas distintas daquelas encontradas em zonas áridas ou tropicais, não
298 apenas em termos de composição, mas também de sensibilidade às perturbações e capacidade
299 de recuperação (Cavicchioli et al., 2019; Knight et al., 2024). A ausência de mapas de atributos
300 microbianos compromete a integração entre biodiversidade subterrânea e serviços
301 ecossistêmicos em avaliações globais, dificultando a formulação de estratégias de mitigação e
302 adaptação às mudanças climáticas baseadas no solo.

303 Diante dessa lacuna, torna-se urgente desenvolver abordagens que consigam capturar a
304 complexidade funcional e espacial da diversidade microbiana em diferentes ecorregiões. A alta
305 variabilidade dos dados biológicos, aliada à natureza não linear das relações entre
306 microrganismos e fatores ambientais, impõe sérias limitações aos métodos estatísticos
307 convencionais, que muitas vezes pressupõem linearidade, independência entre variáveis e
308 distribuições normalizadas. Essas restrições dificultam a detecção de padrões ecológicos
309 complexos e reduzem a capacidade preditiva dos modelos.

310 Nesse cenário, o uso de inteligência artificial, especialmente por meio de algoritmos de
311 aprendizado de máquina, tem se mostrado promissor para modelar fenômenos ecológicos
312 complexos e espacialmente heterogêneos (Moquedace et al., 2024; Siqueira et al., 2024). A
313 aplicação desses modelos a grandes bancos de dados legados surge como alternativa estratégica
314 para prever atributos globais da microbiota do solo, como riqueza e abundância de fungos,
315 além de fornecer novos *insights* sobre os fatores que controlam sua distribuição espacial
316 (Reichstein et al., 2019; Tramontana et al., 2016; Warner et al., 2019).

317 As técnicas de aprendizado de máquina permitem identificar relações lineares e não
318 lineares em um espaço multivariado com uma grande quantidade de preditores ambientais
319 (relevo, clima, solo, entre outros). Isso possibilita realizar previsões com incerteza quantificável
320 em áreas onde informações, como a abundância e riqueza de fungos, não existam. Dessa forma,
321 compreender como a abundância e riqueza fúngica responde a gradientes ambientais em

322 diferentes ecorregiões é essencial para identificar padrões ecológicos consistentes e inferir o
323 potencial funcional dos solos. A riqueza de fungos, frequentemente utilizada como indicador
324 *proxy* da saúde e da multifuncionalidade do solo, representa uma métrica comparativa relevante
325 entre ecossistemas sujeitos a distintas condições ambientais e históricos de uso (Bünemann et
326 al., 2018).

327 Nesse contexto, objetivou-se modelar e mapear a riqueza e abundância de fungos nos
328 primeiros 5 cm de solo em escala global. Especificamente, os objetivos foram: i. mapear a
329 variabilidade da distribuição da riqueza e abundância de fungos globalmente; ii. avaliar quais
330 são os *proxies* ambientais que controlam a riqueza e abundância de fungos globalmente; iii.
331 estratificar a riqueza e abundância de fungos no solo nas diferentes ecorregiões globais.

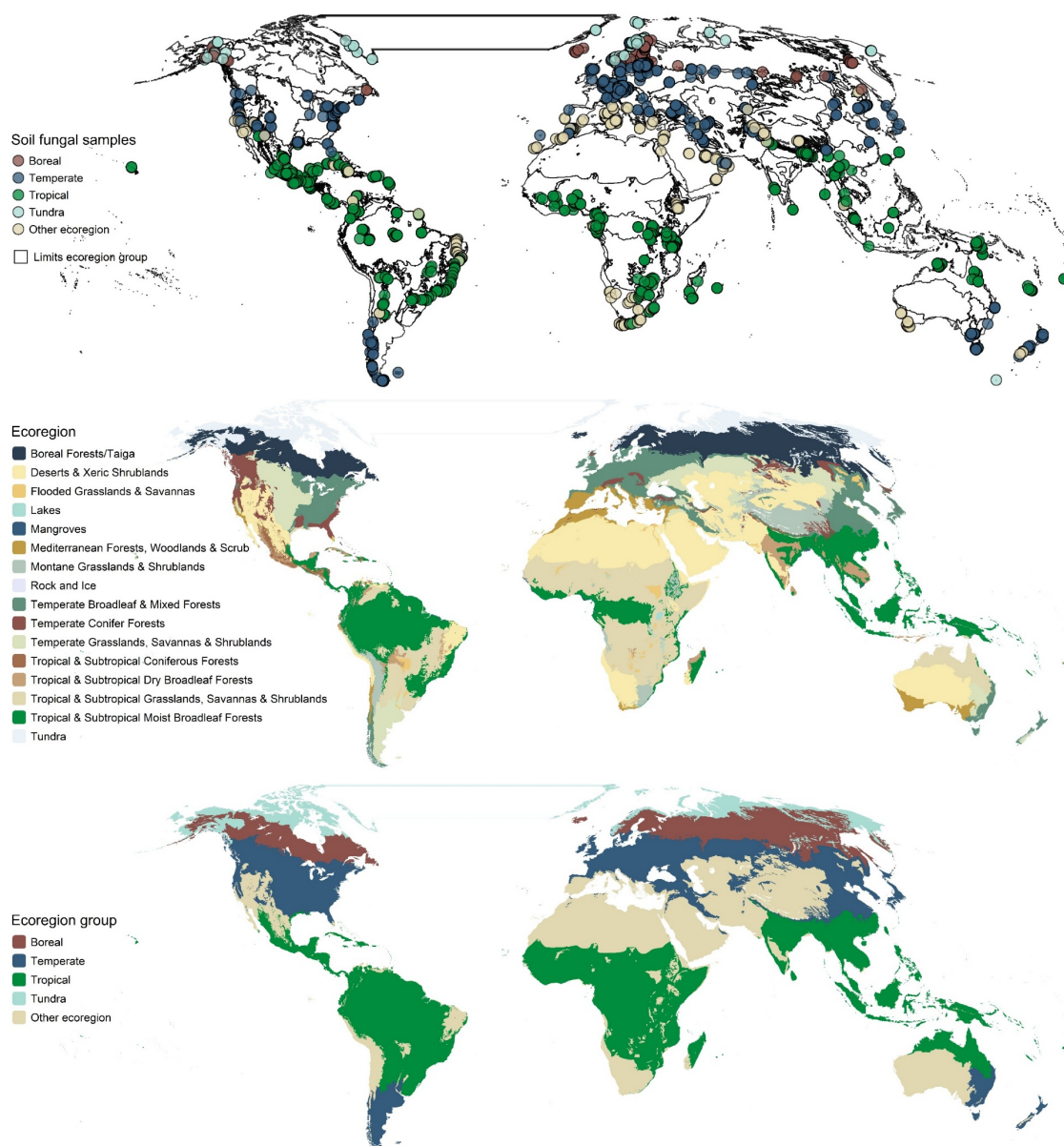
332

333 **2. Material e métodos**

334 **2.1. Base de dados**

335 A base de dados de fungos do solo utilizada foi a “*Global Soil Mycobiome*” (GSM),
336 proveniente do trabalho de Tedersoo et al. (2021). Em uma iniciativa de abrangência global, os
337 autores amostraram mais de 3.250 parcelas distribuídas por 110 países, cobrindo todos os
338 continentes e principais ecorregiões do planeta (Figura 1). As coletas foram realizadas com
339 metodologia padronizada, limitada aos primeiros 5 cm de profundidade do solo, privilegiando
340 ambientes com baixa interferência antrópica, especialmente áreas florestais.

341



342
 343 **Figura 1.** Distribuição das amostras de abundância e riqueza de fungos (Tedersoo et al., 2021)
 344 nos 5 cm de solo nas diferentes ecorregiões globais (Olson et al., 2001a).
 345

346 Como resultado, foram identificadas mais de 700 mil unidades taxonômicas
 347 operacionais (OTUs) de fungos. Para este estudo, foram selecionadas as variáveis de
 348 abundância de fungos do solo (SFA), definida como o número total de sequências obtidas por
 349 parcela, e de riqueza de fungos (SFR), representada pelo número de OTUs distintas por parcela.
 350 As parcelas amostradas apresentavam, em geral, formato retangular (50 × 50 m) ou circular
 351 (56 m de diâmetro), sendo registrado o tamanho efetivo da área amostrada para cada ponto.
 352 Para padronização e viabilização de comparações ecológicas entre diferentes locais, os valores

353 de abundância e riqueza foram divididos pela área da parcela correspondente, resultando na
 354 abundância de fungos expressa em sequências por metro quadrado (sequences m⁻²) e na riqueza
 355 de fungos expressa em OTUs por metro quadrado (OTUs m⁻²).

356

357 2.2. Camadas de preditores ambientais

358 Um atributo ou classe de solo em um ponto específico no espaço resulta de múltiplas
 359 interações ao longo do tempo e do espaço. McBratney et al. (2003) propuseram o modelo
 360 SCORPAN (Sa = f(solo, clima, organismos, relevo, material de origem, tempo, localização
 361 espacial)) como uma estrutura empírica derivada dos modelos mecanicistas clássicos de
 362 formação do solo, como o de Jenny (1945), para representar a formação e variabilidade dos
 363 atributos edáficos. A combinação espacial dos valores desses fatores é potencialmente capaz de
 364 explicar variações em atributos como a abundância e a riqueza fúngica do solo. Para representar
 365 os fatores propostos do modelo SCORPAN nesse trabalho foram escolhidas 50 variáveis
 366 preditoras (Tabela 1).

367

368 **Tabela 1.** Preditores ambientais e suas respectivas referências utilizadas na modelagem da
 369 abundância e riqueza de fungos nos primeiros 5 cm de solo.

Sigla	Preditor	Grupo	Ref.
BIO 1	<i>Mean annual temperature</i>	<i>Climate</i>	(Karger et al., 2017)
BIO 10	<i>Mean temperature of the warmest quarter</i>	<i>Climate</i>	(Karger et al., 2017)
BIO 11	<i>Mean temperature of the coldest quarter</i>	<i>Climate</i>	(Karger et al., 2017)
BIO 12	<i>Annual precipitation</i>	<i>Climate</i>	(Karger et al., 2017)
BIO 13	<i>Precipitation of the wettest month</i>	<i>Climate</i>	(Karger et al., 2017)
BIO 14	<i>Precipitation of the driest month</i>	<i>Climate</i>	(Karger et al., 2017)
BIO 15	<i>Precipitation seasonality</i>	<i>Climate</i>	(Karger et al., 2017)
BIO 16	<i>Precipitation of the wettest quarter</i>	<i>Climate</i>	(Karger et al., 2017)
BIO 17	<i>Precipitation of the driest quarter</i>	<i>Climate</i>	(Karger et al., 2017)
BIO 18	<i>Precipitation of the warmest quarter</i>	<i>Climate</i>	(Karger et al., 2017)
BIO 19	<i>Precipitation of the coldest quarter</i>	<i>Climate</i>	(Karger et al., 2017)
BIO 2	<i>Mean diurnal temperature range</i>	<i>Climate</i>	(Karger et al., 2017)
BIO 3	<i>Isothermality</i>	<i>Climate</i>	(Karger et al., 2017)

Sigla	Preditor	Grupo	Ref.
BIO 4	<i>Temperature seasonality</i>	<i>Climate</i>	(Karger et al., 2017)
BIO 5	<i>Maximum temperature of the warmest month</i>	<i>Climate</i>	(Karger et al., 2017)
BIO 6	<i>Minimum temperature of the coldest month</i>	<i>Climate</i>	(Karger et al., 2017)
BIO 7	<i>Annual temperature range</i>	<i>Climate</i>	(Karger et al., 2017)
BIO 8	<i>Mean temperature of the wettest quarter</i>	<i>Climate</i>	(Karger et al., 2017)
BIO 9	<i>Mean temperature of the driest quarter</i>	<i>Climate</i>	(Karger et al., 2017)
CHT	<i>Canopy height</i>	<i>Organisms</i>	(Tolan et al., 2024)
ET	<i>Evapotranspiration</i>	<i>Organisms</i>	(Steve Running & Qiaozhen Mu, 2015)
EVI	<i>Enhanced vegetation index</i>	<i>Organisms</i>	(Vermote, 2021)
FC	<i>Field capacity</i>	<i>Organisms</i>	(Zhang et al., 2018)
FP	<i>Fraction of photosynthetically active radiation</i>	<i>Organisms</i>	(Running et al., 2021)
GPP	<i>Gross primary production</i>	<i>Organisms</i>	(Running et al., 2021)
LAI	<i>Leaf area index</i>	<i>Organisms</i>	(Myneni et al., 2021)
MSAVI	<i>Modified soil-adjusted vegetation index</i>	<i>Organisms</i>	(Vermote, 2021)
NDMI	<i>Normalized difference moisture index</i>	<i>Organisms</i>	(Vermote, 2021)
NDSI	<i>Normalized difference snow index</i>	<i>Organisms</i>	(Vermote, 2021)
NDVI	<i>Normalized difference vegetation index</i>	<i>Organisms</i>	(Vermote, 2021)
NPP	<i>Net primary productivity</i>	<i>Organisms</i>	(Karger et al., 2017)
PAW	<i>Plant available water</i>	<i>Organisms</i>	(Zhang et al., 2018)
SAVI	<i>Soil-adjusted vegetation index</i>	<i>Organisms</i>	(Vermote, 2021)
SMBC	<i>Microbial biomass carbon</i>	<i>Organisms</i>	(Serna-Chavez et al., 2013)
CEC	<i>Ability to exchange cations</i>	<i>Soil / PM</i>	(Poggio et al., 2021)
Clay	<i>Clay content</i>	<i>Soil / PM</i>	(Poggio et al., 2021)
Fe oxides	<i>Iron oxide</i>	<i>Soil / PM</i>	(Ito & Wagai, 2017)
Gibbsite	<i>Gibbsite</i>	<i>Soil / PM</i>	(Ito & Wagai, 2017)
Illite / Mica	<i>Illite/Mica</i>	<i>Soil / PM</i>	(Ito & Wagai, 2017)
Kaolinite	<i>Kaolinite</i>	<i>Soil / PM</i>	(Ito & Wagai, 2017)
N	<i>Nitrogen content</i>	<i>Soil / PM</i>	(Poggio et al., 2021)
PEAT	<i>Peatland extent</i>	<i>Soil / PM</i>	(Melton et al., 2022)
pH	<i>Hydrogen potential in water</i>	<i>Soil / PM</i>	(Poggio et al., 2021)
Sand	<i>Sand content</i>	<i>Soil / PM</i>	(Poggio et al., 2021)
SM	<i>Soil moisture</i>	<i>Soil / PM</i>	(Muñoz Sabater, 2024)
Smectite	<i>Smectite</i>	<i>Soil / PM</i>	(Ito & Wagai, 2017)
Vermiculite	<i>Vermiculite</i>	<i>Soil / PM</i>	(Ito & Wagai, 2017)
WDT	<i>Water table depth</i>	<i>Soil / PM</i>	(Fan et al., 2013)
DEM	<i>Digital elevation model</i>	<i>Topography</i>	(Hawker et al., 2022)

371 Como representação do clima, utilizaram-se os dados bioclimáticos oriundos do
372 CHELSA v2.1 (Karger et al., 2017), que combinam informações históricas de temperatura e
373 precipitação, proporcionando indicadores mais informativos sobre sazonalidade e eventos
374 extremos do que os dados climáticos brutos isoladamente. Ainda da base CHELSA, foi utilizada
375 a produtividade primária líquida (NPP), representando o componente biótico do modelo
376 SCORPAN, com resolução espacial aproximada de 1 km.

377 A altura do dossel das árvores foi representada a partir do modelo proposto por Tolan et
378 al. (2024), que combinou dados do programa GEDI (Global Ecosystem Dynamics
379 Investigation) com LiDAR aéreo de alta resolução, resultando em uma camada com resolução
380 espacial de aproximadamente 1 m no processo de validação. Esse atributo, também associado
381 ao fator “organismos”, pode contribuir para a compreensão das interações biológicas que
382 influenciam os fungos do solo. Já a evapotranspiração, representando a interação entre solo,
383 clima e vegetação, foi obtida a partir do produto MODIS (Steve Running & Qiaozhen Mu,
384 2015), também incluída como variável preditora.

385 Para representar o relevo, foi utilizado o modelo digital de elevação (DEM) global
386 FABDEM (*Forest And Buildings removed Copernicus DEM*), com resolução espacial de 30 m.
387 Esse modelo é derivado do Copernicus GLO-30 (The European Space Agency, 2022), corrigido
388 por meio de algoritmos de aprendizado de máquina para remover a influência de florestas e
389 edificações, resultando em um modelo de terreno mais próximo do solo exposto (“*bare-earth*”).
390 A correção realizada pelos autores foi baseada em dados LiDAR de 12 países e em variáveis
391 ambientais globais, o que garantiu ampla aplicabilidade geográfica e precisão superior em áreas
392 urbanas e vegetadas (Hawker et al., 2022). A caracterização do solo e do material de origem
393 incluiu a umidade do solo média dos últimos 20 anos, com resolução de ~11 km (Muñoz
394 Sabater, 2024).

395 A concentração de minerais de argila no solo foi obtida a partir do conjunto de dados
396 global desenvolvido por Ito & Wagai (2017), que integrou registros da literatura científica com
397 análises mineralógicas por difração de raios X. Os autores classificaram os minerais em dez
398 grupos principais (gibbsite, caulinita, illita/mica, esmectita, vermiculita, clorita, óxidos de ferro,
399 quartzo, minerais não cristalinos e outros) e espacializaram sua distribuição com base nas
400 ordens da *Soil Taxonomy*. O resultado foi a geração de mapas de abundância relativa e absoluta
401 desses minerais por profundidade e ordem de solo, com resolução espacial aproximada de 9 km.

402 A extensão global das áreas turfosas foi representada pelo produto Peat-ML (Melton et
403 al., 2022), que fornece um mapa contínuo de cobertura fracionária de turfeiras em escala global,
404 também com resolução de aproximadamente 9 km. Esse produto foi gerado com o uso de
405 algoritmos de aprendizado de máquina (LightGBM), treinados com dados regionais validados
406 e preditores ambientais relacionados ao clima, relevo, atributos do solo e índices espectrais de
407 vegetação.

408 Adicionalmente, um conjunto de variáveis relacionadas à vegetação e à superfície
409 terrestre foi obtido por meio da plataforma *Google Earth Engine*, com base em dados MODIS
410 e outros sensores, incluindo: evapotranspiração, índice de vegetação aprimorado, fração de
411 radiação fotossinteticamente ativa, produtividade primária bruta, índice de área foliar, índice de
412 vegetação ajustado ao solo, índice de umidade da vegetação, índice de neve e o índice de
413 vegetação por diferença normalizada. Para assegurar a qualidade desses dados, aplicaram-se
414 filtros para remoção de nuvens, sombras, *pixels* com problemas e baixa qualidade, obtendo-se
415 a média dos últimos 20 anos para cada variável.

416 Todas as variáveis ambientais utilizadas, com suas respectivas fontes, estão listadas na
417 Tabela 1. Em função das limitações impostas pela distribuição das amostras fúngicas e pela
418 cobertura incompleta de preditores ambientais em regiões polares, a Antártica foi excluída das
419 análises.

420 Todas as camadas de preditores ambientais foram harmonizadas para uma resolução
421 espacial de 1 km e reprojctadas para o sistema de projeção Goode Homolosine - Land
422 (ESRI:54052), aplicado ao *Datum* WGS84. A escolha dessa projeção baseia-se na sua
423 efetividade em representar as zonas terrestres de forma mais precisa em termos de área (Sousa
424 et al., 2019).

425

426 **2.3. Modelagem geoespacial**

427 **2.3.1. Processos de modelagem e mapeamento**

428 A modelagem e o mapeamento da abundância e riqueza de fungos no solo seguiram as
429 etapas sequenciais: (i) seleção de preditores ambientais; (ii) ajuste do modelo; (iii) predição dos
430 mapas; (iv) avaliação da importância dos preditores. Todo o processo de modelagem espacial e
431 a preparação dos mapas foram conduzidos no ambiente R (R Core Team, 2025).

432

433 **2.3.2. Seleção de variáveis**

434 Para construir um modelo mais parcimonioso, com menor esforço computacional e
435 maior poder explicativo, foram adotados três processos sequenciais de seleção de variáveis.
436 Inicialmente, os preditores ambientais com valores de variância próximos de zero foram
437 removidos utilizando a função “*nearZeroVar*” do pacote *caret* (Kuhn, 2022) no ambiente R (R
438 Core Team, 2025). Preditores ambientais que carregam pouca variabilidade foram considerados
439 de pouca contribuição para o ajuste do modelo.

440 Os preditores remanescentes foram submetidos a uma remoção por correlação. Este
441 processo consistiu em remover preditores altamente correlacionadas. Para isso, construiu-se
442 uma matriz de correlação utilizando o coeficiente de correlação de Spearman e pares de
443 variáveis com coeficientes $> |0,95|$ tiveram uma das variáveis removidas. A variável do par

444 removida foi a que apresentou maior correlação média com as demais, utilizando a função
445 “*findCorrelation*” do pacote *caret*.

446 As variáveis remanescentes de ambos os processos foram submetidas ao algoritmo
447 iterativo *recursive feature elimination* (RFE). O algoritmo avalia a importância de
448 subconjuntos de preditores em um processo de eliminação reversa. Foram testados
449 subconjuntos discretos de preditores, variando de 2 até o total disponível resultante da
450 eliminação por correlação. O menor subconjunto com as melhores métricas de performance foi
451 o adotado para a modelagem.

452

453 **2.3.3. Ajuste do modelo e predição dos mapas**

454 O algoritmo utilizado para a modelagem foi o quantile random forest (qrf), proposto por
455 Meinshausen (2017). Trata-se de um método baseado no princípio do *bootstrap aggregating*,
456 no qual diversas árvores de decisão são construídas a partir de amostras aleatórias dos dados
457 originais, com reposição. Em cada árvore, apenas um subconjunto aleatório de preditores é
458 considerado a cada divisão dos nós, o que promove diversidade estrutural entre as árvores e
459 reduz a variância do modelo agregado.

460 No qrf, diferentemente do random forest que estima apenas valores médios ou medianas
461 condicionais, todas as observações da variável resposta associadas a cada folha terminal de cada
462 árvore são armazenadas. Durante a predição, para cada nova entrada, o modelo percorre as
463 árvores, identifica as folhas correspondentes e extrai os valores armazenados. A agregação
464 desses valores ao longo de todas as árvores permite a reconstrução empírica da distribuição
465 condicional da variável de interesse. Com isso, o qrf viabiliza a extração direta de quantis
466 condicionais, como os percentis 5%, 50% e 95%, permitindo a geração de intervalos de
467 predição não paramétricos.

468 Essa abordagem é especialmente adequada para dados com distribuição assimétrica ou
469 heterocedástica, como é comum em atributos do solo, pois não requer pressupostos de
470 normalidade. Contudo, sua capacidade de estimar quantis extremos com precisão pode ser
471 limitada em regiões com baixa densidade amostral, o que exige atenção ao ajuste de
472 hiperparâmetros, como a profundidade das árvores e o número mínimo de observações por
473 folha.

474 O qrf tem se mostrado eficaz na modelagem e previsão de propriedades do solo, com
475 aplicações bem-sucedidas em escalas regionais (Moquedace et al., 2024; Siqueira et al., 2023),
476 nacionais (Gomes et al., 2019), continentais (Hengl et al., 2015; Siqueira et al., 2023, 2024) e
477 globais (Hengl et al., 2017; Poggio et al., 2021), evidenciando sua robustez, flexibilidade e
478 capacidade de quantificar incertezas de forma empírica. O processo de ajuste do modelo neste
479 trabalho seguiu dois métodos diferentes: um para a previsão dos mapas e outro para
480 quantificação da acurácia.

481 O modelo para a previsão dos mapas foi ajustado com o quantitativo total das amostras
482 disponíveis. O processo incluiu validação cruzada repetida (10 *folds*) e otimização de
483 hiperparâmetros com o parâmetro “*tuneLength*” da função “*train*” do *caret*, com valor 10. O
484 argumento *tuneLength* controla a quantidade de valores de hiperparâmetros que o algoritmo
485 avaliará durante a busca por hiperparâmetros ótimos. No caso do qrf, o hiperparâmetro
486 otimizado foi o “*mtry*”, que define a quantidade de variáveis selecionadas aleatoriamente para
487 a divisão de um nó. Após o ajuste do modelo, foram realizadas previsões dos mapas globais
488 médios, percentil 25% e 75%, e o coeficiente de variação da abundância e riqueza de fungos do
489 solo.

490 Para quantificar a acurácia, adotamos um processo de *holdout* repetido com amostragem
491 aleatória estratificada na separação dos dados para treinamento e teste de performance. Os
492 dados foram divididos em 80% para treinamento com validação cruzada (10 *folds*) repetida 10

493 vezes e 20% para teste, que faz a avaliação da performance do modelo em dados inéditos para
 494 ele. As métricas de performance adotadas foram o erro médio absoluto – MAE (1), o coeficiente
 495 de correlação de concordância – ρ_c (2) e a raiz quadrada do erro médio – RMSE (3).
 496

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

$$\rho_c = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2 + (\mu_x - \mu_y)^2} \quad (2)$$

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

497 Em que: n representa o número de amostras; ρ_c é o coeficiente de correlação entre as duas variáveis, μ_x e μ_y são as
 498 médias para as duas variáveis, σ_x^2 e σ_y^2 e são as variâncias correspondentes; P_i e O_i representam os valores previstos
 499 e observados no local i, respectivamente.
 500

501 Desde a seleção de variáveis até o ajuste do modelo, o processo foi repetido 100 vezes
 502 sob diferentes arranjos de variáveis e conjuntos de treino e teste. A performance final foi
 503 calculada como a média das 100 performances dos conjuntos de testes. O modelo final foi
 504 ajustado com todos os dados, sem a separação do conjunto de teste e é esperado que ele tenha
 505 uma performance superior àquela calculada pelo processo de separação de treino e teste. Foi
 506 adicionada uma restrição espacial em ambos os modelos (modelo para predição e modelo para
 507 estimativa de performance), tanto no *holdout* quanto na validação cruzada, de tal forma que
 508 dados temporais com as mesmas coordenadas não se misturaram entre os *datasets* de treino e
 509 teste, e nos *folds* da validação cruzada evitando efeitos de *overfitting*. Como forma adicional de
 510 verificação da performance do modelo, foi calculado um modelo nulo que consistiu em
 511 preencher os valores preditos com a média e, posteriormente, calcular as métricas de erro (MAE
 512 e RMSE). O modelo nulo é um modelo simples, como no caso presente a média aritmética. Ele

513 é o modelo cuja performance tem que ser ultrapassada pelos modelos muito mais sofisticados
514 de *machine learning*.

515 Além da avaliação tradicional de desempenho, também foi calculado o Índice de
516 Dissimilaridade (DI) para os modelos ajustados. O DI quantifica a distância, no espaço das
517 variáveis preditoras, entre cada ponto de predição e os dados utilizados no treinamento,
518 ponderando as variáveis conforme sua importância no modelo. A construção do DI seguiu a
519 metodologia proposta por Meyer and Pebesma (2021), implementadas com o pacote CAST
520 (Meyer et al., 2025) em que a dissimilaridade foi baseada na menor distância ponderada entre
521 novos pontos e o conjunto de treino, normalizada pela média das distâncias observadas entre os
522 dados de treinamento.

523 O modelo de importância de variáveis utilizado no cálculo do DI foi derivado do ajuste
524 do modelo global. As mesmas restrições de independência espacial aplicadas na divisão de
525 treino e teste foram mantidas no cálculo do DI, garantindo consistência metodológica. Valores
526 mais elevados de DI indicam maior extrapolação ambiental e, portanto, maior potencial de
527 incerteza nas predições.

528

529 **2.3.4. Importância dos preditores**

530 A importância relativa dos preditores para a construção dos modelos de abundância e
531 riqueza de fungos foi calculada utilizando a função “*varImp*” do pacote *caret*. Os mapas médios
532 da abundância e riqueza de fungos no solo foram estratificados nos diferentes ecorregiões
533 globais. As ecorregiões utilizadas foram uma simplificação das ecorregiões propostas por Olson
534 et al. (2001).

535

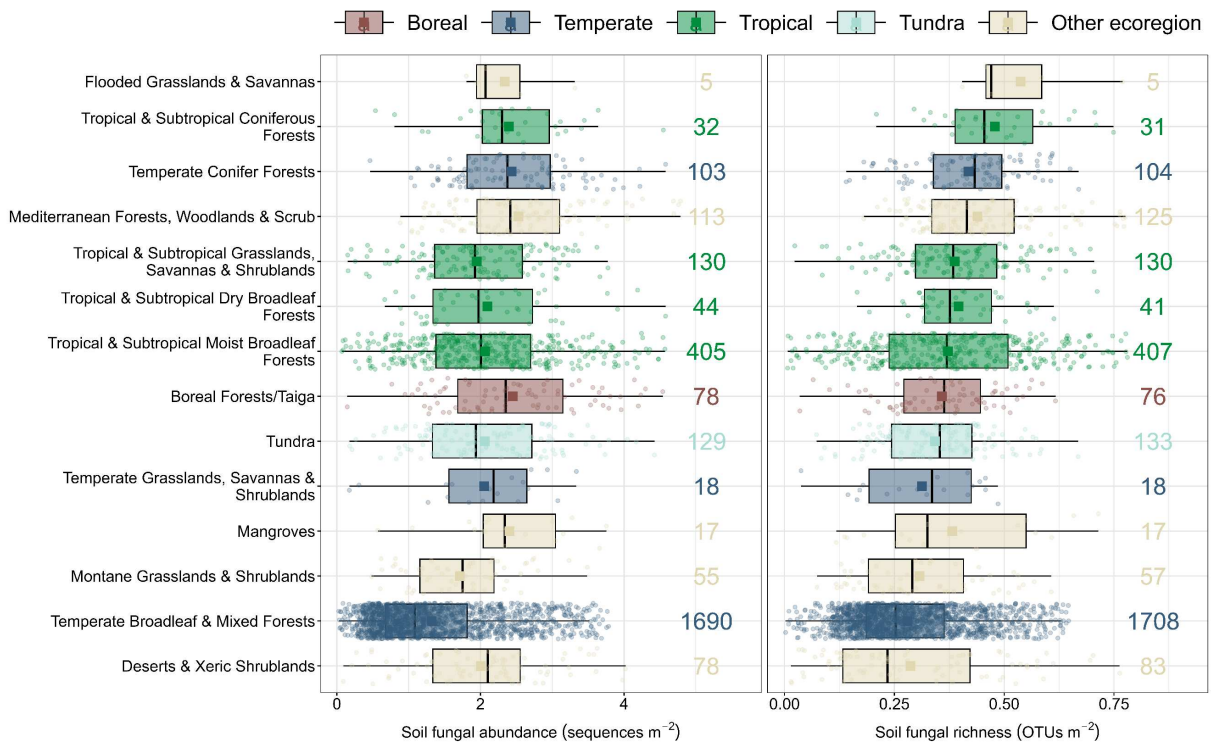
536 3. Resultados

537 3.1. Banco de dados e modelagem

538 Nos valores mensurados, as maiores medianas de abundância de fungos foram
539 observadas nas Florestas, Bosques e Arbustais mediterrâneos, enquanto a maior riqueza foi
540 observada nas Savanas e Campos Inundados (Figura 2). Em relação às medianas das
541 ecorregiões agrupadas, a ecorregião boreal apresenta a maior abundância, enquanto a
542 ecorregião tropical se destaca pela riqueza de fungos no solo.

543 O banco de dados de Tedersoo et al. (2021), embora seja uma iniciativa ambiciosa e
544 abrangente, subamostra algumas áreas e superamostra outras. Nota-se que a ecorregião de
545 Florestas Temperadas alcançou mais de 1700 amostras para ambas as variáveis, enquanto
546 regiões de Savanas e Campos Inundados tiveram menos de 10 amostras. Quanto ao
547 agrupamento das ecorregiões, a ecorregião boreal apresentou o menor número de amostras
548 (Abundância = 78, Riqueza = 76), e a ecorregião temperada, a maior (Abundância = 1690,
549 Riqueza = 1708).

550

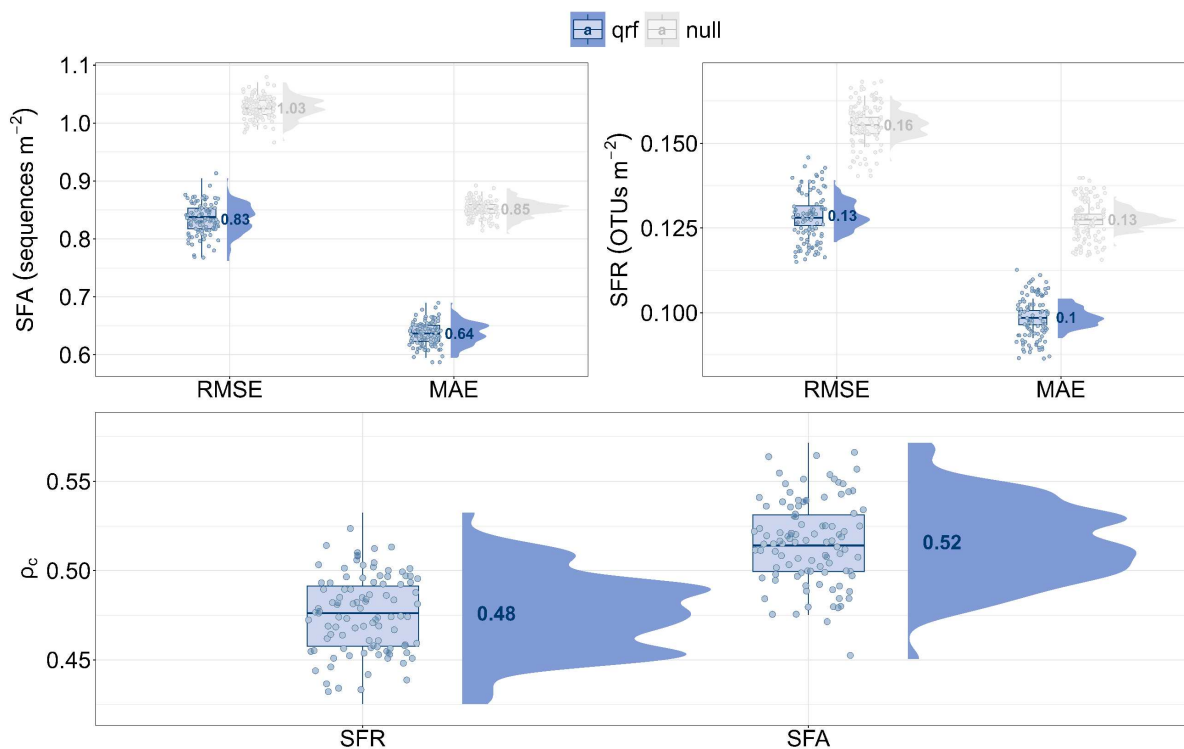


551
552
553
554
555

Figura 2. Distribuição das amostras de abundância e riqueza de fungos no solo em diferentes ecorregiões de acordo com o banco de dados de Tedersoo et al. (2021). Valores a direita indicam a quantidade total de amostras em cada ecorregião.

556
557
558
559
560
561
562

O qrf apresentou desempenho similar no ajuste para riqueza e abundância de fungos no solo globalmente, com desempenhos ligeiramente maiores para riqueza em comparação com abundância (Riqueza ρ_c médio = 0,52; Abundância ρ_c médio = 0,49) (Figura 3). A amplitude de variação entre os valores de desempenho das 100 execuções foi relativamente pequena, e a diferença de desempenho do modelo ajustado em relação ao modelo nulo foi expressivamente maior.



563
 564 **Figura 3.** Coeficiente de correlação de concordância (ρ_c), erro médio absoluto (MAE) e raiz
 565 quadrada do erro-médio (RMSE) das 100 execuções do quantile random forest (qrf) e modelo
 566 nulo (média) para a riqueza (SFR) e abundância (SFA) de fungos nos primeiros 5 cm do solo.
 567 Quadrados coloridos representam os valores médios.
 568

569 3.2. Distribuição espacial da abundância e riqueza de fungos no solo

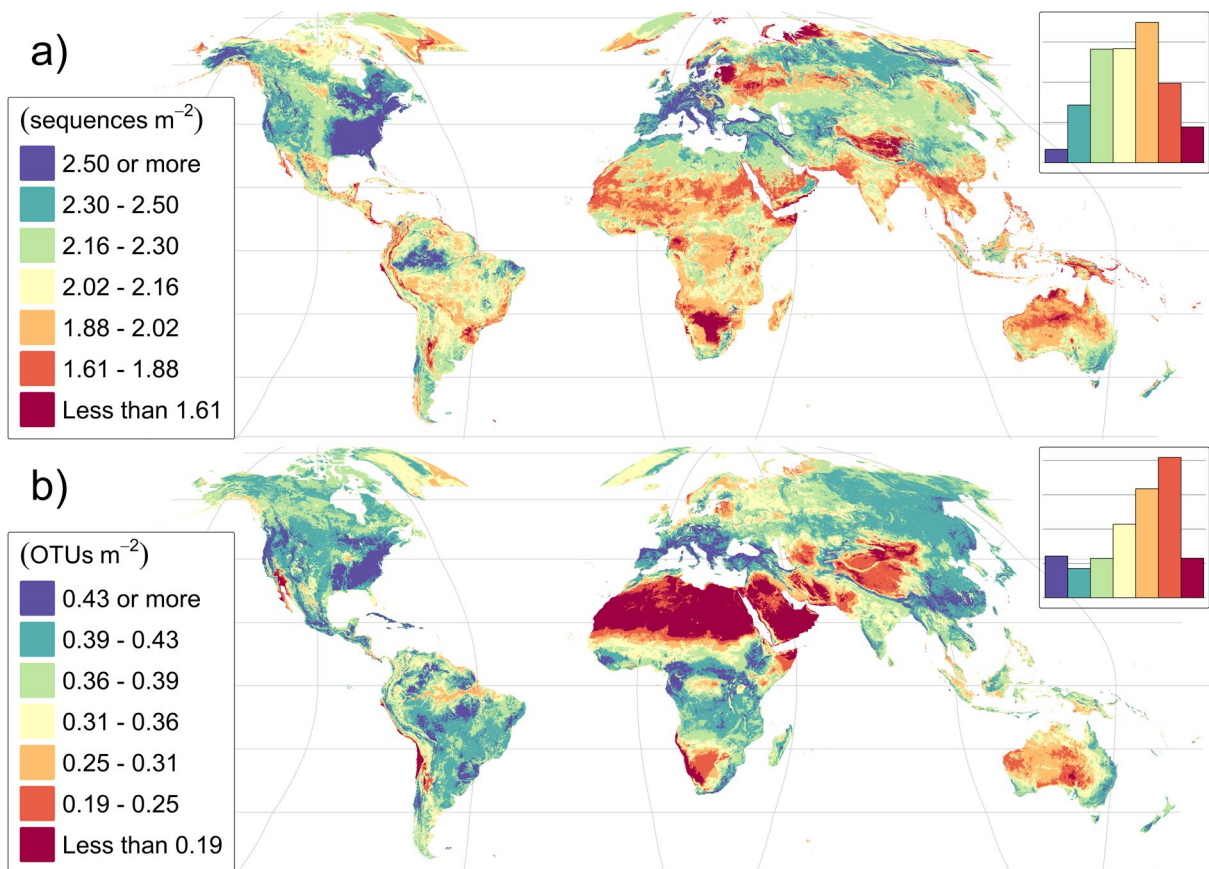
570 A distribuição espacial da diversidade fúngica nos primeiros 5 cm do solo indicou
 571 padrões distintos entre riqueza e abundância, ambos com forte estrutura biogeográfica e
 572 controlados por diferentes conjuntos de fatores ambientais (Figura 4). A riqueza de fungos foi
 573 mais elevada em ecorregiões tropicais, como a Floresta Amazônica, Florestas Tropicais do
 574 Congo, Mata Atlântica e regiões do Sudeste Asiático. Em contrapartida, os menores valores
 575 foram observados em zonas áridas e semiáridas (incluindo os desertos do Saara, Árábico,
 576 Atacama, Kalahari, Namíbia e a Austrália Central) além de áreas montanhosas e frias, como os
 577 Andes e o Planalto Tibetano. Esses padrões refletiram a influência de grandes zonas
 578 ecoclimáticas e gradientes combinados de temperatura, umidade e cobertura vegetal.

579 A abundância de fungos apresentou tendências espaciais parcialmente distintas daquelas
 580 observadas para a riqueza. Ecorregiões de clima temperado e frio úmido, como as florestas

581 boreais do Canadá, Escandinávia e Rússia, indicaram elevada abundância, apesar de valores
 582 relativamente baixos de riqueza. Já em regiões tropicais, como o Sudeste Asiático e porções da
 583 América do Sul, observou-se a coexistência de alta riqueza e alta abundância. Por sua vez, áreas
 584 áridas e semiáridas mantiveram baixos valores em ambas as métricas.

585 Em conjunto, os mapas apontaram que riqueza e abundância fúngica responderam de
 586 maneira complementar às condições ambientais. A riqueza foi maximizada em regiões tropicais
 587 com elevada heterogeneidade ecológica, enquanto a abundância atingiu seus maiores valores
 588 em ambientes úmidos e frios, com acúmulo de matéria orgânica e menor atividade
 589 decompositora. Esses padrões reforçam a necessidade de tratar essas duas dimensões da
 590 diversidade de forma independente em abordagens ecológicas e preditivas.

591



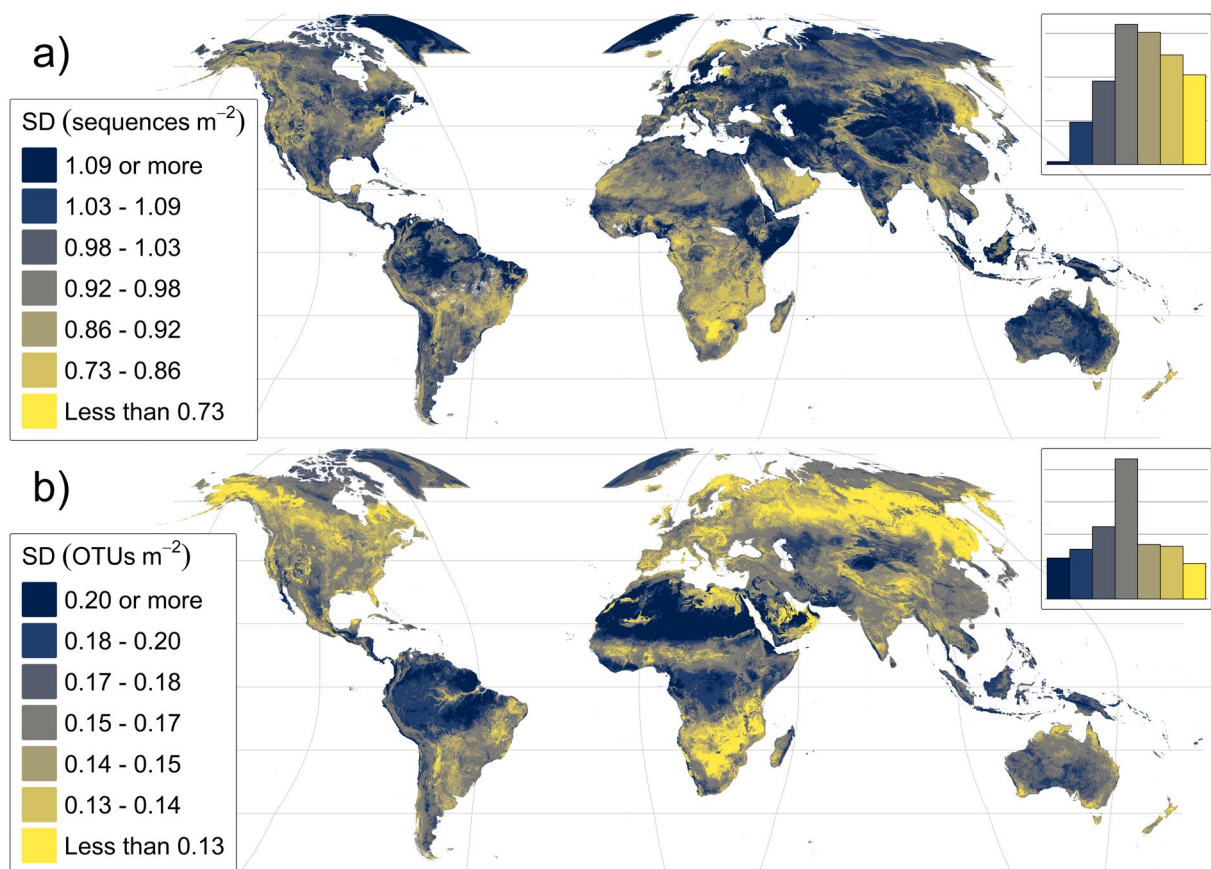
592
 593
 594
 595

Figura 4. Mapas médios da abundância (a) e riqueza (b) de fungos nos primeiros 5 cm de solo modelados com quantile random forest.

596 Os mapas de incerteza (Figura 5) indicaram ampla variação no desvio padrão das
 597 estimativas de diversidade fúngica. Para a riqueza, os valores oscilaram entre $\leq 0,13$ e $\geq 0,20$
 598 OTUs m^{-2} , com os maiores desvios concentrados em regiões desérticas e subdesérticas (como
 599 o Saara, Península Arábica, Irã, deserto da Austrália, Atacama e Kalahari), além de porções da
 600 Amazônia ocidental e oriental e áreas ao sul da América do Sul.

601 Na abundância, observaram-se padrões distintos, com destaque para o leste da África,
 602 onde os desvios padrão ultrapassaram 1 sequences m^{-2} . Elevações nos desvios também
 603 ocorreram na Ásia Central, em partes do leste europeu e da Oceania, bem como em áreas ao sul
 604 da América do Sul. Em ambas as métricas, a incerteza foi mais acentuada em zonas de transição
 605 climática ou com baixa cobertura de dados, refletindo variação espacial considerável nas
 606 previsões do modelo.

607



608

609 **Figura 5.** Mapas do desvio padrão da abundância (a) e riqueza (b) de fungos nos primeiros 5
 610 cm de solo modelados com quantile random forest.

611

612

613

614

615

616

617

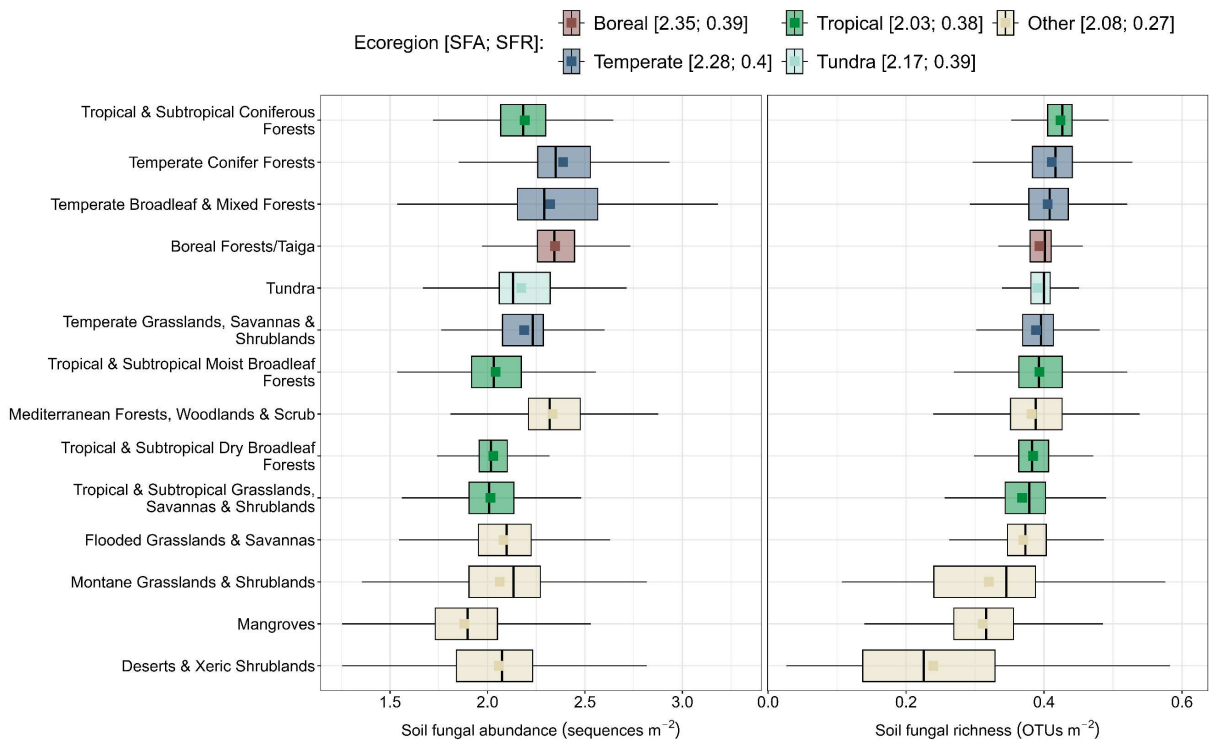
618

619

620

621

Apesar da similaridade nas tendências gerais entre riqueza e abundância de fungos nas diferentes ecorregiões, observou-se que regiões com maior riqueza não apresentaram, necessariamente, os maiores valores de abundância, e vice-versa (Figura 6). A ecorregião boreal destacou-se com a maior abundância média estimada, enquanto as ecorregiões tropical e temperada registraram os maiores valores de riqueza. Em termos de variação intra-ecorregião, a região tropical apresentou uma amplitude relativamente menor quando comparada à região temperada. Já em ambientes com condições ambientais mais severas, como a tundra e os desertos, os valores de riqueza e abundância apresentaram elevada variabilidade, com amplitudes mais acentuadas, indicando forte heterogeneidade nesses sistemas.



622

623

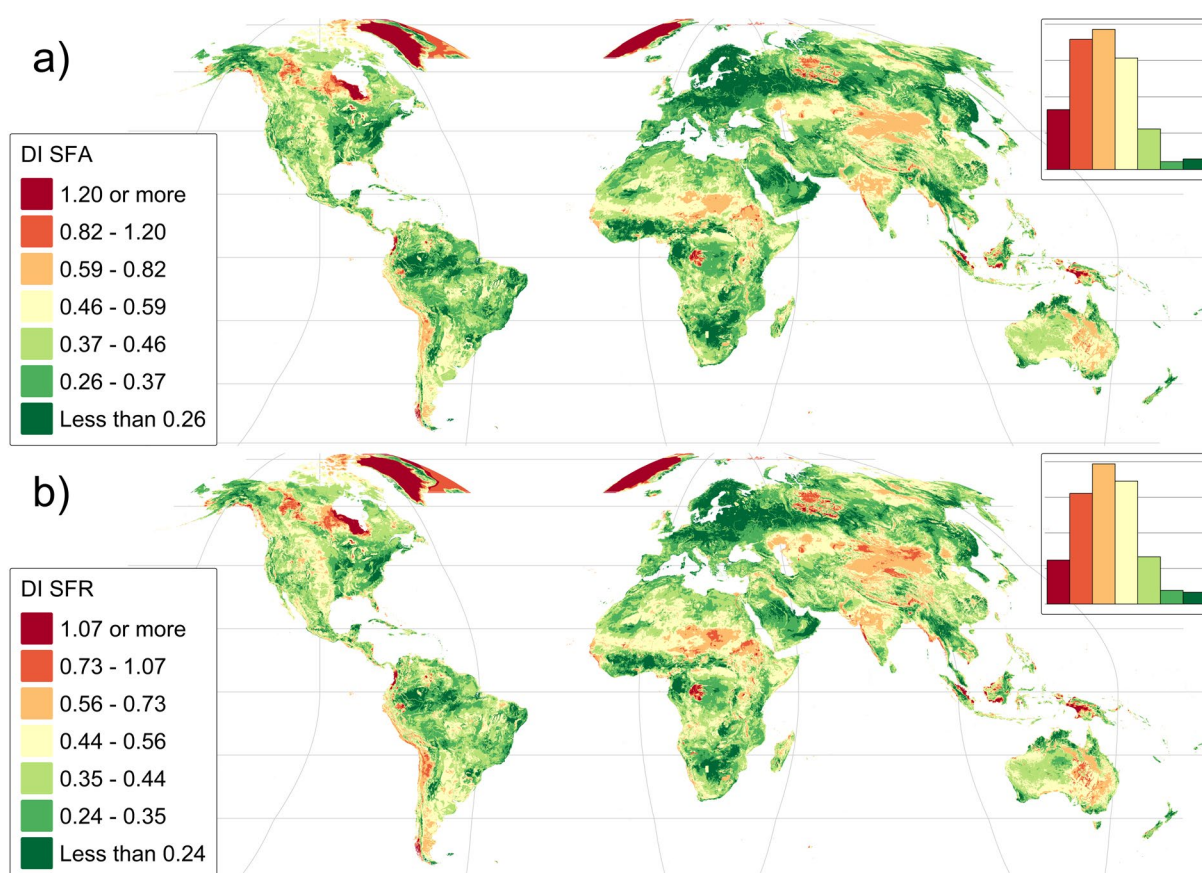
624

625

626

Figura 6. Abundância e riqueza de fungos nos primeiros 5 cm do solo modelados com quantile random forest nos diferentes ecorregiões globais. Valores entre colchetes representam a média, Ab = Abundância de fungos no solo e Ri = Riqueza de fungos no solo.

627 O Índice de Dissimilaridade (DI) para a abundância e para a riqueza de fungos
628 apresentou padrões espaciais distintos, com maiores valores registrados em regiões de alta
629 latitude, como a tundra ártica, o Planalto Tibetano e as áreas montanhosas dos Andes, além de
630 zonas áridas como o deserto do Saara, a Península Arábica e a Austrália Central (Figura 7).
631 Valores mais baixos de DI concentraram-se em regiões tropicais e subtropicais, como a Floresta
632 Amazônica, a Bacia do Congo e o Sudeste Asiático.
633



634 **Figura 7.** Mapas do Índice de Dissimilaridade (DI) da abundância (a) e riqueza (b) de fungos
635 nos primeiros 5 cm de solo modelados com quantile random forest.
636
637

638 A distribuição espacial dos valores de DI apresentou correspondência parcial com os
639 padrões de incerteza descritos nos mapas de desvio padrão (Figura 5), onde as maiores
640 incertezas para a riqueza foram observadas em zonas áridas e, para a abundância, em regiões

641 como o leste da África, a Ásia Central e a Oceania. Nessas áreas, a elevação simultânea dos
642 desvios padrão e dos valores de DI indicou maior extrapolação do modelo.

643 Nas regiões tropicais, especialmente na Amazônia e no Sudeste Asiático, observaram-
644 se baixos valores de DI combinados a baixos desvios padrão, consistentes com as áreas de maior
645 riqueza e abundância fúngica previamente identificadas. Em contraste, ambientes de clima frio
646 e árido, que já haviam apresentado baixos valores absolutos de diversidade e alta variabilidade
647 intra-ecorregional, concentraram também os maiores valores de dissimilaridade ambiental em
648 relação aos dados de treinamento.

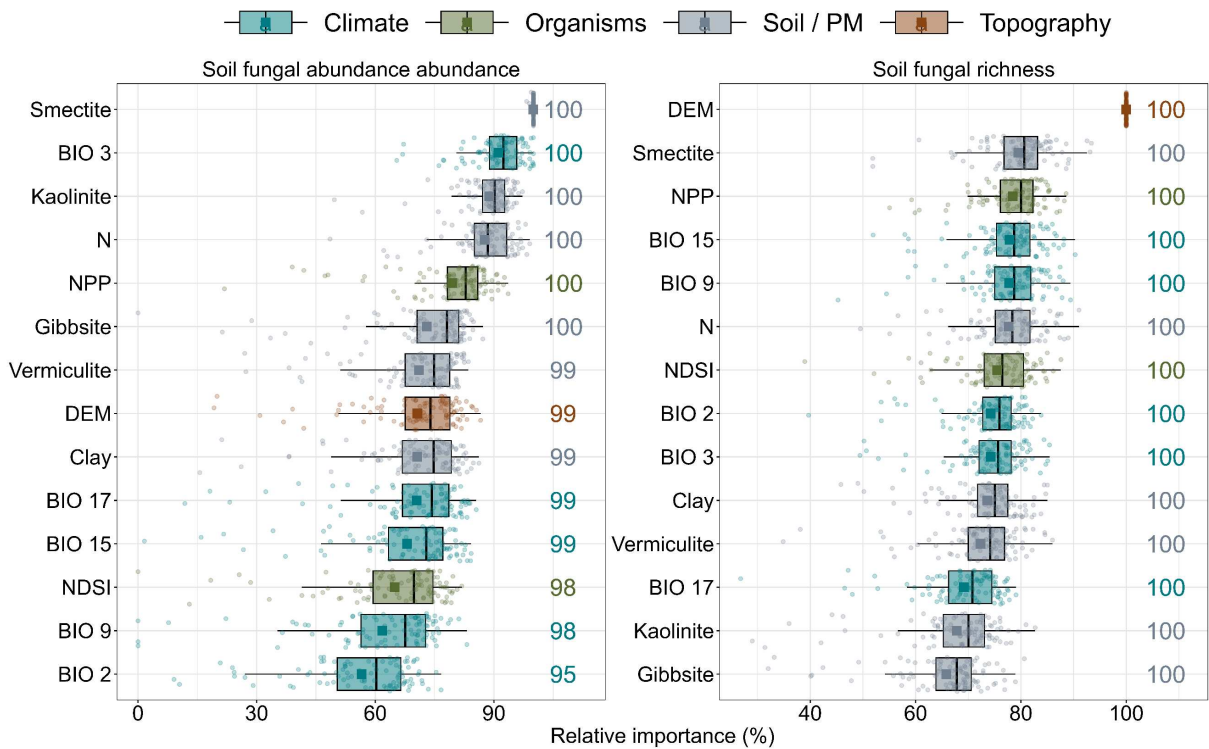
649

650

651 **3.3. Moduladores de fungos no solo e a interação entre riqueza e abundância**

652 Para a abundância, destacaram-se como os principais preditores a esmectita, o índice de
653 isotermalidade (BIO 3), a caulinita, o teor de nitrogênio do solo e a produtividade primária
654 líquida (NPP), todos com alta frequência de seleção (Figura 8). A presença de minerais como
655 gibbsita e vermiculita, além do modelo digital de elevação (DEM), também apresentou elevada
656 contribuição (>85%). De maneira geral, variáveis climáticas, relacionadas ao material de
657 origem do solo e à vegetação estiveram entre os preditores mais importantes para a abundância
658 fúngica.

659



660
661
662
663
664
665
666

Figura 8. Importância relativa e frequência de seleção dos preditores mais importantes para o modelo quantile random forest na modelagem da riqueza e abundância de fungos nos 5 cm superficiais do solo globalmente. Os números à direita indicam a frequência de seleção do preditor, e os quadrados coloridos representam a média da importância relativa. MO = Material de origem.

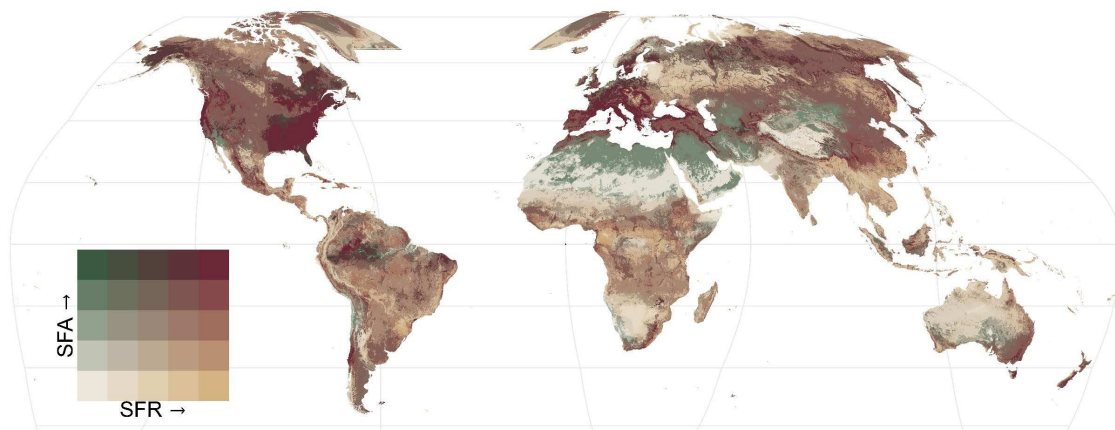
667
668
669
670
671
672
673
674
675

Para a riqueza, os preditores com maior importância média incluíram o DEM, a esmectita, a NPP, a precipitação do trimestre mais seco (BIO 17) e a temperatura média do trimestre mais seco (BIO 9), todos com frequência de seleção igual a 100%. Preditores biológicos (como NPP e nitrogênio), climáticos e relacionados ao material de origem também dominaram a composição dos fatores mais relevantes, com destaque adicional para o índice espectral NDSI e o teor de argila. Os resultados indicaram que os preditores mais influentes variaram entre os modelos de abundância e riqueza, com sobreposição parcial entre os grupos de variáveis selecionadas e predomínio de fatores climáticos e edáficos em ambas as respostas.

676
677

As maiores concentrações simultâneas de abundância e riqueza de fungos do solo foram observadas predominantemente no leste dos Estados Unidos, na Europa Central e na região central da Amazônia, com ocorrências pontuais em zonas de transição tropicais, como no sul

678 da América do Sul e em áreas meridionais da África (Figura 9). Em contraste, regiões
 679 caracterizadas por alta abundância, mas baixa riqueza, foram identificadas em partes das zonas
 680 temperadas, especialmente no centro-oeste dos Estados Unidos, no leste europeu, no norte do
 681 deserto do Saara e na região central do Oriente Médio.
 682



683 **Figura 9.** Mapa bivariado da riqueza (SFR) e abundância (SFA) de fungos nos primeiros 5 cm
 684 de solo modelados com quantile random forest - qrf.
 685
 686

687 Os menores valores de abundância e riqueza concentraram-se em ambientes áridos e
 688 frios, incluindo a região central do Saara, a Península Arábica, a Austrália central, o sul da
 689 Argentina, os Andes, a Groenlândia e extensas áreas da Sibéria e do Himalaia. Por outro lado,
 690 áreas com alta riqueza, mas baixa abundância, foram detectadas principalmente em florestas
 691 tropicais úmidas, destacando-se a Bacia do Congo, o Sudeste Asiático e partes da Amazônia.
 692

693 4. Discussão

694 4.1. Distribuição espacial de fungos no solo

695 Embora ambos os modelos tenham apresentado desempenho semelhante, a modelagem
 696 da riqueza apresentou ligeira superioridade em relação à abundância, o que pode refletir sua
 697 maior estabilidade frente à heterogeneidade amostral. A alta riqueza e abundância encontradas
 698 nas regiões de florestas tropicais podem ser explicadas pela diversidade acima do solo. As

699 florestas tropicais são conhecidas por conter uma composição diversa de espécies arbóreas, o
700 que implica em alta variabilidade bioquímica dos componentes provenientes das folhas e
701 exsudatos das raízes (Yaffar et al., 2024). Esses fatores impactam diretamente a comunidade
702 bacteriana e fúngica do solo, conferindo uma heterogeneidade de compostos disponíveis para a
703 biota do solo. Consequentemente, essa heterogeneidade propicia uma diversidade de nichos e
704 diversifica a comunidade microbiana (Buscardo et al., 2024).

705 Essas associações simbióticas entre plantas e fungos desempenham papel central na
706 construção e manutenção do carbono orgânico no solo, ao favorecerem maior acúmulo e
707 permanência da matéria orgânica. (Emilia Hannula & Morriën, 2022). No entanto, a
708 contribuição fúngica para a permanência do carbono não se limita às interações micorrízicas ou
709 à produção de glomalina. O micélio das hifas fúngicas atua diretamente no enovelamento de
710 partículas minerais e orgânicas, o que pode promover a formação e a estabilidade de agregados
711 do solo, uma estrutura física essencial para a proteção da matéria orgânica contra a ação de
712 microrganismos e enzimas degradativas (Rillig & Mummey, 2006). Além disso, fungos
713 decompositores liberam compostos extracelulares com propriedades adesivas e estabilizantes,
714 que reforçam a agregação e criam microambientes menos favoráveis à decomposição rápida da
715 matéria orgânica (J. Lehmann & Kleber, 2015). Essa atuação múltipla confere aos fungos um
716 papel funcional de destaque na dinâmica do carbono, especialmente em ambientes manejados.
717 Em solos agrícolas, por exemplo, estima-se que a biomassa fúngica represente até 40% da
718 matéria orgânica total (Liang et al., 2019), indicando que o manejo da atividade fúngica pode
719 ser estratégico para aumentar a resiliência dos estoques de carbono.

720 Ao contrário da relativa estabilidade climática das florestas tropicais, as regiões áridas
721 e desérticas enfrentam sazonalidades severas e climas considerados inóspitos para a maioria da
722 vida (Noy-Meir, 1973). As condições adversas, especialmente a escassez hídrica e as
723 temperaturas extremas, tornam esses ambientes locais onde a comunidade fúngica é

724 especificamente adaptada, reduzindo a riqueza e aumentando a abundância em proporções
725 relativas (Ameen et al., 2022; Kang et al., 2023). Em regiões com condições extremas, como o
726 deserto do Saara, Austrália e Arábia Saudita, a riqueza de fungos é ainda menor, e a abundância
727 limita-se a uma comunidade fúngica muito específica e adaptada, com características como
728 morfologia com proteções, melanina nas paredes celulares, esporos de paredes espessas e hifas
729 capazes de atravessar os poros do solo cheios de ar para acessar nutrientes e água (Alwakeel et
730 al., 2023; Barnard et al., 2013). Os mapas nessas regiões corroboram a literatura, evidenciando
731 uma distribuição limitada e específica das comunidades fúngicas.

732

733 **4.2. Incertezas espaciais na distribuição de fungos no solo**

734 A transição sul-norte entre o Sahel e o Saara indicou uma redução esperada nos valores
735 de diversidade fúngica, tanto para riqueza quanto para abundância, acompanhando os
736 gradientes de aridez e severidade edáfica da região. Enquanto os valores de riqueza
737 apresentaram queda mais acentuada, a abundância exibiu um padrão mais distribuído, refletindo
738 possíveis diferenças na resposta ecológica de biomassa *versus* diversidade funcional frente à
739 limitação hídrica. Nas regiões temperadas da Europa, os valores de ambas as métricas
740 aumentaram, compatíveis com a maior estabilidade climática e a disponibilidade de matéria
741 orgânica (Tedersoo et al., 2014, 2021).

742 As regiões desérticas e de tundra continuaram associadas aos maiores desvios padrão
743 nas estimativas, sugerindo maior incerteza nas predições para ambientes com condições
744 edafoclimáticas extremas. Na Amazônia, a incerteza na riqueza manteve-se elevada,
745 possivelmente refletindo a heterogeneidade ambiental intrínseca às regiões tropicais.

746 A análise do Índice de Dissimilaridade (DI) reforçou os padrões de incerteza espacial
747 observados. Regiões áridas e frias, como o Saara, o Planalto Tibetano, a tundra ártica e o deserto
748 da Austrália, apresentaram simultaneamente altos desvios padrão e elevados valores de DI,

749 indicando extrapolação ambiental acentuada e maior fragilidade preditiva (Meyer & Pebesma,
750 2021). Essas áreas, caracterizadas por condições edafoclimáticas extremas e baixa
751 representatividade amostral, configuraram-se como zonas de maior extrapolação dos modelos
752 aqui ajustados.

753 Em contraste, regiões tropicais, especialmente como a Amazônia e partes do Sudeste
754 Asiático exibiram moderados a baixos valores de DI, mesmo em face de limitações amostrais
755 conhecidas. Esse padrão reforça que, embora a cobertura geográfica na região tropical tenha
756 sido limitada, a variabilidade ambiental capturada pela amostragem foi suficiente para permitir
757 o reconhecimento dos padrões pelo modelo. Modelos de aprendizado de máquina não exigem
758 necessariamente ampla cobertura espacial, mas sim a representação adequada da diversidade
759 ambiental nos dados de treinamento (Olden et al., 2008; Slik et al., 2015). Quando combinações
760 de variáveis ambientais semelhantes às dos trópicos estão presentes em outras regiões, o modelo
761 é capaz de reconhecer e reproduzir essas relações com baixa dissimilaridade e menores
762 incertezas.

763 Essa observação destaca a importância de design inteligentes de amostragem voltadas
764 para maximizar a variabilidade ambiental, como o uso do método *conditioned Latin Hypercube*
765 *Sampling* (cLHS), em oposição a abordagens aleatórias ou baseadas em grades regulares
766 (Minasny & McBratney, 2006). A adoção de delineamentos amostrais ambientalmente
767 estratificados é fundamental para fortalecer a área de aplicabilidade dos modelos e reduzir a
768 extrapolação, especialmente em ecorregiões complexas e logisticamente desafiadores como
769 parte dos trópicos.

770

771 **4.3. Estratificação dos fungos no solo e seus principais controladores**

772 Nas análises estratificadas por ecorregiões, observa-se que a ecorregião boreal
773 (caracterizado por altas latitudes, clima frio e relativa homogeneidade ambiental) tendeu a

774 apresentar elevados níveis de biomassa fúngica, mas com baixa variação entre os dados. Essa
775 uniformidade reflete não apenas as condições climáticas relativamente estáveis, mas também a
776 estrutura mais simplificada da vegetação, que reduz a variabilidade microambiental do solo (J.
777 Li et al., 2019), esse padrão confirma que biomassa fúngica elevada não implica,
778 necessariamente, maior diversidade funcional, especialmente em regiões homogêneas como as
779 florestas boreais. Em contraste, os trópicos, embora também ricos e abundantes em fungos,
780 exibiram uma amplitude maior nos valores estimados. Essa alta variabilidade reflete a
781 heterogeneidade ambiental e a complexidade estrutural das florestas tropicais, fatores que
782 influenciam diretamente a diversidade de nichos e a atividade fúngica no solo.

783 A influência combinada do clima e da vegetação torna-se ainda mais evidente em
784 ecorregiões extremas, como desertos e tundras. Apesar da baixa produtividade primária, essas
785 regiões apresentaram alta variabilidade nos valores de riqueza e abundância fúngica, o que
786 sugere a presença de nichos especializados e adaptações funcionais específicas à limitação
787 hídrica ou térmica (Ameen et al., 2022; Barnard et al., 2013).

788 O DEM, que apresentou 100% de frequência de seleção para riqueza, também merece
789 destaque. O DEM atua como *proxy* para diversos fatores ambientais, como drenagem, regime
790 térmico, acúmulo de matéria orgânica e heterogeneidade topográfica. Em ambientes tropicais
791 montanhosos, como Andes e Himalaia, ou mesmo em regiões de planície com variações
792 altimétricas sutis, mas ecologicamente relevantes, o relevo influencia a distribuição de fungos
793 ao moldar microclimas, o escoamento de água e a deposição de matéria orgânica (Chen et al.,
794 2022; Z. Li et al., 2024).

795 A importância atribuída aos preditores climáticos no modelo reforça o papel
796 determinante do clima, em particular, da sazonalidade térmica, na distribuição global dos
797 fungos do solo. A isotermalidade (BIO 3), um dos preditores mais relevantes, expressa a relação
798 entre a amplitude térmica diária e a anual. Altos valores de BIO 3 indicam grandes oscilações

799 de temperatura ao longo do dia em comparação com a variação sazonal. Essa métrica é
800 especialmente informativa para florestas tropicais, onde a sazonalidade, embora menos
801 marcada em termos de temperatura, regula fortemente os ciclos fenológicos das plantas (Malhi
802 et al., 2014), a atividade fotossintética (Guan et al., 2015) e a dinâmica da serrapilheira (Schaap
803 et al., 2021).

804 Durante a estação chuvosa, há um aumento expressivo na produção de biomassa vegetal,
805 especialmente de raízes finas, além de maior deposição de serrapilheira. Esses processos
806 favorecem a atividade microbiana, promovem o crescimento fúngico e intensificam a
807 respiração do solo, a decomposição e a mineralização da matéria orgânica (Cornejo et al., 1994).
808 No entanto, mesmo nos períodos de menor precipitação, a combinação entre temperaturas
809 elevadas e umidade residual mantém uma atividade biológica substancial, especialmente em
810 florestas tropicais, onde a sazonalidade hídrica é menos extrema (Fujii et al., 2018). Essa
811 persistência funcional ao longo do ano reforça o papel da sazonalidade moderada combinada a
812 altas temperaturas como um importante modulador da abundância e riqueza fúngica, o que é
813 coerente com os padrões climáticos capturados pelos preditores isothermalidade e sazonalidade
814 no modelo.

815 Relativo aos preditores do grupo Soil/PM, os atributos relacionados à mineralogia e ao
816 teor de nitrogênio se destacaram. Isso se explica pela influência principal desses minerais e do
817 teor de N na estrutura e fertilidade do solo. A conectividade dos poros entre as estruturas do
818 solo facilita a aeração e a condução de água e crescimento das hifas fúngicas, que, por sua vez,
819 condicionam a colonização de fungos (A. Lehmann et al., 2020). Além disso, a disponibilidade
820 de nutrientes é outro fator-chave para a biomassa fúngica (Rola et al., 2023), tanto em termos
821 de riqueza quanto de abundância. Por exemplo, Han et al. (2020) encontraram que com o
822 aumento das quantidades de N no solo, tanto a abundância quanto a riqueza de fungos
823 micorrízicos diminuíram significativamente. Ma et al. (2023) relataram resultados similares ao

824 avaliar o fósforo disponível no solo; conforme a quantidade de fósforo aumentou, a diversidade
825 e abundância de fungos diminuíram. Além disso, os diferentes tamanhos de agregados no solo
826 fornecem superfícies de interação microbiana essenciais para os fungos, o que também
827 influencia sua distribuição e diversidade (Kuzyakov & Blagodatskaya, 2015).

828

829 **4.4. Interação entre abundância e riqueza de fungos no solo**

830 Apesar de o clima ser um moderador importante da microbiota do solo, pois sem
831 temperatura e umidade as interações tróficas podem ser atenuadas, o mapa bivariado reafirma
832 a importância da interação entre os fatores edáficos e a vegetação no controle da riqueza e
833 abundância de fungos no solo. Ao contrário da comunidade bacteriana, que é fortemente
834 controlada pelo clima (Labouyrie et al., 2023), os fungos no solo são mais moderados pela
835 vegetação e pelo solo (Bahram et al., 2018; Fierer et al., 2009; Tedersoo et al., 2014; Yamauchi
836 et al., 2021; Yang et al., 2024; Zeng et al., 2020). A importância atribuída às variáveis
837 relacionadas à vegetação e ao solo pelo modelo reafirma isso. O sudeste dos Estados Unidos, a
838 Europa e a Amazônia, ao apresentarem elevada riqueza e abundância, são consistentes com
839 essas hipóteses.

840 O fato de os fungos responderem mais à vegetação e ao solo se deve ao fato de que
841 grande parte dos fungos possui associações com plantas (por exemplo, endofitismo e
842 mutualismo por fungos micorrízicos (Hooper et al., 2000; Prober et al., 2015; Urbanová et al.,
843 2015), especialmente árvores. No entanto, isso é fortemente modulado por atributos do solo,
844 como fertilidade, estrutura e pH (Fierer et al., 2009; Yang et al., 2024; Zeng et al., 2020). As
845 florestas tropicais, com alta diversidade arbórea e aporte de matéria orgânica no solo, e as
846 florestas temperadas, com microclimas amenos e solos com *hotspots* de biodiversidade, ajudam
847 a explicar os altos valores dessas áreas.

848 O uso intensivo do solo, particularmente em sistemas agrícolas, pode ter impactos
849 significativos na riqueza e abundância de fungos do solo. A aplicação frequente de fertilizantes
850 e agrotóxicos pode alterar a composição química do solo e afetar negativamente as
851 comunidades fúngicas. Fertilizantes ricos em nitrogênio e fósforo, embora possam aumentar a
852 produtividade das culturas, podem reduzir a diversidade de fungos micorrízicos arbusculares,
853 que são essenciais para a saúde das plantas (Han et al., 2020; Ma et al., 2023). Além disso,
854 agrotóxicos podem ter efeitos tóxicos diretos sobre os fungos do solo, reduzindo sua biomassa
855 e alterando as interações ecológicas no solo (Rola et al., 2023).

856 Além da detecção de padrões globais na distribuição da abundância e riqueza de fungos
857 do solo, os mapas gerados neste estudo podem servir como referência de base para o
858 monitoramento de mudanças futuras. Ao identificar áreas com alta diversidade fúngica
859 associadas à vegetação natural preservada (como florestas tropicais úmidas, porções da Bacia
860 do Congo, Amazônia e Sudeste Asiático) os resultados oferecem marcos espaciais para avaliar
861 impactos de mudanças no uso da terra, intensificação agrícola ou variações climáticas. Dessa
862 forma, os mapas não apenas descrevem o estado atual da biomassa fúngica, mas também
863 apontam regiões prioritárias para estudos de longo prazo e áreas-sentinela para políticas de
864 conservação e restauração.

865 Esses achados reforçam que estratégias conservacionistas voltadas à proteção da
866 cobertura vegetal, aliadas a práticas adequadas de manejo do solo, são mais eficazes para
867 garantir a multifuncionalidade e a saúde dos solos, promovendo comunidades fúngicas diversas
868 e funcionais. A adoção de práticas sustentáveis, como agroecossistemas de base agroecológica,
869 pode mitigar os efeitos negativos do uso intensivo da terra e favorecer um ecossistema do solo
870 mais resiliente e equilibrado. No entanto, estudos em larga escala que investiguem de forma
871 integrada os efeitos de fatores bióticos, climáticos e de manejo sobre a diversidade e
872 funcionalidade do solo ainda são limitados (Bahram et al., 2018; Tedersoo et al., 2014).

873 As análises realizadas podem, portanto, estabelecer as bases para um novo paradigma
874 na preservação das comunidades microbianas do solo e na promoção dos serviços
875 ecossistêmicos fornecidos pelos grupos funcionais do solo. Os mapas gerados podem subsidiar
876 informações essenciais para modelos globais de carbono no solo, sendo utilizados como uma
877 covariável representativa dos fungos do solo. Dados relacionados à microbiologia do solo são
878 frequentemente negligenciados em tais modelos (Gomes et al., 2019; Hengl et al., 2015, 2017;
879 Moquedace et al., 2024; Poggio et al., 2021; Siqueira et al., 2024). Além disso, esses dados
880 podem fundamentar a compreensão das funções ecológicas dos fungos do solo no contexto das
881 mudanças climáticas.

882 Os solos são habitats biologicamente ativos e complexos, repletos de organismos vivos
883 que sustentam funções essenciais para os ecossistemas terrestres (Maestre et al., 2012; Wagg et
884 al., 2014). Esses organismos promovem a ciclagem de nutrientes, o armazenamento e a
885 renovação do carbono, a retenção de água, a estabilidade da estrutura física, a regulação da
886 diversidade e produtividade das plantas, além do controle biótico e da atenuação de compostos
887 potencialmente tóxicos (Blume et al., 2015). A capacidade do solo em manter essas funções ao
888 longo do tempo define seu estado de saúde, entendido como a habilidade contínua de sustentar
889 a produtividade biológica, promover a qualidade ambiental e preservar a saúde de plantas,
890 animais e seres humanos (Doran & Zeiss, 2000).

891 Nesse contexto, indicadores biológicos como a abundância e a riqueza de fungos do solo
892 assumem papel central na avaliação funcional da saúde do solo. A riqueza fúngica está
893 associada à redundância funcional e à resiliência ecológica, enquanto a abundância reflete a
894 intensidade de processos fundamentais, como decomposição, formação de agregados e
895 permanência da matéria orgânica. Ao mapear espacialmente esses atributos em escala global,
896 este estudo propõe uma abordagem funcional para compreender e monitorar a saúde do solo,
897 associando diretamente proxies biológicos à provisão de serviços ecossistêmicos.

898 Essa perspectiva está alinhada ao conceito de segurança do solo (*soil security*), que
899 reconhece a importância estratégica do solo para a segurança alimentar, hídrica e climática (A.
900 McBratney et al., 2014). A identificação de regiões com comunidades fúngicas diversas e
901 abundantes pode orientar ações de conservação, restauração e manejo sustentável, contribuindo
902 para proteger a funcionalidade do solo diante das crescentes pressões antrópicas e das mudanças
903 ambientais globais.

904

905 **4.5. Limitação e futuras direções sobre mapeamento da biodiversidade do solo em** 906 **escala global**

907 A elevada concentração de dados em regiões temperadas resulta de uma combinação de
908 fatores logísticos, operacionais e estruturais. A amostragem exigiu o deslocamento a longas
909 distâncias, coleta de material fresco e posterior análise molecular em laboratórios com
910 infraestrutura especializada — um processo dispendioso e tecnicamente exigente (Tedersoo et
911 al., 2021). Essas limitações restringiram a representatividade espacial do banco de dados,
912 mesmo diante dos esforços colaborativos para abarcar todos as ecorregiões e fitofisionomias.

913 Como consequência, regiões como savanas, manguezais e grande parte das áreas
914 tropicais permaneceram subamostradas, o que impactou diretamente os resultados do
915 mapeamento e comprometeu, em certa medida, a acurácia das estimativas.

916 Relativo à performance dos modelos ajustados, as métricas sugerem um ajuste
917 adequado, em linha com outros autores que utilizaram o *quantile random forest* para modelar
918 atributos do solo (Gomes et al., 2019; Moquedace et al., 2024; Siqueira et al., 2023, 2024).
919 Atributos relacionados à parte viva do solo, como os fungos, apresentam elevada variação
920 natural, o que, associado à interação com múltiplos fatores, dificulta o estabelecimento de
921 padrões e o alcance de modelos com graus de acurácia mais elevados. van den Hoogen et al.
922 (2019), por exemplo, alcançaram R^2 menores que 0,5 na validação cruzada ao modelar a

923 abundância de nematoides no solo em escala global. Esse desbalanceamento amostral, como
924 observado entre ecorregiões temperadas e savanas, compromete a robustez dos padrões
925 identificados para determinadas regiões e impõe cautela à extrapolação ecológica dos modelos
926 nessas áreas.

927

928 **5. Conclusão**

929 A distribuição desbalanceada de dados, especialmente em regiões tropicais e de climas
930 extremos, compromete a compreensão e a modelagem da biodiversidade fúngica do solo.
931 Apesar do bom desempenho obtido com o uso do algoritmo quantile random forest, a alta
932 variabilidade intrínseca da comunidade fúngica e a escassez de amostras representativas
933 limitam a precisão das estimativas, sobretudo em ecorregiões subamostradas. Nas florestas
934 tropicais, a elevada riqueza e abundância de fungos está associada à complexidade estrutural da
935 vegetação, à diversidade de interações microbianas e à intensa deposição de matéria orgânica,
936 que promove a oferta contínua de recursos e nichos funcionais.

937 Por outro lado, regiões áridas e desérticas, marcadas por estresse hídrico e baixa
938 cobertura vegetal, tendem a abrigar comunidades fúngicas mais especializadas e de menor
939 diversidade. Fatores edáficos como a mineralogia do solo e a disponibilidade de nutrientes
940 desempenham papel determinante nessa dinâmica, influenciando diretamente a formação de
941 agregados, a retenção de matéria orgânica e a atividade enzimática microbiana, que por sua vez
942 afetam a colonização e persistência de fungos sapróbios e simbiontes.

943 As incertezas observadas nos mapas reforçam a necessidade de ampliar a
944 representatividade espacial das amostras, especialmente em ecossistemas negligenciados.
945 Práticas agrícolas sustentáveis tornam-se fundamentais não apenas para manter a saúde do solo,
946 mas também para preservar comunidades fúngicas diversas e funcionais, essenciais à
947 multifuncionalidade do sistema edáfico. Este estudo reforça a importância de integrar dados

948 microbiológicos em modelos globais de carbono e destaca a urgência de abordagens
949 conservacionistas voltadas à biodiversidade do solo e aos serviços ecossistêmicos a ela
950 associados.

951

952 **Referências**

953 Almario, J., Muller, D., Défago, G., & Moëgne-Loccoz, Y. (2014). Rhizosphere ecology and
954 phytprotection in soils naturally suppressive to Thielaviopsis black root rot of tobacco.
955 *Environmental Microbiology*, 16(7), 1949–1960. [https://doi.org/10.1111/1462-](https://doi.org/10.1111/1462-2920.12459)
956 2920.12459

957 Alwakeel, S., Alothman, N., Ameen, F., Alotaibi, M., Mohammed, A. E., & Alhomaidi, E.
958 (2023). Stress-driven metabolites of desert soil fungi. *Biotechnology and Genetic*
959 *Engineering Reviews*. <https://doi.org/10.1080/02648725.2023.2182537>

960 Ameen, F., AlNadhari, S., Yassin, M. A., Al-Sabri, A., Almansob, A., Alqahtani, N., &
961 Stephenson, S. L. (2022). Desert soil fungi isolated from Saudi Arabia: cultivable fungal
962 community and biochemical production. *Saudi Journal of Biological Sciences*, 29(4),
963 2409–2420. <https://doi.org/10.1016/j.sjbs.2021.12.011>

964 Bahram, M., Hildebrand, F., Forslund, S. K., Anderson, J. L., Soudzilovskaia, N. A., Bodegom,
965 P. M., Bengtsson-Palme, J., Anslan, S., Coelho, L. P., Harend, H., Huerta-Cepas, J.,
966 Medema, M. H., Maltz, M. R., Mundra, S., Olsson, P. A., Pent, M., Pölme, S., Sunagawa,
967 S., Ryberg, M., ... Bork, P. (2018). Structure and function of the global topsoil
968 microbiome. *Nature* 2018 560:7717, 560(7717), 233–237.
969 <https://doi.org/10.1038/s41586-018-0386-6>

970 Baldrian, P. (2019). The known and the unknown in soil microbial ecology. *FEMS Microbiology*
971 *Ecology*, 95(2), 5. <https://doi.org/10.1093/femsec/fiz005>

- 972 Barnard, R. L., Osborne, C. A., & Firestone, M. K. (2013). Responses of soil bacterial and
973 fungal communities to extreme desiccation and rewetting. *ISME Journal*, 7(11), 2229–
974 2241. <https://doi.org/10.1038/ismej.2013.104>
- 975 Blackwell, M. (2011). The fungi: 1, 2, 3 ... 5.1 million species? *American Journal of Botany*,
976 98(3), 426–438. <https://doi.org/10.3732/ajb.1000298>
- 977 Blume, H. P., Brümmer, G. W., Fleige, H., Horn, R., Kandeler, E., Kögel-Knabner, I.,
978 Kretzschmar, R., Stahr, K., & Wilke, B. M. (2015). Scheffer/schachtschabel soil science.
979 *Scheffer/Schachtschabel Soil Science*, 1–618. [https://doi.org/10.1007/978-3-642-30942-](https://doi.org/10.1007/978-3-642-30942-7/COVER)
980 [7/COVER](https://doi.org/10.1007/978-3-642-30942-7/COVER)
- 981 Bünemann, E. K., Bongiorno, G., Bai, Z., Creamer, R. E., De Deyn, G., de Goede, R., Fleskens,
982 L., Geissen, V., Kuyper, T. W., Mäder, P., Pulleman, M., Sukkel, W., van Groenigen, J. W.,
983 & Brussaard, L. (2018). Soil quality – A critical review. *Soil Biology and Biochemistry*,
984 120, 105–125. <https://doi.org/10.1016/j.soilbio.2018.01.030>
- 985 Buscardo, E., Geml, J., & Nagy, L. (2024). Seasonal dependence of deterministic versus
986 stochastic processes influencing soil fungal community composition in a lowland
987 Amazonian rain forest. *Communications Earth and Environment*, 5(1), 1–11.
988 <https://doi.org/10.1038/s43247-024-01273-2>
- 989 Cavicchioli, R., Ripple, W. J., Timmis, K. N., Azam, F., Bakken, L. R., Baylis, M., Behrenfeld,
990 M. J., Boetius, A., Boyd, P. W., Classen, A. T., Crowther, T. W., Danovaro, R., Foreman,
991 C. M., Huisman, J., Hutchins, D. A., Jansson, J. K., Karl, D. M., Koskella, B., Mark Welch,
992 D. B., ... Webster, N. S. (2019). Scientists' warning to humanity: microorganisms and
993 climate change. *Nature Reviews Microbiology*, 17(9), 569–586.
994 <https://doi.org/10.1038/s41579-019-0222-5>
- 995 Chen, J., Shi, Z., Liu, S., Zhang, M., Cao, X., Chen, M., Xu, G., Xing, H., Li, F., & Feng, Q.
996 (2022). Altitudinal Variation Influences Soil Fungal Community Composition and

- 997 Diversity in Alpine–Gorge Region on the Eastern Qinghai–Tibetan Plateau. *Journal of*
998 *Fungi*, 8(8), 807. <https://doi.org/10.3390/jof8080807>
- 999 Cornejo, F. H., Varela, A., & Wright, S. J. (1994). Tropical Forest Litter Decomposition under
1000 Seasonal Drought: Nutrient Release, Fungi and Bacteria. *Oikos*, 70(2), 183.
1001 <https://doi.org/10.2307/3545629>
- 1002 De Vries, F. T., & Shade, A. (2013). Controls on soil microbial community stability under
1003 climate change. *Frontiers in Microbiology*, 4(SEP), 52465.
1004 <https://doi.org/10.3389/fmicb.2013.00265>
- 1005 Delgado-Baquerizo, M., Maestre, F. T., Reich, P. B., Jeffries, T. C., Gaitan, J. J., Encinar, D.,
1006 Berdugo, M., Campbell, C. D., & Singh, B. K. (2016). Microbial diversity drives
1007 multifunctionality in terrestrial ecosystems. *Nature Communications*, 7(1), 1–8.
1008 <https://doi.org/10.1038/ncomms10541>
- 1009 Doran, J. W., & Zeiss, M. R. (2000). Soil health and sustainability: Managing the biotic
1010 component of soil quality. *Applied Soil Ecology*, 15(1), 3–11.
1011 [https://doi.org/10.1016/S0929-1393\(00\)00067-6](https://doi.org/10.1016/S0929-1393(00)00067-6)
- 1012 Downing, A. L., & Leibold, M. A. (2010). Species richness facilitates ecosystem resilience in
1013 aquatic food webs. *Freshwater Biology*, 55(10), 2123–2137.
1014 <https://doi.org/10.1111/j.1365-2427.2010.02472.x>
- 1015 Emilia Hannula, S., & Morriën, E. (2022). Will fungi solve the carbon dilemma? *Geoderma*,
1016 413, 115767. <https://doi.org/10.1016/j.geoderma.2022.115767>
- 1017 Fan, Y., Li, H., & Miguez-Macho, G. (2013). Global patterns of groundwater table depth.
1018 *Science*, 339(6122), 940–943. <https://doi.org/10.1126/science.1229881>
- 1019 Fierer, N., Strickland, M. S., Liptzin, D., Bradford, M. A., & Cleveland, C. C. (2009). Global
1020 patterns in belowground communities. *Ecology Letters*, 12(11), 1238–1249.
1021 <https://doi.org/10.1111/j.1461-0248.2009.01360.x>

- 1022 Fitter, A. H. (2005). Darkness visible: Reflections on underground ecology. *Journal of Ecology*,
1023 93(2), 231–243. <https://doi.org/10.1111/j.0022-0477.2005.00990.x>
- 1024 Fujii, K., Shibata, M., Kitajima, K., Ichie, T., Kitayama, K., & Turner, B. L. (2018). Plant–soil
1025 interactions maintain biodiversity and functions of tropical forest ecosystems. *Ecological*
1026 *Research*, 33(1), 149–160. <https://doi.org/10.1007/s11284-017-1511-y>
- 1027 Gaston, K. J. (2000). Global patterns in biodiversity. *Nature*, 405(6783), 220–227.
1028 <https://doi.org/10.1038/35012228>
- 1029 Gessner, M., Gulis, V., Kuehn, K., Chauvet, E., & Suberkropp, K. (2007). Fungal Decomposers
1030 of Plant Litter in Aquatic Ecosystems. Em C. Kubicek & I. Druzhinina (Orgs.),
1031 *Environmental and Microbial Relationships* (p. 301–324). Springer Berlin Heidelberg.
1032 https://doi.org/10.1007/978-3-540-71840-6_17
- 1033 Gomes, L. C., Faria, R. M., de Souza, E., Veloso, G. V., Schaefer, C. E. G. R., & Filho, E. I. F.
1034 (2019). Modelling and mapping soil organic carbon stocks in Brazil. *Geoderma*,
1035 340(January), 337–350. <https://doi.org/10.1016/j.geoderma.2019.01.007>
- 1036 Guan, K., Pan, M., Li, H., Wolf, A., Wu, J., Medvigy, D., Caylor, K. K., Sheffield, J., Wood, E.
1037 F., Malhi, Y., Liang, M., Kimball, J. S., Saleska, S. R., Berry, J., Joiner, J., & Lyapustin, A.
1038 I. (2015). Photosynthetic seasonality of global tropical forests constrained by
1039 hydroclimate. *Nature Geoscience*, 8(4), 284–289. <https://doi.org/10.1038/ngeo2382>
- 1040 Han, Y., Feng, J., Han, M., & Zhu, B. (2020). Responses of arbuscular mycorrhizal fungi to
1041 nitrogen addition: A meta-analysis. *Global Change Biology*, 26(12), 7229–7241.
1042 <https://doi.org/10.1111/gcb.15369>
- 1043 Hawker, L., Uhe, P., Paulo, L., Sosa, J., Savage, J., Sampson, C., & Neal, J. (2022). A 30 m
1044 global map of elevation with forests and buildings removed. *Environmental Research*
1045 *Letters*, 17(2), 024016. <https://doi.org/10.1088/1748-9326/ac4d4f>

- 1046 Hengl, T., De Jesus, J. M., Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A.,
1047 Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A.,
1048 Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I.,
1049 Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based
1050 on machine learning. *PLoS ONE*, *12*(2), e0169748.
1051 <https://doi.org/10.1371/journal.pone.0169748>
- 1052 Hengl, T., Heuvelink, G. B. M., Kempen, B., Leenaars, J. G. B., Walsh, M. G., Shepherd, K.
1053 D., Sila, A., MacMillan, R. A., De Jesus, J. M., Tamene, L., & Tondoh, J. E. (2015).
1054 Mapping soil properties of Africa at 250 m resolution: Random forests significantly
1055 improve current predictions. *PLoS ONE*, *10*(6), e0125814.
1056 <https://doi.org/10.1371/journal.pone.0125814>
- 1057 Hooper, D. U., Bignell, D. E., Brown, V. K., Brussaard, L., Dangerfield, J. M., Wall, D. H.,
1058 Wardle, D. A., Coleman, D. C., Giller, K. E., Lavelle, P., Van Der Putten, W. H., De Ruiter,
1059 P. C., Rusek, J., Silver, W. L., Tiedje, J. M., & Wolters, V. (2000). Interactions between
1060 aboveground and belowground biodiversity in terrestrial ecosystems: Patterns,
1061 mechanisms, and feedbacks. *BioScience*, *50*(12), 1049–1061.
1062 [https://doi.org/10.1641/0006-3568\(2000\)050\[1049:IBAABB\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2000)050[1049:IBAABB]2.0.CO;2)
- 1063 Horner-Devine, M. C., Lage, M., Hughes, J. B., & Bohannon, B. J. M. (2004). A taxa-area
1064 relationship for bacteria. *Nature*, *432*(7018), 750–753.
1065 <https://doi.org/10.1038/nature03073>
- 1066 Ito, A., & Wagai, R. (2017). Global distribution of clay-size minerals on land surface for
1067 biogeochemical and climatological studies. *Scientific Data*, *4*(1), 170103.
1068 <https://doi.org/10.1038/sdata.2017.103>
- 1069 Jenny, H. (1945). Factors of Soil Formation: A System of Quantitative Pedology. *Geographical*
1070 *Review*, *35*(2), 336. <https://doi.org/10.2307/211491>

- 1071 Kang, P., Pan, Y., Ran, Y., Li, W., Shao, M., Zhang, Y., Ji, Q., & Ding, X. (2023). Soil
1072 saprophytic fungi could be used as an important ecological indicator for land management
1073 in desert steppe. *Ecological Indicators*, *150*, 110224.
1074 <https://doi.org/10.1016/j.ecolind.2023.110224>
- 1075 Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann,
1076 N. E., Linder, H. P., & Kessler, M. (2017). Climatologies at high resolution for the earth's
1077 land surface areas. *Scientific Data*, *4*(1), 170122. <https://doi.org/10.1038/sdata.2017.122>
- 1078 Knight, C. G., Nicolitch, O., Griffiths, R. I., Goodall, T., Jones, B., Weser, C., Langridge, H.,
1079 Davison, J., Dellavalle, A., Eisenhauer, N., Gongalsky, K. B., Hector, A., Jardine, E.,
1080 Kardol, P., Maestre, F. T., Schädler, M., Semchenko, M., Stevens, C., Tsiafouli, M., ... de
1081 Vries, F. T. (2024). Soil microbiomes show consistent and predictable responses to extreme
1082 events. *Nature*, *636*(8043), 690–696. <https://doi.org/10.1038/s41586-024-08185-3>
- 1083 Kuhn, M. (2022). *caret: Classification and Regression Training* (6.0-91).
1084 <https://doi.org/10.1887/0750303123/b365c43>
- 1085 Kuzyakov, Y., & Blagodatskaya, E. (2015). Microbial hotspots and hot moments in soil:
1086 Concept & review. *Soil Biology and Biochemistry*, *83*, 184–199.
1087 <https://doi.org/10.1016/j.soilbio.2015.01.025>
- 1088 Labouyrie, M., Ballabio, C., Romero, F., Panagos, P., Jones, A., Schmid, M. W., Mikryukov, V.,
1089 Dulya, O., Tedersoo, L., Bahram, M., Lugato, E., van der Heijden, M. G. A., & Orgiazzi,
1090 A. (2023). Patterns in soil microbial diversity across Europe. *Nature Communications*,
1091 *14*(1), 1–21. <https://doi.org/10.1038/s41467-023-37937-4>
- 1092 Larson, M. S. (1991). Health-Related Messages Embedded in Prime-Time Television
1093 Entertainment. *Health Communication*, *3*(3), 175–184.
1094 https://doi.org/10.1207/s15327027hc0303_3

- 1095 Lehmann, A., Zheng, W., Ryo, M., Soutschek, K., Roy, J., Rongstock, R., Maaß, S., & Rillig,
1096 M. C. (2020). Fungal Traits Important for Soil Aggregation. *Frontiers in Microbiology*,
1097 10, 495825. <https://doi.org/10.3389/fmicb.2019.02904>
- 1098 Lehmann, J., Bossio, D. A., Kögel-Knabner, I., & Rillig, M. C. (2020). The concept and future
1099 prospects of soil health. *Nature Reviews Earth and Environment*, 1(10), 544–553.
1100 <https://doi.org/10.1038/s43017-020-0080-8>
- 1101 Lehmann, J., & Kleber, M. (2015). The contentious nature of soil organic matter. *Nature*,
1102 528(7580), 60–68. <https://doi.org/10.1038/nature16069>
- 1103 Li, J., Delgado-Baquerizo, M., Wang, J. T., Hu, H. W., Cai, Z. J., Zhu, Y. N., & Singh, B. K.
1104 (2019). Fungal richness contributes to multifunctionality in boreal forest soil. *Soil Biology*
1105 *and Biochemistry*, 136, 107526. <https://doi.org/10.1016/j.soilbio.2019.107526>
- 1106 Li, Z., Wang, Z., Zhang, W., Zhu, J., Chen, B., Jiang, L., Xu, D., Li, W., Liu, J., & He, Z. (2024).
1107 Soil Environments Regulate Dominant Soil Fungal Communities along an Elevational
1108 Gradient in Subtropical Forests. *Forests*, 15(4), 643. <https://doi.org/10.3390/f15040643>
- 1109 Liang, C., Amelung, W., Lehmann, J., & Kästner, M. (2019). Quantitative assessment of
1110 microbial necromass contribution to soil organic matter. *Global Change Biology*, 25(11),
1111 3578–3590. <https://doi.org/10.1111/gcb.14781>
- 1112 Locey, K. J., & Lennon, J. T. (2016). Scaling laws predict global microbial diversity.
1113 *Proceedings of the National Academy of Sciences of the United States of America*, 113(21),
1114 5970–5975. <https://doi.org/10.1073/pnas.1521291113>
- 1115 Ma, X., Xu, X., Geng, Q., Luo, Y., Ju, C., Li, Q., & Zhou, Y. (2023). Global arbuscular
1116 mycorrhizal fungal diversity and abundance decreases with soil available phosphorus.
1117 *Global Ecology and Biogeography*, 32(8), 1423–1434. <https://doi.org/10.1111/geb.13704>
- 1118 Maestre, F. T., Quero, J. L., Gotelli, N. J., Escudero, A., Ochoa, V., Delgado-Baquerizo, M.,
1119 García-Gómez, M., Bowker, M. A., Soliveres, S., Escolar, C., García-Palacios, P.,

- 1120 Berdugo, M., Valencia, E., Gozalo, B., Gallardo, A., Aguilera, L., Arredondo, T., Blones,
1121 J., Boeken, B., ... Zaady, E. (2012). Plant species richness and ecosystem
1122 multifunctionality in global drylands. *Science*, 335(6065), 214–218.
1123 <https://doi.org/10.1126/science.1215442>
- 1124 Malhi, Y., Farfán Amézquita, F., Doughty, C. E., Silva-Espejo, J. E., Girardin, C. A. J., Metcalfe,
1125 D. B., Aragão, L. E. O. C., Huaraca-Quispe, L. P., Alzamora-Taype, I., Eguiluz-Mora, L.,
1126 Marthews, T. R., Halladay, K., Quesada, C. A., Robertson, A. L., Fisher, J. B., Zaragoza-
1127 Castells, J., Rojas-Villagra, C. M., Pelaez-Tapia, Y., Salinas, N., ... Phillips, O. L. (2014).
1128 The productivity, metabolism and carbon cycle of two lowland tropical forest plots in
1129 south-western Amazonia, Peru. *Plant Ecology and Diversity*, 7(1–2), 85–105.
1130 <https://doi.org/10.1080/17550874.2013.820805>
- 1131 McBratney, A. B., Mendonça Santos, M. L., & Minasny, B. (2003). On digital soil mapping.
1132 *Geoderma*, 117(1–2), 3–52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)
- 1133 McBratney, A., Field, D. J., & Koch, A. (2014). The dimensions of soil security. *Geoderma*,
1134 213, 203–213. <https://doi.org/10.1016/j.geoderma.2013.08.013>
- 1135 Meinshausen, N. (2017). *quantregForest: Quantile Regression Forests* (1.3-7). [https://cran.r-](https://cran.r-project.org/package=quantregForest)
1136 [project.org/package=quantregForest](https://cran.r-project.org/package=quantregForest)
- 1137 Melton, J. R., Chan, E., Millard, K., Fortier, M., Winton, R. S., Martín-López, J. M., Cadillo-
1138 Quiroz, H., Kidd, D., & Verchot, L. V. (2022). A map of global peatland extent created
1139 using machine learning (Peat-ML). *Geoscientific Model Development*, 15(12), 4709–
1140 4738. <https://doi.org/10.5194/gmd-15-4709-2022>
- 1141 Meyer, H., Milà, C., Ludwig, M., Linnenbrink, J., & Schumacher, F. (2025). *CAST: “caret”*
1142 *Applications for Spatial-Temporal Models* (1.0.3).

- 1143 Meyer, H., & Pebesma, E. (2021). Predicting into unknown space? Estimating the area of
1144 applicability of spatial prediction models. *Methods in Ecology and Evolution*, 12(9), 1620–
1145 1633. <https://doi.org/10.1111/2041-210X.13650>
- 1146 Minasny, B., & McBratney, A. B. (2006). A conditioned Latin hypercube method for sampling
1147 in the presence of ancillary information. *Computers and Geosciences*, 32(9), 1378–1388.
1148 <https://doi.org/10.1016/j.cageo.2005.12.009>
- 1149 Moquedace, C. M., Baldi, C. G. O., Siqueira, R. G., Cardoso, I. M., Souza, E. F. M. de, Fontes,
1150 R. L. F., Francelino, M. R., Gomes, L. C., & Fernandes-Filho, E. I. (2024). High-resolution
1151 mapping of soil carbon stocks in the western Amazon. *Geoderma Regional*, 36, e00773.
1152 <https://doi.org/10.1016/j.geodrs.2024.e00773>
- 1153 Muñoz Sabater, J. (2024). ERA5-Land hourly data from 1950 to present. Em *Copernicus*
1154 *Climate Change Service (C3S) Climate Data Store (CDS)* (p. 1–10). ECMWF.
1155 <https://doi.org/10.24381/cds.e2161bac>
- 1156 Myneni, R., Knyazikhin, Y., & Park, T. (2021). MOD15A2H MODIS/Terra Leaf Area
1157 Index/FPAR 8-Day L4 Global 500m SIN Grid V061. Em *NASA EOSDIS Land Processes*
1158 *DAAC*. <https://doi.org/10.5067/MODIS/MCD15A3H.061>
- 1159 Noy-Meir, I. (1973). Desert Ecosystems: Environment and Producers. *Annual Review of*
1160 *Ecology and Systematics*, 4(1), 25–51.
1161 <https://doi.org/10.1146/annurev.es.04.110173.000325>
- 1162 Olden, J. D., Lawler, J. J., & Poff, N. L. (2008). Machine learning methods without tears: A
1163 primer for ecologists. *Quarterly Review of Biology*, 83(2), 171–193.
1164 <https://doi.org/10.1086/587826>
- 1165 Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N.,
1166 Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J.,
1167 Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., &

- 1168 Kassem, K. R. (2001a). Terrestrial Ecoregions of the World: A New Map of Life on Earth:
1169 A new global map of terrestrial ecoregions provides an innovative tool for conserving
1170 biodiversity. *BioScience*, *51*(11), 933–938. [https://doi.org/10.1641/0006-](https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2)
1171 [3568\(2001\)051\[0933:TEOTWA\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2)
- 1172 Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N.,
1173 Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J.,
1174 Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., &
1175 Kassem, K. R. (2001b). Terrestrial Ecoregions of the World: A New Map of Life on Earth:
1176 A new global map of terrestrial ecoregions provides an innovative tool for conserving
1177 biodiversity. *BioScience*, *51*(11), 933–938. [https://doi.org/10.1641/0006-](https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2)
1178 [3568\(2001\)051\[0933:TEOTWA\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2)
- 1179 Poggio, L., De Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., &
1180 Rossiter, D. (2021). SoilGrids 2.0: Producing soil information for the globe with quantified
1181 spatial uncertainty. *Soil*, *7*(1), 217–240. <https://doi.org/10.5194/soil-7-217-2021>
- 1182 Prober, S. M., Leff, J. W., Bates, S. T., Borer, E. T., Firn, J., Harpole, W. S., Lind, E. M.,
1183 Seabloom, E. W., Adler, P. B., Bakker, J. D., Cleland, E. E., Decrappeo, N. M., Delorenze,
1184 E., Hagenah, N., Hautier, Y., Hofmockel, K. S., Kirkman, K. P., Knops, J. M. H., La Pierre,
1185 K. J., ... Fierer, N. (2015). Plant diversity predicts beta but not alpha diversity of soil
1186 microbes across grasslands worldwide. *Ecology Letters*, *18*(1), 85–95.
1187 <https://doi.org/10.1111/ele.12381>
- 1188 R Core Team. (2025). R: A Language and Environment for Statistical Computing. Em *R*
1189 *Foundation for Statistical Computing* (4.4.3). <http://www.r-project.org>
- 1190 Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvahalais, N., & Prabhat.
1191 (2019). Deep learning and process understanding for data-driven Earth system science.
1192 *Nature*, *566*(7743), 195–204. <https://doi.org/10.1038/s41586-019-0912-1>

- 1193 Rillig, M. C., & Mummey, D. L. (2006). Mycorrhizas and soil structure. *New Phytologist*,
1194 171(1), 41–53. <https://doi.org/10.1111/J.1469-8137.2006.01750.X>
- 1195 Rola, K., Rożek, K., Chowaniec, K., Błaszowski, J., Gielas, I., Stanek, M., Wietrzyk-Pełka,
1196 P., Węgrzyn, M., Fałowska, P., Dziurawicz, P., Nicia, P., Bejger, R., Zadrozny, P., Pliszko,
1197 A., Zalewska-Gałosz, J., & Zubek, S. (2023). Vascular plant and cryptogam abundance as
1198 well as soil chemical properties shape microbial communities in the successional gradient
1199 of glacier foreland soils. *Science of the Total Environment*, 860, 160550.
1200 <https://doi.org/10.1016/j.scitotenv.2022.160550>
- 1201 Rosa, L. H., Gonçalves, V. N., Barreto, D. L. C., Francelino, M. R., Baldi, C. G. O., Mello, D.
1202 C., Santos, K. C. R., Lopes, F. A. C., Carvalho-Silva, M., Convey, P., & Câmara, P. E. A.
1203 S. (2025). Carbon Dioxide Fluxes Associated with Prokaryotic and Eukaryotic
1204 Communities in Ice-Free Areas on King George Island, Maritime Antarctica. *DNA*, 5(1),
1205 15. <https://doi.org/10.3390/dna5010015>
- 1206 Rosa, L. H., Vaz, A. B. M., Caligiorne, R. B., Campolina, S., & Rosa, C. A. (2009). Endophytic
1207 fungi associated with the Antarctic grass *Deschampsia antarctica* Desv. (Poaceae). *Polar*
1208 *Biology*, 32(2), 161–167. <https://doi.org/10.1007/s00300-008-0515-z>
- 1209 Rosenzweig, M. L. (1995). Species Diversity in Space and Time. *Species Diversity in Space*
1210 *and Time*. <https://doi.org/10.1017/cbo9780511623387>
- 1211 Running, S., Mu, Q., & Zhao, M. (2021). MODIS/Terra Gross Primary Productivity 8-Day L4
1212 Global 500m SIN Grid V061, NASA EOSDIS Land Processes DAAC [data set]. Em *NASA*
1213 *EOSDIS Land Processes Distributed Active Archive Center*.
1214 <https://doi.org/10.5067/MODIS/MOD17A2H.061>
- 1215 Schaap, K. J., Fuchslueger, L., Hoosbeek, M. R., Hofhansl, F., Martins, N. P., Valverde-
1216 Barrantes, O. J., Hartley, I. P., Lugli, L. F., & Quesada, C. A. (2021). Litter inputs and
1217 phosphatase activity affect the temporal variability of organic phosphorus in a tropical

- 1218 forest soil in the Central Amazon. *Plant and Soil*, 469(1–2), 423–441.
1219 <https://doi.org/10.1007/s11104-021-05146-x>
- 1220 Serna-Chavez, H. M., Fierer, N., & Van Bodegom, P. M. (2013). Global drivers and patterns of
1221 microbial abundance in soil. *Global Ecology and Biogeography*, 22(10), 1162–1172.
1222 <https://doi.org/10.1111/geb.12070>
- 1223 Siqueira, R. G., Moquedace, C. M., Fernandes-Filho, E. I., Schaefer, C. E. G. R., Francelino,
1224 M. R., Sacramento, I. F., & Michel, R. F. M. (2024). Modelling and prediction of major
1225 soil chemical properties with Random Forest: Machine learning as tool to understand soil-
1226 environment relationships in Antarctica. *Catena*, 235, 107677.
1227 <https://doi.org/10.1016/j.catena.2023.107677>
- 1228 Siqueira, R. G., Moquedace, C. M., Francelino, M. R., Schaefer, C. E. G. R., & Fernandes-
1229 Filho, E. I. (2023). Machine learning applied for Antarctic soil mapping: Spatial prediction
1230 of soil texture for Maritime Antarctica and Northern Antarctic Peninsula. *Geoderma*, 432,
1231 116405. <https://doi.org/10.1016/j.geoderma.2023.116405>
- 1232 Slik, J. W. F., Arroyo-Rodríguez, V., Aiba, S. I., Alvarez-Loayza, P., Alves, L. F., Ashton, P.,
1233 Balvanera, P., Bastian, M. L., Bellingham, P. J., Van Den Berg, E., Bernacci, L., Da
1234 Conceição Bispo, P., Blanc, L., Böhning-Gaese, K., Boeckx, P., Bongers, F., Boyle, B.,
1235 Bradford, M., Brearley, F. Q., ... Venticinque, E. M. (2015). An estimate of the number of
1236 tropical tree species. *Proceedings of the National Academy of Sciences of the United States*
1237 *of America*, 112(24), 7472–7477. <https://doi.org/10.1073/pnas.1423147112>
- 1238 Sousa, L. M., Poggio, L., & Kempen, B. (2019). Comparison of FOSS4G Supported Equal-
1239 Area Projections Using Discrete Distortion Indicatrices. *ISPRS International Journal of*
1240 *Geo-Information*, 8(8), 351. <https://doi.org/10.3390/ijgi8080351>

- 1241 Steve Running, & Qiaozhen Mu. (2015). *MOD16A2 - MODIS/Terra Net Evapotranspiration*
1242 *8-Day L4 Global 500m SIN Grid*. NASA EOSDIS Land Processes Distributed Active
1243 Archive Center. <https://doi.org/10.5067/MODIS/MOD16A2GF.061>
- 1244 Tedersoo, L., Bahram, M., Pöhlme, S., Kõljalg, U., Yorou, N. S., Wijesundera, R., Ruiz, L. V.,
1245 Vasco-Palacios, A. M., Thu, P. Q., Suija, A., Smith, M. E., Sharp, C., Saluveer, E., Saitta,
1246 A., Rosas, M., Riit, T., Ratkowsky, D., Pritsch, K., Põldmaa, K., ... Abarenkov, K. (2014).
1247 Global diversity and geography of soil fungi. *Science*, *346*(6213).
1248 https://doi.org/10.1126/SCIENCE.1256688/SUPPL_FILE/TEDERSOO-SM.PDF
- 1249 Tedersoo, L., Mikryukov, V., Anslan, S., Bahram, M., Khalid, A. N., Corrales, A., Agan, A.,
1250 Vasco-Palacios, A. M., Saitta, A., Antonelli, A., Rinaldi, A. C., Verbeken, A., Sulisty, B.
1251 P., Tamgnoue, B., Furneaux, B., Ritter, C. D., Nyamukondiwa, C., Sharp, C., Marín, C.,
1252 ... Abarenkov, K. (2021). The Global Soil Mycobiome consortium dataset for boosting
1253 fungal diversity research. *Fungal Diversity*, *111*(1), 573–588.
1254 <https://doi.org/10.1007/s13225-021-00493-7>
- 1255 The European Space Agency. (2022). Copernicus Digital Elevation Model Handbook. Em
1256 *GSICS Quarterly* (Número 1, p. 36 p). <https://doi.org/10.5270/ESA-c5d3d65>
- 1257 Tolan, J., Yang, H. I., Nosarzewski, B., Couairon, G., Vo, H. V., Brandt, J., Spore, J., Majumdar,
1258 S., Haziza, D., Vamaraju, J., Moutakanni, T., Bojanowski, P., Johns, T., White, B., Tiecke,
1259 T., & Couprie, C. (2024). Very high resolution canopy height maps from RGB imagery
1260 using self-supervised vision transformer and convolutional decoder trained on aerial lidar.
1261 *Remote Sensing of Environment*, *300*, 113888. <https://doi.org/10.1016/j.rse.2023.113888>
- 1262 Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein,
1263 M., Arain, M. A., Cescatti, A., Kiely, G., Merbold, L., Serrano-Ortiz, P., Sickert, S., Wolf,
1264 S., & Papale, D. (2016). Predicting carbon dioxide and energy fluxes across global

- 1265 FLUXNET sites with regression algorithms. *Biogeosciences*, 13(14), 4291–4313.
1266 <https://doi.org/10.5194/bg-13-4291-2016>
- 1267 Urbanová, M., Šnajdr, J., & Baldrian, P. (2015). Composition of fungal and bacterial
1268 communities in forest litter and soil is largely determined by dominant trees. *Soil Biology
1269 and Biochemistry*, 84, 53–64. <https://doi.org/10.1016/j.soilbio.2015.02.011>
- 1270 van den Hoogen, J., Geisen, S., Routh, D., Ferris, H., Traunspurger, W., Wardle, D. A., de
1271 Goede, R. G. M., Adams, B. J., Ahmad, W., Andriuzzi, W. S., Bardgett, R. D., Bonkowski,
1272 M., Campos-Herrera, R., Cares, J. E., Caruso, T., de Brito Caixeta, L., Chen, X., Costa, S.
1273 R., Creamer, R., ... Crowther, T. W. (2019). Soil nematode abundance and functional
1274 group composition at a global scale. *Nature*, 572(7768), 194–198.
1275 <https://doi.org/10.1038/s41586-019-1418-6>
- 1276 Vermote, E. (2021). MOD09GA, MODIS/Terra Surface Reflectance Daily L2G Global 1km
1277 and 500m Dataset. Em *Nasa Eosdis Lp Daac*.
1278 <https://doi.org/10.5067/MODIS/MOD09GA.061>
- 1279 Wagg, C., Bender, S. F., Widmer, F., & Van Der Heijden, M. G. A. (2014). Soil biodiversity and
1280 soil community composition determine ecosystem multifunctionality. *Proceedings of the
1281 National Academy of Sciences of the United States of America*, 111(14), 5266–5270.
1282 <https://doi.org/10.1073/pnas.1320054111>
- 1283 Warkentin, B. P. (1995). The changing concept of soil quality. *Journal of Soil and Water
1284 Conservation*, 50(3), 226–228. <https://www.jsowonline.org/content/50/3/226>
- 1285 Warner, D. L., Bond-Lamberty, B., Jian, J., Stell, E., & Vargas, R. (2019). Spatial Predictions
1286 and Associated Uncertainty of Annual Soil Respiration at the Global Scale. *Global
1287 Biogeochemical Cycles*, 33(12), 1733–1745. <https://doi.org/10.1029/2019GB006264>
- 1288 Yaffar, D., Lugli, L. F., Wong, M. Y., Norby, R. J., Addo-Danso, S. D., Arnaud, M., Cordeiro,
1289 A. L., Dietterich, L. H., Diaz-Toribio, M. H., Lee, M. Y., Ghimire, O. P., Smith-Martin, C.

- 1290 M., Toro, L., Andersen, K., McCulloch, L. A., Meier, I. C., Powers, J. S., Sanchez-Julia,
1291 M., Soper, F. M., & Cusack, D. F. (2024). Tropical root responses to global changes: A
1292 synthesis. *Global Change Biology*, *30*(7), e17420. <https://doi.org/10.1111/gcb.17420>
- 1293 Yamauchi, D. H., Garces, H. G., Teixeira, M. D. M., Rodrigues, G. F. B., Ullmann, L. S., Garces,
1294 A. G., Hebelers-Barbosa, F., & Bagagli, E. (2021). Soil mycobiome is shaped by vegetation
1295 and microhabitats: A regional-scale study in Southeastern Brazil. *Journal of Fungi*, *7*(8),
1296 587. <https://doi.org/10.3390/jof7080587>
- 1297 Yang, B., Yang, Z., He, K., Zhou, W., & Feng, W. (2024). Soil Fungal Community Diversity,
1298 Co-Occurrence Networks, and Assembly Processes under Diverse Forest Ecosystems.
1299 *Microorganisms*, *12*(9), 1915. <https://doi.org/10.3390/microorganisms12091915>
- 1300 Zeng, Q., Liu, Y., Xiao, L., & An, S. (2020). Climate and soil properties regulate soil fungal
1301 communities on the Loess Plateau. *Pedobiologia*, *81–82*, 150668.
1302 <https://doi.org/10.1016/j.pedobi.2020.150668>
- 1303 Zhang, Y., Schaap, M. G., & Zha, Y. (2018). A High-Resolution Global Map of Soil Hydraulic
1304 Properties Produced by a Hierarchical Parameterization of a Physically Based Water
1305 Retention Model. *Water Resources Research*, *54*(12), 9774–9790.
1306 <https://doi.org/10.1029/2018WR023539>

1307 **Artigo 2: Mapeamento global de traços funcionais de raízes: insights sobre dinâmicas**
1308 **edáficas e climáticas**

1309 1. Introdução

1310 A ecologia tem avançado na compreensão das diversas características dos ecossistemas,
1311 com o estudo dos traços funcionais emergindo como uma abordagem promissora para entender
1312 como esses sistemas respondem a pressões ambientais (Lee & Andersen, 2023). Em particular,
1313 a investigação dos traços funcionais das raízes tem revelado estratégias variadas das plantas na
1314 aquisição de recursos, sua história evolutiva e crescimento (Grime et al., 1997; Lavorel et al.,
1315 2007; Violle et al., 2007).

1316 As raízes desempenham um papel fundamental no crescimento das plantas e na
1317 produtividade dos ecossistemas, além de influenciar diretamente várias funções ecológicas.
1318 Elas são essenciais para a absorção de nutrientes e água, bem como para a fixação das plantas
1319 no solo (Brunner & Godbold, 2007). As raízes também desempenham um papel em processos
1320 subterrâneos, como hospedar comunidades de simbiontes micorrízicos e outros microrganismos
1321 na rizosfera, que influenciam a decomposição da serrapilheira e a ciclagem de nutrientes,
1322 essenciais para a produtividade do ecossistema (Gadgil & Gadgil, 1971; Taylor et al., 2016).
1323 Além disso, as características das raízes são indicadores-chave da resposta das plantas às
1324 mudanças ambientais, refletindo seu impacto nos processos do ecossistema (Bardgett & Van
1325 Der Putten, 2014; Freschet, Roumet, et al., 2021).

1326 No entanto, o foco predominante da ecologia funcional tem sido em órgãos acima do
1327 solo, como folhas e caules, enquanto os componentes subterrâneos, como raízes e a rizosfera,
1328 têm sido menos estudados (Bardgett et al., 2014; Cadotte et al., 2009; Díaz et al., 2007; Roscher
1329 et al., 2012). Esse desequilíbrio gera uma lacuna de conhecimento que abrange desde raízes
1330 individuais até ecossistemas inteiros. Isso é atribuído à dificuldade de acesso e observação dos
1331 órgãos vegetais subterrâneos e às complexas interações entre raízes, micorrizas e
1332 microrganismos do solo, que regulam os *feedbacks* planta-solo (Finzi et al., 2015; Fujii et al.,
1333 2018). Além disso, há dificuldades na coleta de dados e na vinculação das características

1334 funcionais das raízes à sua distribuição e dinâmica de crescimento nas paisagens (McCormack
1335 et al., 2015; Pregitzer et al., 2002).

1336 Apesar da importância das raízes, suas funções dinâmicas são frequentemente sub-
1337 representadas em modelos de superfície terrestre que englobam processos e dinâmicas de
1338 fundamental importância para os ecossistemas atuais, como os modelos de ciclo de carbono e
1339 os modelos de resposta da vegetação às mudanças climáticas (Iversen, 2014; Ostle et al., 2009;
1340 Woodward & Osborne, 2000). Sabe-se menos sobre as características das raízes em comparação
1341 com as folhas, especialmente no que diz respeito ao controle sobre a absorção de recursos em
1342 gradientes climáticos e edáficos (Bergmann et al., 2017; Freschet, Roumet, et al., 2021;
1343 Tumber-Dávila et al., 2022).

1344 Dado que as raízes precisam coexistir e adquirir nutrientes e água em diversas
1345 ecorregiões, que variam de florestas tropicais relativamente estáveis a desertos altamente
1346 sazonais ou florestas boreais (Joswig et al., 2022), há uma necessidade urgente de entender
1347 como os traços funcionais das raízes estão distribuídos globalmente. A diversidade na forma e
1348 função das raízes e o grau de associação com fungos micorrízicos simbióticos levantam
1349 questões sobre as adaptações das plantas em diferentes condições ecológicas (Jian et al., 2022;
1350 Öpik et al., 2006).

1351 Desta forma, melhorar nossa compreensão sobre a distribuição espacial dos traços
1352 funcionais das raízes é essencial para os modelos globais de clima, vegetação e carbono do solo,
1353 além de projetar a dinâmica futura dos ecossistemas. A falta de dados detalhados sobre essa
1354 distribuição cria lacunas nos modelos globais, limitando sua precisão e a confiabilidade dos
1355 processos de tomada de decisão baseadas nestes modelos (van den Hoogen et al., 2019).

1356 No entanto, grande parte do conhecimento atual sobre as características das raízes e
1357 comunidades da rizosfera provém de estudos em regiões temperadas do hemisfério norte, que
1358 representam menos de 40% da superfície terrestre vegetada, enquanto os trópicos, zonas áridas

1359 e regiões boreais ocupam a maior parte da área restante (Clark et al., 2013; Guerrero-Ramírez
1360 et al., 2021; Lee & Andersen, 2023; Pan et al., 2011).

1361 Esse desequilíbrio limita severamente nossa compreensão sobre as estratégias
1362 edafoecológicas das plantas em escala global. Em particular, negligencia a diversidade de
1363 respostas adaptativas das raízes frente a condições edáficas extremas, como solos com baixa
1364 concentração de nutrientes, ácidos, sazonalmente secos ou com restrições hídricas, que são
1365 predominantes fora das zonas temperadas (Bardgett & Van Der Putten, 2014b) . Essa lacuna
1366 restringe o entendimento de mecanismos fundamentais relacionados à aquisição de recursos, à
1367 plasticidade fenotípica, às associações micorrízicas e à resiliência funcional das plantas.

1368 Com a intensificação das mudanças climáticas, espera-se que o estresse hídrico e a
1369 variabilidade na disponibilidade de nutrientes se agravem, ampliando a importância de
1370 compreender essas respostas em ambientes sub-representados (Yaffar et al., 2024). Além disso,
1371 a ausência de dados abrangentes sobre raízes compromete a parametrização de modelos
1372 preditivos (Freschet, Roumet, et al., 2021), os quais tendem a superestimar padrões de zonas
1373 bem amostradas e ignorar processos críticos em regiões ecologicamente vulneráveis.

1374 Avanços recentes em tecnologias de *software*, *hardware* e métodos de análise espacial,
1375 como o uso de aprendizado de máquina, têm permitido a criação de mapas mais precisos e a
1376 quantificação da incerteza associada (Mello et al., 2025; Moquedace et al., 2024; Siqueira et
1377 al., 2023, 2024). Iniciativas como o SoilGrids exemplificam como grandes conjuntos de dados
1378 e métodos avançados podem ser integrados para fornecer informações detalhadas sobre os
1379 atributos do solo em escala global (Poggio et al., 2021).

1380 Bancos de dados globais, como os compilados por Guerrero-Ramírez et al. (2021), são
1381 fundamentais para melhorar a modelagem global de variáveis ecológicas. Esses recursos não
1382 só permitem a integração de dados ambientais complexos, como clima e relevo, mas também
1383 facilitam a verificação das relações entre esses fatores e as variáveis-resposta. Em suma, uma

1384 compreensão aprimorada das interações entre os traços funcionais das raízes e os processos
1385 ambientais é essencial para uma gestão mais eficaz dos bens naturais e para uma compreensão
1386 holística da multifuncionalidade do solo. Objetivou-se nesta pesquisa modelar e mapear
1387 diversos traços funcionais de raízes globalmente. Especificamente objetivou-se: i. avaliar a
1388 distribuição espacial dos traços funcionais de raízes nos diferentes ecossistemas do mundo; ii.
1389 investigar as relações edáficas e climáticas com os traços funcionais de raízes.

1390

1391 **2. Material e métodos**

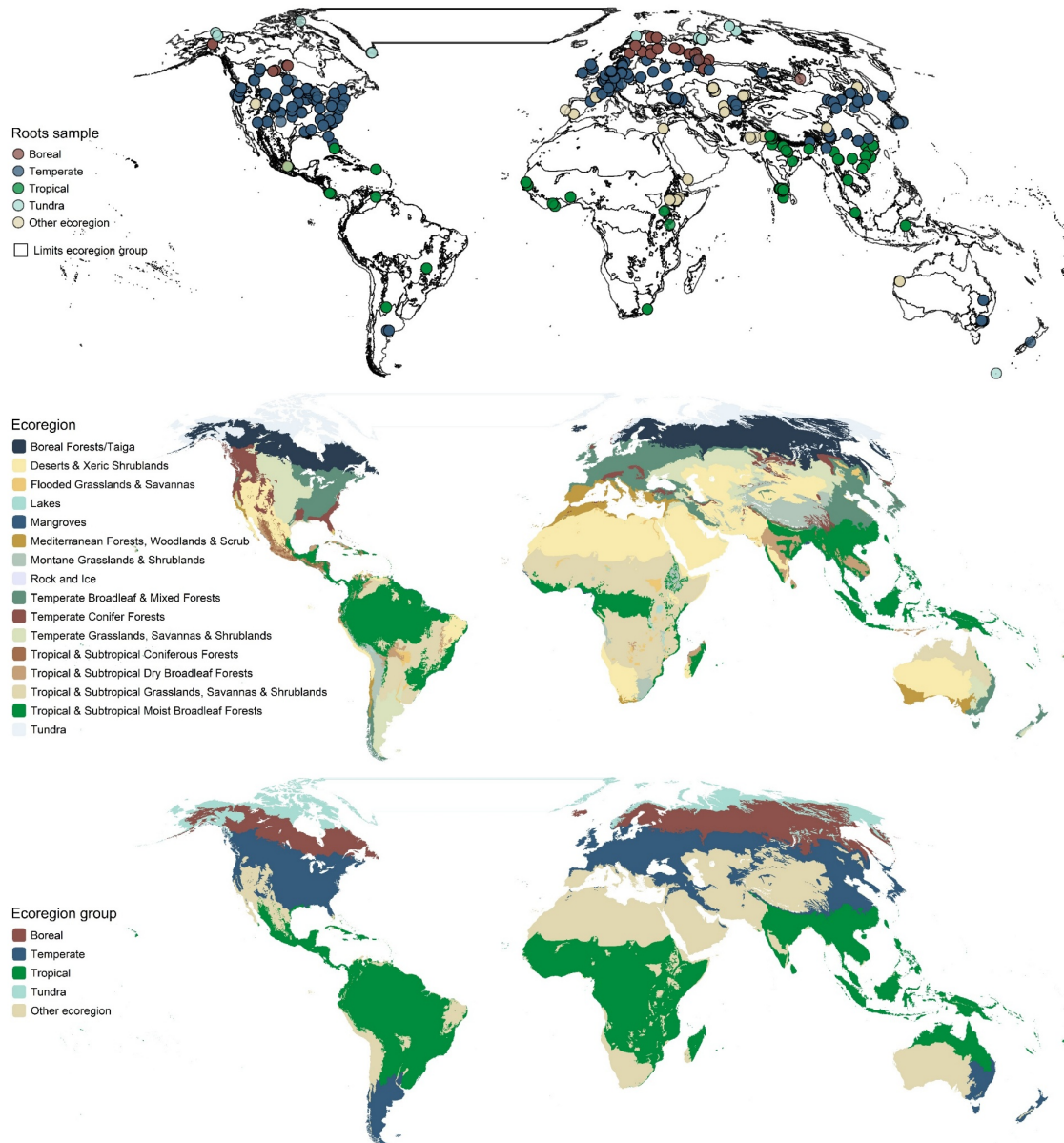
1392 **2.1. Banco de dados**

1393 Para modelar traços radiculares em escala global, utilizou-se o banco de dados Global
1394 Root Traits Database (GRooT), desenvolvido por Guerrero-Ramírez et al. (2021). Esse
1395 repositório integra informações oriundas de múltiplas fontes, incluindo as bases FRED e TRY,
1396 literatura científica e conjuntos inéditos, por meio de um processo sistemático de curadoria,
1397 padronização e controle de qualidade.

1398 A harmonização realizada pelos autores envolveu a uniformização de nomes de traços,
1399 unidades de medida e categorias taxonômicas, com base nas diretrizes do manual de traços
1400 radiculares de Freschet et al. (2021a), permitindo a integração consistente dos dados
1401 independentemente da metodologia original. O GRooT reúne mais de 114 mil registros
1402 padronizados, obtidos entre 1911 e 2019 em estudos de campo e experimentos controlados,
1403 abrangendo mais de 6.200 espécies pertencentes a 254 famílias botânicas e representando uma
1404 ampla diversidade de ecorregiões.

1405 Para o presente estudo, foram considerados exclusivamente os registros obtidos em
1406 condições de campo, os quais passaram por filtros adicionais para assegurar a consistência dos
1407 dados, como a remoção de entradas inconsistentes (*outliers*) ou sem coordenadas geográficas
1408 (Figura 1).

1409



1410

1411

1412

1413

1414

Figura 1. Distribuição das amostras de concentração de carbono nas raízes (RCC), colonização micorrízica das raízes (RMC), respiração específica das raízes (SRR) e taxa de renovação radicular (RTR) em diferentes ecorregiões globais.

1415

1416

1417

1418

1419

A Antártica foi excluída do mapeamento devido à escassez de amostras e à limitada disponibilidade de preditores ambientais adequados. A partir do GRooT, foram selecionados os seguintes traços para análise: intensidade de colonização micorrízica nas raízes, concentração de carbono nas raízes, taxa específica de respiração radicular e taxa de renovação radicular (Tabela 1).

1420

1421 **Tabela 1.** Descrição dos traços radiculares selecionados a partir do banco de dados GRooT
 1422 (Guerrero-Ramírez et al., 2021) para a modelagem e mapeamento.

Acronym	Trait	Unit	Group	Definitions	Samples
RMC	Root mycorrhizal colonization intensity	%	Microbial associations	Percentage length of roots colonized by fungi	6.606
RCC	Root carbon concentration	mg g ⁻¹	Chemistry	Mass of C per root dry mass	5.049
SRR	Specific root respiration	nmol g ⁻¹ s ⁻¹	Physiology & respiration	Amount of CO ₂ released or O ₂ absorbed by root per unit root dry mass per unit of time	1.794
RTR	Root turnover rate	year ⁻¹	Dynamics & decomposition	Root dry mass production per dry mass of a given pool of roots over a period	523

1423

1424 2.2. Preditores ambientais

1425 Foram utilizadas mais de 50 preditores ambientais (Tabela 2) que potencialmente se
 1426 relacionam com os traços funcionais de raízes. Esses preditores são elementos que expressam
 1427 quantitativamente uma extensão dos fatores de formação do solo (McBratney et al., 2003).

1428

1429 **Tabela 2.** Preditores ambientais e suas respectivas referências utilizadas na modelagem da
 1430 intensidade de colonização micorrízica nas raízes, concentração de carbono nas raízes, taxa
 1431 específica de respiração radicular e taxa de renovação radicular.

Sigla	Preditor	Grupo	Ref.
BIO 1	Mean annual temperature	Climate	(Karger et al., 2017)
BIO 10	Mean temperature of the warmest quarter	Climate	(Karger et al., 2017)
BIO 11	Mean temperature of the coldest quarter	Climate	(Karger et al., 2017)
BIO 12	Annual precipitation	Climate	(Karger et al., 2017)
BIO 13	Precipitation of the wettest month	Climate	(Karger et al., 2017)
BIO 14	Precipitation of the driest month	Climate	(Karger et al., 2017)
BIO 15	Precipitation seasonality	Climate	(Karger et al., 2017)
BIO 16	Precipitation of the wettest quarter	Climate	(Karger et al., 2017)
BIO 17	Precipitation of the driest quarter	Climate	(Karger et al., 2017)
BIO 18	Precipitation of the warmest quarter	Climate	(Karger et al., 2017)
BIO 19	Precipitation of the coldest quarter	Climate	(Karger et al., 2017)
BIO 2	Mean diurnal temperature range	Climate	(Karger et al., 2017)
BIO 3	Isothermality	Climate	(Karger et al., 2017)
BIO 4	Temperature seasonality	Climate	(Karger et al., 2017)
BIO 5	Maximum temperature of the warmest month	Climate	(Karger et al., 2017)

Sigla	Preditor	Grupo	Ref.
BIO 6	Minimum temperature of the coldest month	Climate	(Karger et al., 2017)
BIO 7	Annual temperature range	Climate	(Karger et al., 2017)
BIO 8	Mean temperature of the wettest quarter	Climate	(Karger et al., 2017)
BIO 9	Mean temperature of the driest quarter	Climate	(Karger et al., 2017)
CHT	Canopy height	Organisms	(Tolan et al., 2024)
ET	Evapotranspiration	Organisms	(Steve Running & Qiaozhen Mu, 2015)
EVI	Enhanced vegetation index	Organisms	(Vermote, 2021)
FC	Field capacity	Organisms	(Zhang et al., 2018)
FP	Fraction of photosynthetically active radiation	Organisms	(Running et al., 2021)
GPP	Gross primary production	Organisms	(Running et al., 2021)
LAI	Leaf area index	Organisms	(Myneni et al., 2021)
MSAVI	Modified soil-adjusted vegetation index	Organisms	(Vermote, 2021)
NDMI	Normalized difference moisture index	Organisms	(Vermote, 2021)
NDSI	Normalized difference snow index	Organisms	(Vermote, 2021)
NDVI	Normalized difference vegetation index	Organisms	(Vermote, 2021)
NPP	Net primary productivity	Organisms	(Karger et al., 2017)
PAW	Plant available water	Organisms	(Zhang et al., 2018)
SAVI	Soil-adjusted vegetation index	Organisms	(Vermote, 2021)
SMBC	Microbial biomass carbon	Organisms	(Serna-Chavez et al., 2013)
CEC	Ability to exchange cations	Soil / PM	(Poggio et al., 2021)
Clay	Clay content	Soil / PM	(Poggio et al., 2021)
Fe oxides	Iron oxide	Soil / PM	(Ito & Wagai, 2017)
Gibbsite	Gibbsite	Soil / PM	(Ito & Wagai, 2017)
Illite / Mica	Illite/Mica	Soil / PM	(Ito & Wagai, 2017)
Kaolinite	Kaolinite	Soil / PM	(Ito & Wagai, 2017)
N	Nitrogen content	Soil / PM	(Poggio et al., 2021)
PEAT	Peatland extent	Soil / PM	(Melton et al., 2022)
pH	Hydrogen potential in water	Soil / PM	(Poggio et al., 2021)
Sand	Sand content	Soil / PM	(Poggio et al., 2021)
SM	Soil moisture	Soil / PM	(Muñoz Sabater, 2024)
Smectite	Smectite	Soil / PM	(Ito & Wagai, 2017)
Vermiculite	Vermiculite	Soil / PM	(Ito & Wagai, 2017)
WDT	Water table depth	Soil / PM	(Fan et al., 2013)
DEM	Digital elevation model	Topography	(Hawker et al., 2022)

1432

1433 Para representar o relevo, foi utilizado o modelo digital de elevação (FABDEM), com

1434 resolução espacial de 30 metros (Hawker et al., 2022). Para além do relevo, foram incorporadas

1435 variáveis climáticas derivadas do conjunto CHELSA v2.1, incluindo as 19 variáveis
1436 bioclimáticas padronizadas. Essas variáveis sintetizam padrões de temperatura e precipitação
1437 ao longo do tempo, permitindo uma caracterização mais abrangente do ambiente climático do
1438 que o uso isolado de variáveis mensais. Muitas delas descrevem aspectos da sazonalidade. A
1439 vegetação foi representada pela produtividade primária líquida (NPP), também proveniente do
1440 CHELSA v2.1 (Karger et al., 2017).

1441 Complementando os dados climáticos e de vegetação, foram também utilizadas
1442 variáveis oriundas do *Moderate Resolution Imaging Spectroradiometer* (MODIS). Estas
1443 variáveis, que se relacionam potencialmente com os traços funcionais de raízes, foram
1444 utilizadas uma média dos últimos 20 anos de imagens livres de nuvens dos produtos MODIS
1445 (Tabela 2). Além disso, foram incorporados dados de alta resolução das copas das árvores
1446 (Tolan et al., 2024) e uma média histórica de 20 anos da umidade do solo oriunda do
1447 reprocessamento do ERA5 (Muñoz Sabater, 2024).

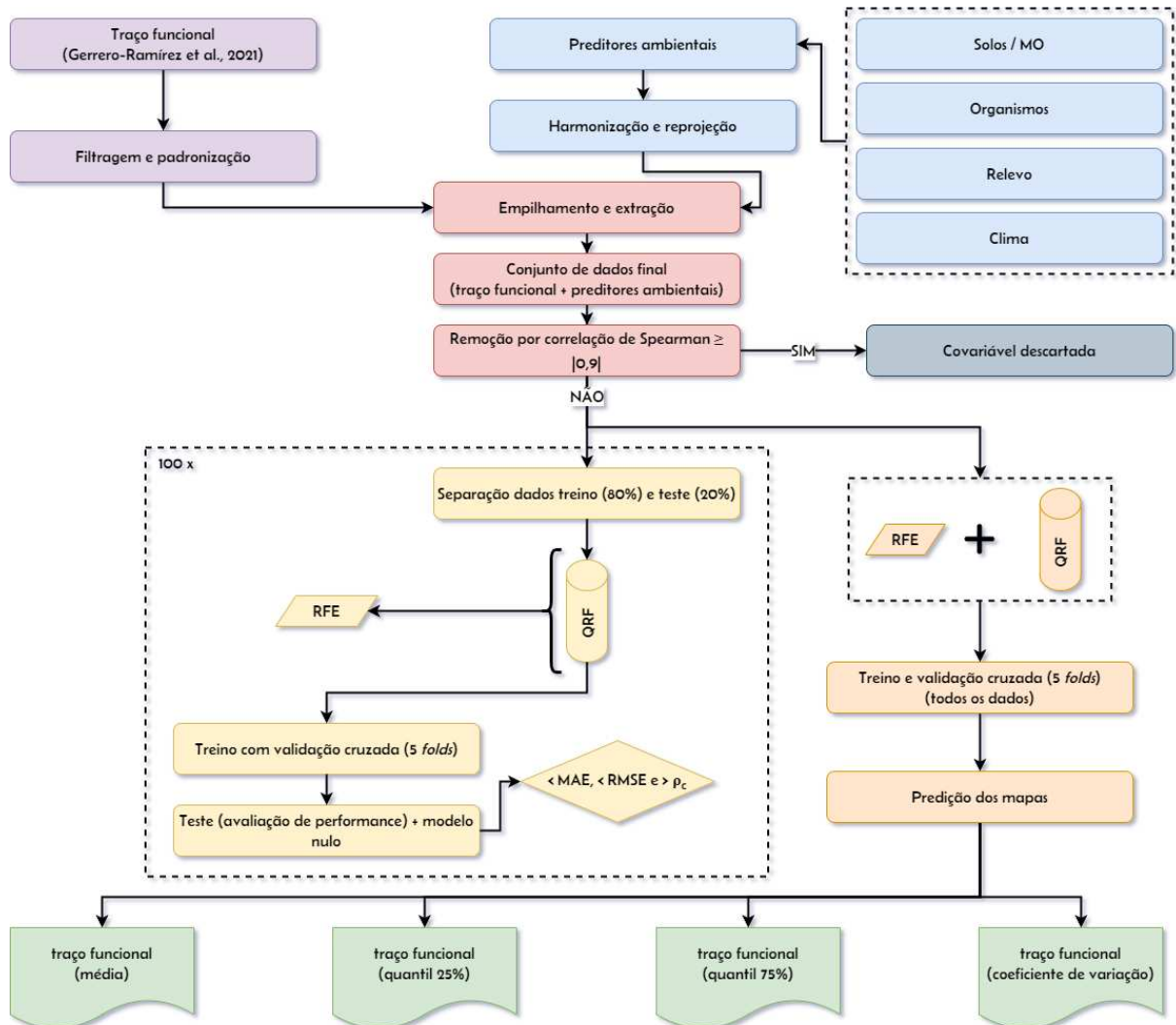
1448 Devido à diversidade de resoluções espaciais das variáveis, todas os preditores
1449 ambientais foram harmonizadas para uma resolução espacial aproximada de 1 km.
1450 Posteriormente, essas variáveis foram reprojctadas para o sistema de projeção Goode's
1451 Homolosine (ESRI:54052), aplicado ao *Datum* WGS84. A escolha dessa projeção baseia-se na
1452 sua efetividade em representar regiões terrestres de forma mais precisa em termos de área
1453 (Sousa et al., 2019).

1454

1455 **2.3. Modelagem geoespacial**

1456 A sequência metodológica para modelagem e mapeamento seguiu etapas sequenciais,
1457 consistindo em: i. seleção de variáveis; ii. ajuste do modelo; iii. avaliação de performance; iv.
1458 quantificação da importância dos preditores ambientais; e v. predição dos mapas médio e de
1459 incerteza (Figura 2).

1460



1461
1462
1463
1464
1465

Figura 2. Sequência metodológica adotada para a modelagem e o mapeamento da concentração de carbono nas raízes (RCC), colonização micorrízica das raízes (RMC), respiração específica das raízes (SRR) e taxa de renovação radicular (RTR) em escala global.

1466

2.3.1. Seleção de preditores

1467

1468

1469

1470

1471

Para construir um modelo de aprendizado de máquina mais parcimonioso e que consuma menos esforço computacional no ajuste, foram adotados dois métodos de seleção de preditores. Esses métodos visaram reduzir o volume dos dados para processamento e aumentar o poder explicativo do modelo. Os dois métodos foram aplicados de forma sequencial: primeiro baseado em correlação e depois em importância.

1472 Para a seleção inicial de preditores, foi construída uma matriz de correlação utilizando
1473 o coeficiente de Spearman. Adotou-se um limiar de correlação de $|0,9|$ para identificar
1474 redundâncias. Quando pares de variáveis apresentaram correlação acima desse limite, foi
1475 removido o preditor com a maior média de correlações absolutas em relação aos demais. Esse
1476 procedimento foi implementado com o pacote caret (Kuhn, 2008).

1477 Após a etapa de seleção por correlação, os preditores remanescentes foram submetidos
1478 ao algoritmo recursive feature elimination (RFE), que seleciona subconjuntos ótimos de
1479 variáveis com base na importância para o modelo. O algoritmo opera de forma iterativa,
1480 avaliando subconjuntos de tamanho variável (de dois até o total disponível) e retendo aquele
1481 que apresentou o melhor desempenho segundo as métricas avaliadas.

1482

1483 **2.3.2. Ajuste do modelo e predição dos mapas**

1484 O algoritmo utilizado para a modelagem foi o quantile random forest (qrf)
1485 (Meinshausen, 2017). Entre os diversos algoritmos de inteligência artificial, incluindo redes
1486 neurais, o random forest se destaca no mapeamento de atributos do solo em escala regional
1487 (Moquedace et al., 2024), nacional (Gomes et al., 2019), continental (Hengl et al., 2015) e
1488 global (Hengl et al., 2017; Poggio et al., 2021). O qrf apresenta a vantagem adicional de permitir
1489 o cálculo dos quantis e do desvio padrão dos valores médios/medianos das árvores de regressão.

1490 Para avaliar a capacidade de generalização do modelo ajustado, os dados foram
1491 particionados em 80% para treinamento (com validação cruzada repetida e 10 *folds*) e 20% para
1492 teste externo de desempenho. Esse processo foi repetido 100 vezes para minimizar a
1493 dependência de uma única divisão dos dados, evitando estimativas enviesadas de performance.

1494 A avaliação do modelo considerou quatro métricas: erro médio absoluto (MAE), raiz quadrada
1495 do erro médio (RMSE), coeficiente de correlação de concordância (ρ_c) e coeficiente de

1496 determinação (R^2). As métricas foram calculadas com base na média dos resultados obtidos ao
1497 longo das 100 execuções do modelo.

1498 No processo de predição espacial, foi ajustado um modelo global utilizando a totalidade
1499 do conjunto de dados amostral, com validação cruzada repetida de 10 *folds*. Com base nesse
1500 modelo, foram gerados mapas para os quatro traços radiculares selecionados: intensidade de
1501 colonização micorrízica nas raízes, concentração de carbono nas raízes, taxa específica de
1502 respiração radicular e taxa de renovação radicular. Para cada traço, foram produzidos dois
1503 produtos cartográficos: um mapa da mediana predita e um mapa do desvio padrão das
1504 predições, representando a incerteza associada às estimativas espaciais.

1505 Adicionalmente, calculou-se o índice de dissimilaridade (DI) para cada traço funcional,
1506 conforme a metodologia proposta por Meyer and Pebesma (2021), resultando em mapas
1507 complementares. O cálculo do DI foi especialmente importante diante da baixa quantidade de
1508 amostras disponíveis em escala global, particularmente na faixa tropical, onde se concentram
1509 os ecossistemas de maior biodiversidade do planeta. O DI expressa a distância mínima de cada
1510 *pixel* de predição em relação aos dados de treinamento no espaço multivariado dos preditores,
1511 ponderado pela importância relativa de cada variável no modelo. Para sua estimativa, os
1512 preditores foram previamente escalonados e ponderados, e a distância euclidiana ao ponto de
1513 treinamento mais próximo foi normalizada pela média das distâncias entre as amostras de
1514 treinamento. Isto foi realizado com auxílio da função “*aoa*” do pacote CAST (Meyer et al.,
1515 2025). O DI é uma medida adimensional, em que valores mais próximos de zero indicam alta
1516 similaridade ambiental em relação ao conjunto amostral, enquanto valores positivos mais
1517 elevados refletem condições ambientais progressivamente distintas e, portanto, maior risco de
1518 extrapolação.

1519 Para avaliar a robustez do qrf, sua performance foi comparada à de um modelo nulo,
1520 definido como a simples média aritmética dos valores observados, ou seja, sem qualquer uso

1521 de preditores. A comparação foi realizada por meio do teste não paramétrico de Kruskal-Wallis,
1522 aplicado sobre as métricas de desempenho (MAE e RMSE), com o objetivo de verificar se as
1523 diferenças observadas entre os modelos foram estatisticamente significativas a 5% de
1524 probabilidade.

1525

1526 **2.3.3. Importância dos preditores**

1527 Foi calculada a importância dos preditores para intensidade de colonização micorrízica
1528 nas raízes, concentração de carbono nas raízes, taxa específica de respiração radicular e taxa de
1529 renovação radicular. Esse cálculo definiu o percentual de contribuição de cada preditor na
1530 construção do modelo, auxiliando na compreensão dos fatores que mais impactaram os
1531 diferentes traços funcionais de raízes a nível global.

1532 Em seguida, foi avaliada a dependência parcial que o modelo estabeleceu entre cada
1533 preditor e os traços funcionais. Isso permitiu a avaliação das faixas de valores de cada preditor
1534 que contribuem para o aumento ou a redução dos diferentes traços funcionais.

1535 Todo o processo de modelagem, predição e elaboração de figuras foi realizado em
1536 ambiente R, com o auxílio dos pacotes caret (Kuhn, 2008), doParallel (Microsoft Corporation
1537 & Weston, 2022), ggplot2 (Wickham, 2016), parallelly (Bengtsson, 2023), quantregForest
1538 (Meinshausen, 2017), sf (Pebesma, 2018), terra (Hijmans, 2023), tmap (Tennekes, 2018) e
1539 mdsFuncs (Fernandes Filho et al., 2021).

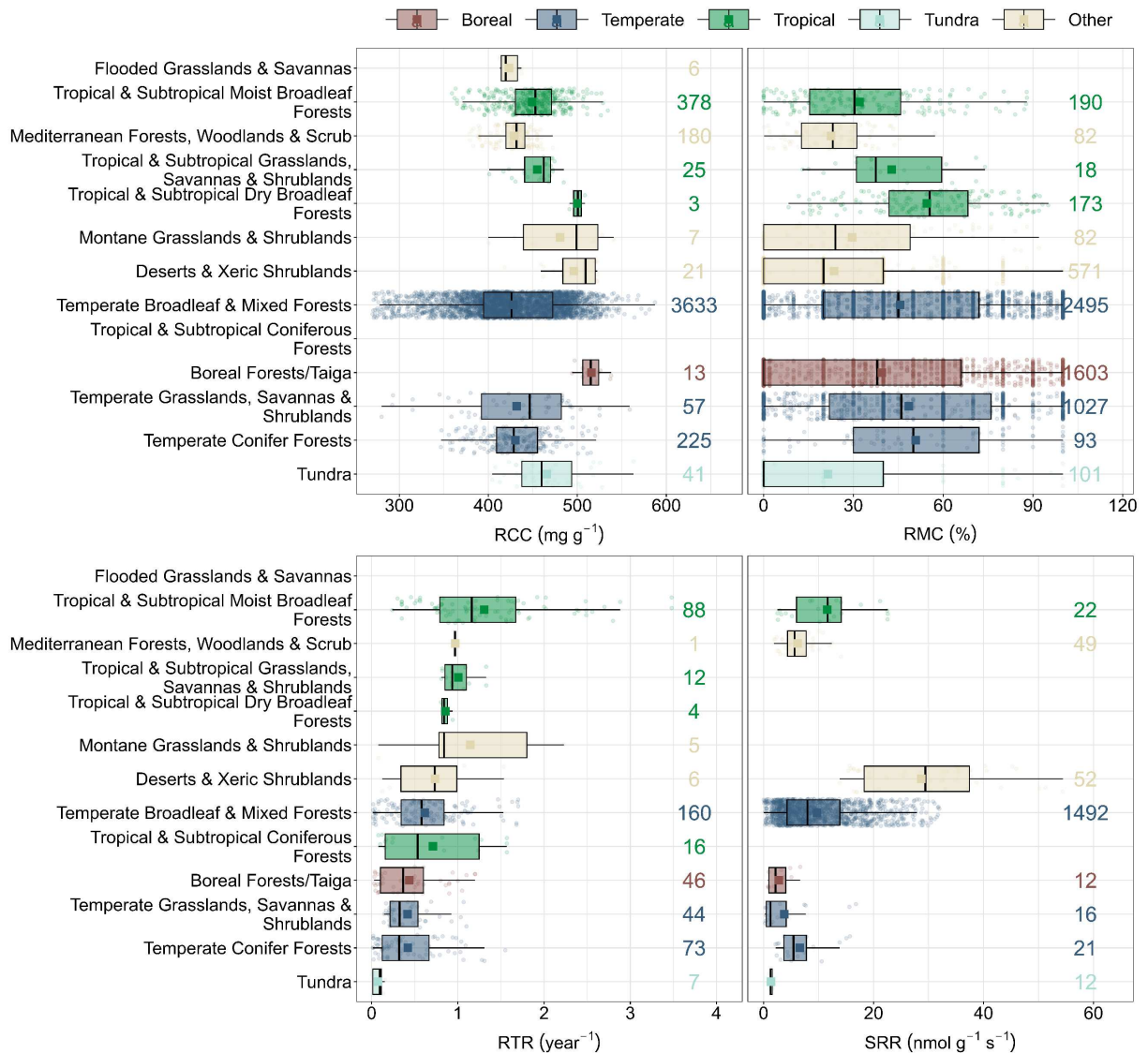
1540

1541 **3. Resultados**

1542 **3.1. Distribuição dos dados e performance do modelo**

1543 Os dados filtrados do GRooT (Guerrero-Ramírez et al., 2021) mostraram uma
1544 expressiva concentração de dados nos ecossistemas temperados para todos os traços funcionais
1545 analisados. Nos ecossistemas com menor representatividade, essa variação ocorre dentro de

1546 cada traço funcional. Por exemplo, para a RCC e a RTR, a menor representatividade está nos
 1547 trópicos, seguidos pelos ecossistemas boreais. Para a RTR, apesar da sub-representação dos
 1548 trópicos, os ecossistemas mais sub-representados são a tundra e a categoria “Other” (Figura 3).
 1549

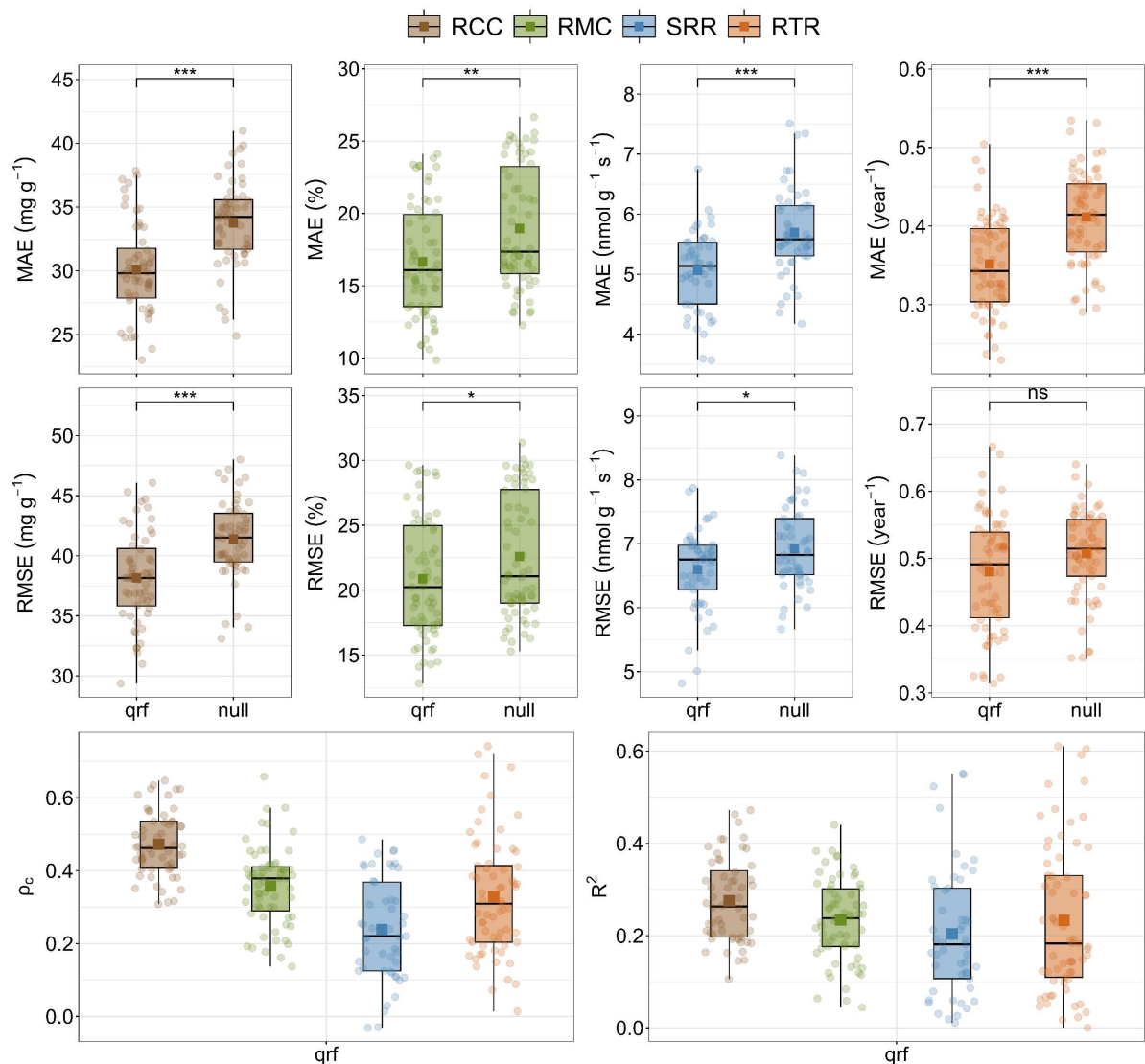


1550
 1551
 1552
 1553
 1554

Figura 3. Distribuição de dados do GRooT (Guerrero-Ramírez et al., 2021) para concentração de carbono nas raízes (RCC), colonização micorrízica das raízes (RMC), respiração específica das raízes (SRR) e taxa de renovação radicular (RTR) em diferentes ecorregiões.

1555 Ao longo das 100 execuções, com exceção do RMSE para o traço RTR, todos os modelos
 1556 ajustados apresentaram erros estatisticamente menores em comparação com o modelo nulo
 1557 (Figura 4), conforme indicado pelo teste de Kruskal-Wallis. As diferenças foram

1558 particularmente acentuadas para o traço RCC (concentração de carbono nas raízes), que
 1559 apresentou desempenho superior nas quatro métricas avaliadas. Em contraste, o traço RTR (taxa
 1560 de renovação radicular) obteve os menores ganhos preditivos em relação ao modelo nulo.
 1561 Considerando o coeficiente de correlação de concordância (ρ_c) e o coeficiente de determinação
 1562 (R^2), a melhor performance mediana foi observada para o RCC, enquanto o RTR apresentou os
 1563 menores valores, refletindo maior incerteza preditiva nesse traço funcional.
 1564



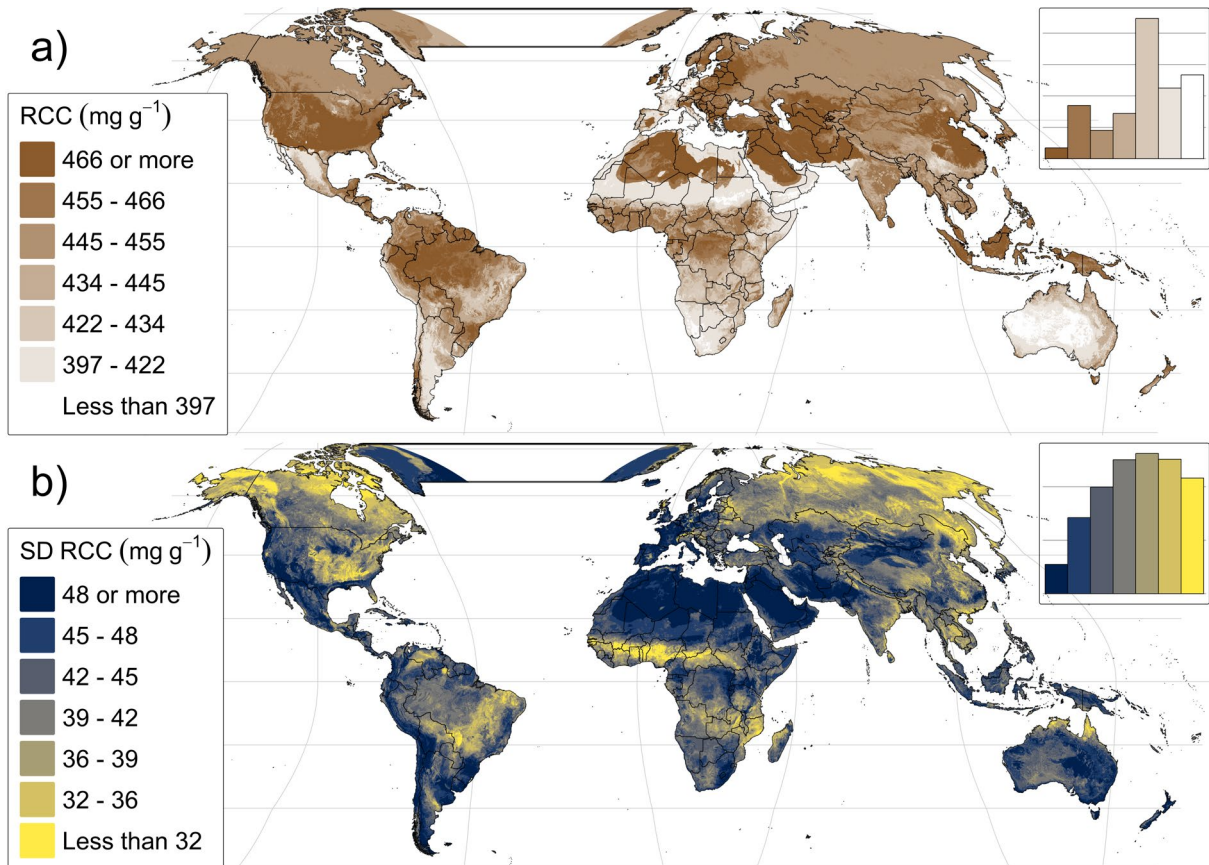
1565
 1566 **Figura 4.** Métricas de desempenho (MAE, RMSE e ρ_c) do modelo quantile random forest para
 1567 os traços radiculares: concentração de carbono nas raízes (RCC), intensidade de colonização
 1568 micorrízica (RMC), respiração específica das raízes (SRR) e taxa de renovação radicular (RTR).
 1569

1570 3.2. Distribuição espacial de diferentes traços funcionais de raízes

1571 O mapa da concentração de carbono nas raízes (RCC) apresentou valores variando de
1572 100 a pouco mais de 500 mg g⁻¹ ao longo do globo (Figura 5). As maiores concentrações foram
1573 observadas em regiões tropicais, com destaque para a Amazônia na América do Sul, o sudeste
1574 dos Estados Unidos e a bacia do Congo no centro do continente africano, todas com valores
1575 superiores a 500 mg g⁻¹. Em contraste, regiões polares e desérticas (como a Groenlândia, o
1576 deserto do Saara e áreas do interior da Austrália) exibiram baixas concentrações desse traço
1577 funcional.

1578 Foi identificada uma predição anômala ao norte do deserto do Saara, com valores
1579 discrepantes em relação ao padrão observado nas regiões vizinhas, sugerindo a presença de um
1580 artefato no processo de predição. As maiores incertezas associadas ao traço RCC concentraram-
1581 se em regiões áridas e desérticas, como o próprio Saara, o noroeste dos Estados Unidos, o
1582 Oriente Médio e partes da Austrália, incluindo a área onde o artefato foi detectado. Nesses
1583 locais, os valores preditos de RCC variaram de baixos a intermediários, sem um padrão espacial
1584 claramente definido. Em contrapartida, as menores incertezas foram observadas no norte do
1585 Canadá, nas regiões sudeste e nordeste do Brasil e no norte da Rússia, indicando maior
1586 confiança nas estimativas do modelo para essas áreas.

1587

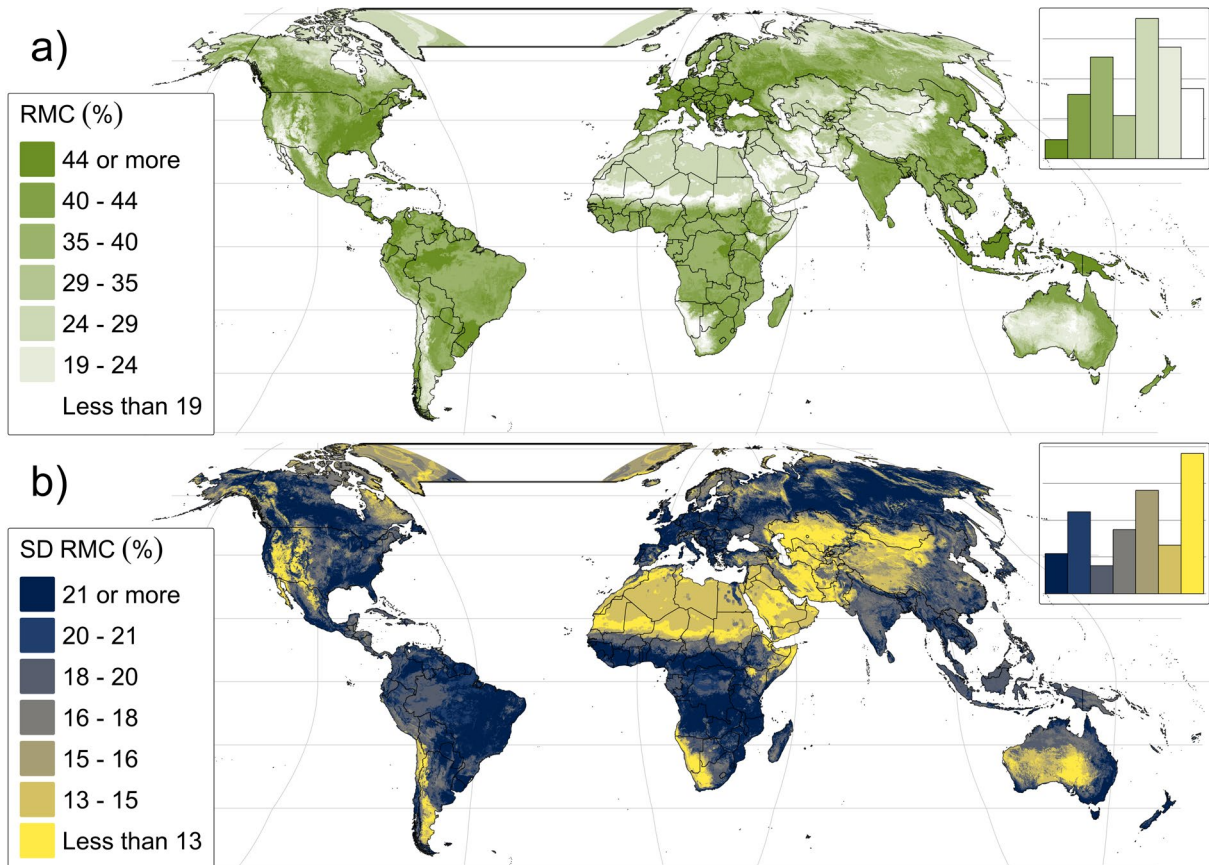


1588
1589
1590
1591

Figura 5. Concentração de carbono nas raízes (RCC) e desvio padrão das previsões (SD RCC) modelados com quantile random forest em escala global.

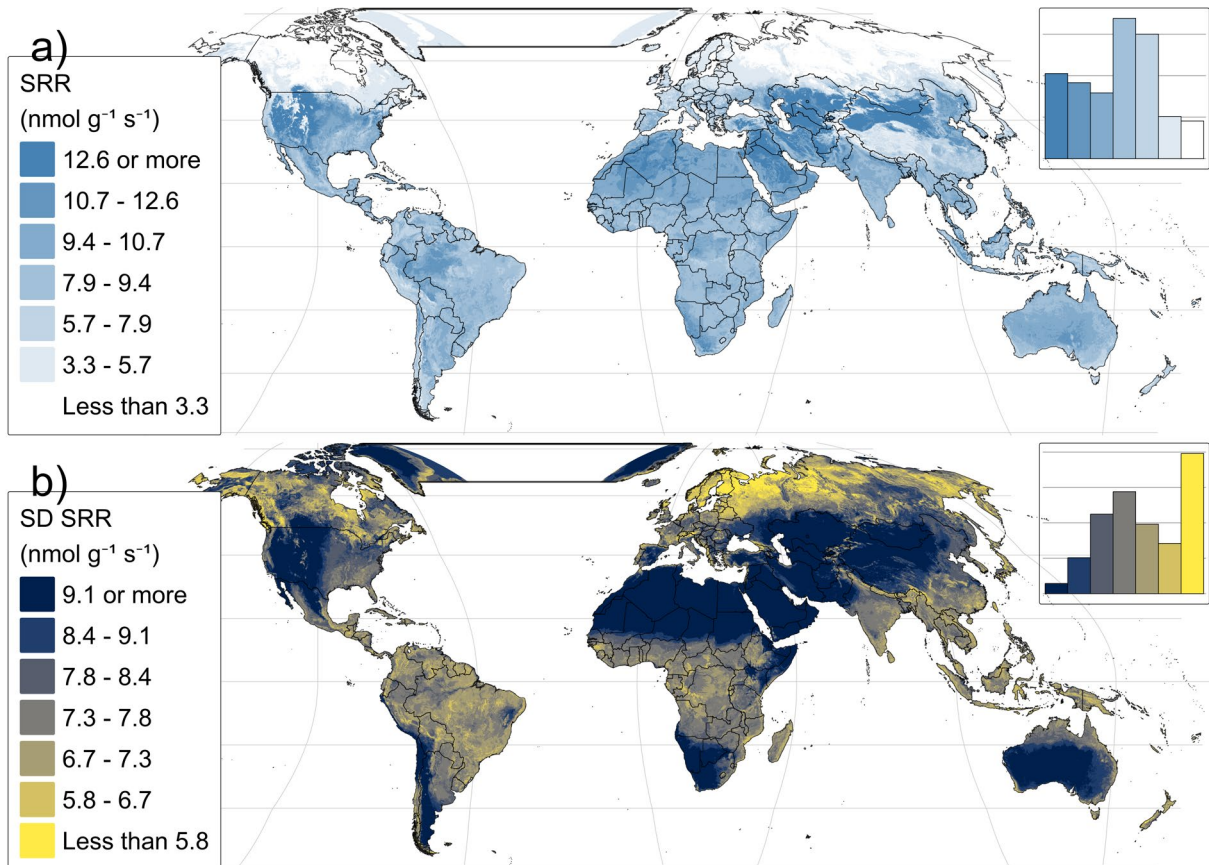
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602

O mapa da intensidade de colonização micorrízica das raízes (RMC) apresentou padrões em parte distintos dos observados para a concentração de carbono nas raízes (RCC), embora também tenha exibido altas taxas de colonização nos trópicos, como nas regiões amazônica, central da África e sudeste asiático, além de ecossistemas temperados, como o sudeste dos Estados Unidos e áreas da Europa (Figura 6). Em contrapartida, as menores taxas de colonização foram observadas principalmente em regiões áridas e desérticas, incluindo os desertos asiáticos, a bacia do Irã e grande parte da Austrália. Diferentemente do padrão observado para RCC, os maiores desvios nas previsões de RMC concentraram-se nas regiões com valores moderados a elevados desse traço, como nos trópicos (especialmente no Brasil e na África central) e em zonas temperadas, como os Estados Unidos, Europa e parte da Rússia.



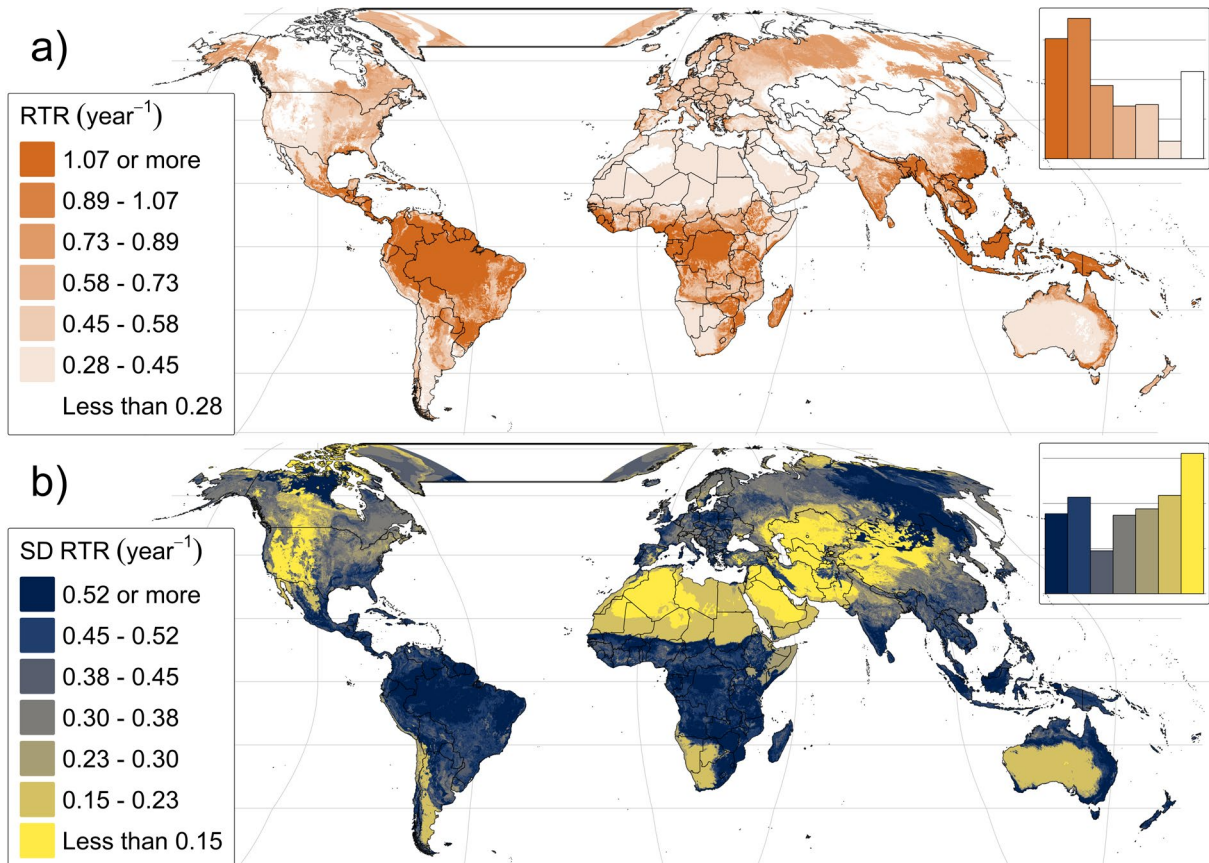
1603 **Figura 6.** Intensidade de colonização micorrízica das raízes (RMC) e desvio padrão das
 1604 previsões (SD RMC) modelados com quantile random forest em escala global.
 1605
 1606

1607 Para a taxa específica de respiração radicular (SRR), observaram-se padrões coerentes
 1608 com as variações ambientais das ecorregiões (Figura 7), com forte tendência latitudinal: os
 1609 valores de SRR diminuíram progressivamente com o aumento da latitude. Altas taxas de
 1610 respiração ($> 25 \text{ nmol g}^{-1} \text{ s}^{-1}$) foram registradas no noroeste intermontano dos Estados Unidos,
 1611 em partes da Rússia e na Mongólia. Regiões com latitudes abaixo de 45°N apresentaram
 1612 predominantemente valores intermediários, enquanto áreas situadas acima de 45°N exibiram
 1613 baixas taxas de SRR, com valores inferiores a $3 \text{ nmol g}^{-1} \text{ s}^{-1}$. Quanto à incerteza das previsões,
 1614 os maiores desvios padrão concentraram-se em zonas áridas, especialmente no deserto do Saara,
 1615 no sudoeste dos Estados Unidos e nos desertos asiáticos, indicando menor confiabilidade nas
 1616 estimativas nessas regiões.



1617
1618 **Figura 7.** Taxa específica de respiração radicular (SRR) e desvio padrão das previsões (SD
1619 SRR) modelados com quantile random forest em escala global.
1620

1621 Para a taxa de renovação radicular (RTR), os maiores valores ($> 1,5 \text{ ano}^{-1}$) foram
1622 estimados para regiões tropicais, com destaque para o sudeste asiático, o centro-leste da África
1623 e o território brasileiro (Figura 8). Em contrapartida, as menores taxas ($< 1 \text{ ano}^{-1}$) seguiram um
1624 padrão latitudinal decrescente em direção ao hemisfério norte, com os valores mais baixos
1625 associados a ecossistemas boreais e de tundra. Além disso, os desertos da Austrália e do Saara
1626 também apresentaram valores reduzidos de RTR. Os maiores desvios padrão das previsões
1627 coincidiram com as áreas de maior RTR, sugerindo maior incerteza associada às estimativas
1628 em regiões de alta rotatividade radicular.
1629

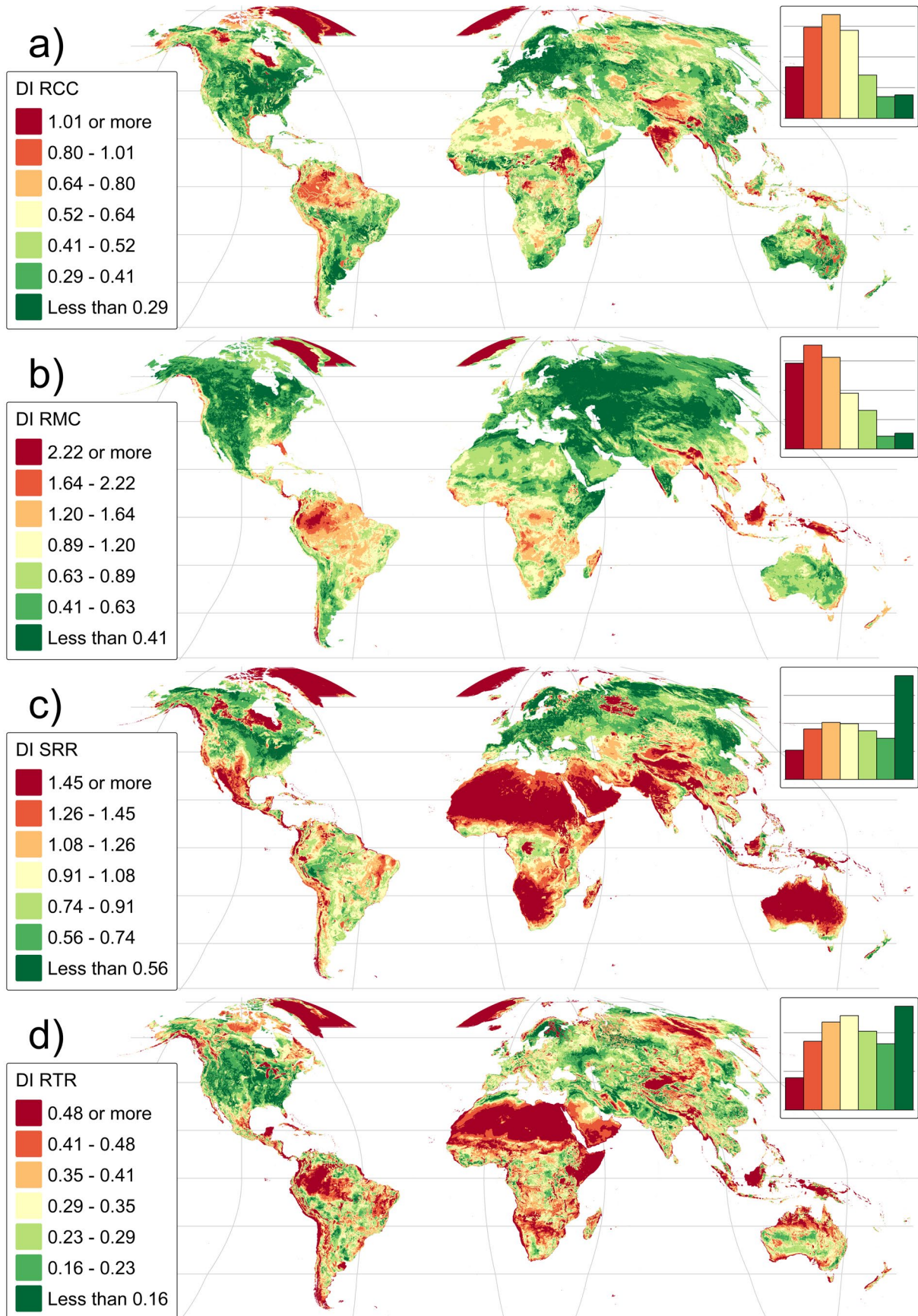


1630
1631
1632
1633

Figura 8. Taxa de renovação radicular (RTR) e desvio padrão das previsões (SD RTR) modelados com quantile random forest em escala global.

1634
1635
1636
1637
1638
1639
1640
1641
1642

O mapeamento do índice de dissimilaridade (DI) para os traços funcionais de raízes indicou padrões espaciais consistentes com as incertezas observadas nas previsões (Figura 9). As maiores dissimilaridades foram registradas em regiões áridas e de altas latitudes, como o deserto do Saara, a Península Arábica, o interior da Austrália, o norte do Canadá e a Groenlândia, coincidindo com áreas de elevada incerteza para RCC, SRR e RTR. Nas florestas tropicais da América do Sul, África Central e sudeste asiático, onde se observaram altos desvios padrão para RMC e RTR, também foram identificados elevados valores de DI. As menores dissimilaridades concentraram-se em regiões temperadas do Hemisfério Norte, em concordância com as áreas de maior confiança preditiva para todos os traços avaliados.

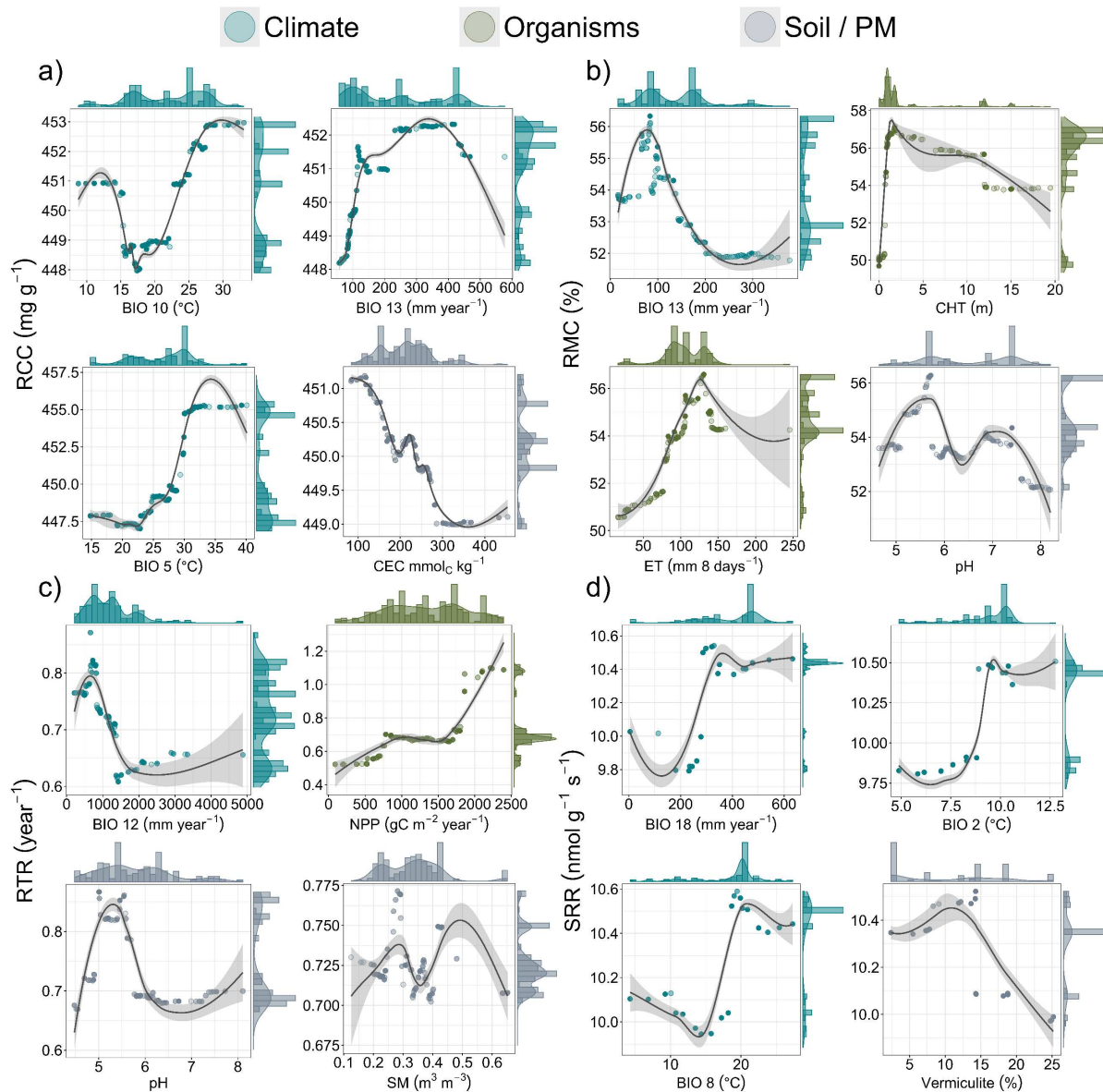


1643
1644
1645
1646

Figura 9. Índice de dissimilaridade (DI) do qrf para a concentração de carbono nas raízes – RCC (a), intensidade de colonização micorrízica – RMC (b), respiração específica das raízes – SRR (c) e taxa de renovação radicular – RTR (d).

1647 3.3. Importância dos preditores e dependência parcial

1648 As análises de dependência parcial da concentração de carbono nas raízes (RCC)
1649 evidenciaram respostas não lineares às principais variáveis climáticas e edáficas (Figura 10).
1650 Observou-se contribuição significativa dos grupos *Climate* e *Soil / PM* — representados por
1651 variáveis de temperatura, precipitação e propriedades do solo — refletindo a complexidade dos
1652 fatores que influenciam a alocação de carbono nas raízes. Entre os preditores climáticos, a
1653 temperatura média do trimestre mais quente (BIO 10) e a temperatura máxima do mês mais
1654 quente (BIO 5) apresentaram padrões semelhantes: a RCC manteve-se estável ou levemente
1655 decrescente até cerca de 15–20 °C, seguido por um aumento acentuado em temperaturas
1656 superiores. A precipitação do mês mais úmido (BIO 13) indicou um platô de RCC entre 250 e
1657 400 mm, com posterior declínio em volumes superiores a 500 mm, sugerindo limitação sob
1658 condições de saturação hídrica. A capacidade de troca catiônica (CEC) apresentou uma relação
1659 inversa com RCC, com redução gradual do traço até aproximadamente 350 mmol_c kg⁻¹, seguida
1660 de leve inflexão.
1661



1662
1663
1664
1665
1666

Figura 10. Curvas de dependência parcial obtidas com quantile random forest para concentração de carbono nas raízes – RCC (a), colonização micorrízica das raízes – RMC (b), respiração específica das raízes – SRR (c) e taxa de renovação radicular – RTR (d).

1667
1668
1669
1670
1671
1672

As dependências parciais da taxa específica de respiração radicular (SRR) indicaram forte influência de variáveis climáticas e edáficas (Figura 10). A precipitação do trimestre mais quente (BIO 18) apresentou relação positiva com SRR, com aumento acentuado até cerca de 300 mm ano⁻¹, seguido de estabilização. A amplitude média diária da temperatura (BIO 2) também mostrou relação positiva com SRR: a partir de ~7,5 °C, a respiração radicular aumentou continuamente, refletindo maior atividade metabólica em ambientes com maior variação

1673 térmica. A temperatura média do trimestre mais úmido (BIO 8) apresentou padrão em dois
1674 estágios, com valores estáveis até cerca de 15 °C e incremento expressivo da SRR em
1675 temperaturas superiores. Entre os preditores edáficos, o teor de vermiculita apresentou relação
1676 inversa com SRR: valores mais elevados (> 15%) foram associados à redução da respiração
1677 específica das raízes, possivelmente refletindo efeitos sobre a estrutura física do solo.

1678 A taxa de renovação radicular (RTR) foi o traço que selecionou o maior número de
1679 variáveis entre os analisados (Figuras S1 - S4). Diferentemente dos demais traços, RTR atribuiu
1680 maior importância a variáveis relacionadas aos organismos e ao clima, com menor peso para
1681 propriedades edáficas. A produtividade primária líquida (NPP) mostrou relação diretamente
1682 proporcional com RTR: até cerca de 1.500 gC m⁻² ano⁻¹ os valores do traço aumentaram
1683 levemente, com elevação mais acentuada a partir desse ponto. A precipitação anual (BIO 12)
1684 demonstrou padrão inverso: RTR foi mais elevada em valores baixos a moderados
1685 (~1.000 mm ano⁻¹), com redução progressiva à medida que os totais aumentaram, sugerindo
1686 limitação sob regimes de alta umidade. O pH do solo revelou relação não linear, com pico de
1687 RTR em torno de pH 5 e redução em condições mais ácidas ou alcalinas. A umidade volumétrica
1688 do solo (SM) apresentou padrão bimodal, com dois picos em 0,25 e 0,45 m³ m⁻³.

1689

1690 **4. Discussão**

1691 **4.1. Dados globais de alta resolução sobre traços funcionais de raízes**

1692 Este estudo representa o primeiro esforço para gerar dados globais, especialmente
1693 explícitos e de alta resolução sobre atributos funcionais de raízes que extrapolam a dimensão
1694 da respiração autotrófica. Enquanto trabalhos anteriores, como o de Jian et al. (2022), focaram
1695 exclusivamente na modelagem da respiração autotrófica, o presente trabalho amplia essa
1696 abordagem ao incluir aspectos relacionados à dinâmica e decomposição (RTR), à associação
1697 com microrganismos simbiotes (RMC), à composição química (RCC) e à fisiologia radicular

1698 (SRR). Essa caracterização multifuncional proporciona uma visão integrada da rizosfera,
1699 promovendo avanços na compreensão das interações entre raízes, solo e fatores ambientais. A
1700 abordagem adotada contribui para superar a lacuna histórica de representação das funções
1701 radiculares em estudos ecológicos em escala global (Lee & Andersen, 2023; Li et al., 2022;
1702 Rustad et al., 2000).

1703 A sub-representação dos dados, especialmente nas regiões tropicais, é amplamente
1704 reconhecida na literatura. Historicamente, a ecologia funcional concentrou seus esforços em
1705 compreender os órgãos das plantas acima do solo, enquanto a rizosfera permaneceu
1706 negligenciada (Lee & Andersen, 2023). Esse desequilíbrio é, em grande parte, atribuído à maior
1707 complexidade envolvida no acesso, observação e estudo direto das raízes em campo. Em
1708 contraste, a parte aérea das plantas tem sido amplamente estudada, favorecida pela
1709 disponibilidade de sensores remotos, inventários florestais e métodos padronizados de coleta,
1710 o que possibilitou a geração de um número considerável de mapas e modelos globais de
1711 estrutura da vegetação (J. M. Chen et al., 2005; Dechant et al., 2024; Fisher & Mustard, 2007;
1712 Kier et al., 2005; Liu et al., 2024; Moreno-Martínez et al., 2018; Williamson et al., 2024; Wolf
1713 et al., 2022; Zhao et al., 2024). Essa assimetria metodológica e tecnológica contribuiu para uma
1714 lacuna significativa no entendimento integrado entre componentes acima e abaixo do solo.

1715 Embora as pesquisas relacionadas às raízes em campo sejam limitadas, estudos em
1716 laboratório e em casas de vegetação em níveis micro ocorrem com mais frequência (Comas et
1717 al., 2013; Fang & Xiong, 2015; Gruber et al., 2013). Além disso, a intensidade de pesquisas nos
1718 ecossistemas temperados em comparação com os trópicos contribui para a cobertura limitada
1719 de informações sobre traços funcionais nos trópicos (Guerrero-Ramírez et al., 2021).

1720

1721 **4.2. Relações ambientais e traços funcionais de raízes**

1722 Para todos os traços estudados, as correlações ambientais detectadas pelo modelo
1723 apontam fortes relações dos traços funcionais das raízes com a vegetação. Além da ancoragem,
1724 a função principal das raízes é a absorção de nutrientes, principalmente nitrogênio (N) e fósforo
1725 (P), além de água do solo (Lambers et al., 2008). Isso faz com que as raízes em todo o planeta
1726 desenvolvam características e estratégias para alcançar esses objetivos e responder às
1727 perturbações, especialmente às mudanças no espaço e no tempo (Bardgett et al., 2014).

1728 Nesse contexto, os trópicos, especialmente as florestas tropicais como a Amazônia e a
1729 floresta da bacia do Congo, apresentam solos altamente intemperizados e com baixa
1730 disponibilidade de nutrientes (Quesada et al., 2010, 2012). A escassez de recursos minerais
1731 nesses ambientes, aliada à elevada produtividade primária resultante das condições climáticas
1732 favoráveis (altas temperaturas e elevada precipitação), estimula o investimento das plantas na
1733 construção de raízes. Além das florestas tropicais, formações savânicas tropicais, como o
1734 Cerrado brasileiro e o miombo africano, também se destacam por combinar solos oligotróficos
1735 com elevada biomassa subterrânea, reflexo de estratégias adaptativas voltadas à aquisição
1736 eficiente de nutrientes e água em condições edafoclimáticas limitantes (Lehmann et al., 2014;
1737 Ryan et al., 2011; Tomlinson et al., 2012). Esse processo envolve o investimento de carbono
1738 em arquitetura e morfologia, como o aumento do comprimento, e características bióticas que
1739 envolvem interação com a biota do solo (fungos, rizóbios, entre outros) para solubilizar
1740 nutrientes (Bardgett et al., 2014).

1741 Todos esses processos, que muitas vezes são limitados em regiões com forte
1742 sazonalidade, onde há escassez hídrica ou restrições térmicas extremas (tanto por frio quanto
1743 por calor excessivo), ajudam a explicar os maiores valores de concentração de carbono nas
1744 raízes (RCC) e taxa de renovação radicular (RTR) nas regiões tropicais, em contraste com
1745 desertos, ecossistemas boreais e tundras. Nesse contexto, destacam-se variáveis associadas à

1746 vegetação, como a produtividade primária líquida (NPP) e a altura média do dossel (CHT), que
1747 apresentaram relações diretamente proporcionais com RCC e RTR, refletindo o papel da
1748 estrutura e do vigor da vegetação sobre o investimento radicular. A evapotranspiração (ET)
1749 também emergiu como variável relevante, indicando que fluxos intensos de água e energia estão
1750 associados a maior atividade e renovação do sistema radicular em regiões de alta produtividade.

1751 As altas taxas estimadas de colonização micorrízica radicular (RMC) nos ecossistemas
1752 tropicais e temperados são resultado de interações entre vegetação, fatores edáficos e
1753 climáticos. Apesar do banco de dados do GRooT compilar dados gerais de colonização
1754 micorrízica, predominam os dados de micorrizas arbusculares (Guerrero-Ramírez et al., 2021).

1755 O sudeste asiático e os ecossistemas boreais são conhecidos pela dominância de plantas
1756 hospedeiras para a colonização (Read, 1991). No entanto, os demais ecossistemas aqui
1757 apontados como regiões com colonização moderada a alta não apresentam essa dominância, o
1758 que sugere a forte influência das condições climáticas (temperatura e umidade) e edáficas (pH
1759 do solo, disponibilidade de N, textura, por exemplo) (Soudzilovskaia et al., 2015).

1760 Complementando esses achados, estudos anteriores indicaram que a diversidade
1761 morfológica das raízes é proporcionalmente maior, acompanhando a diversidade de plantas dos
1762 ecossistemas, como, por exemplo, nos trópicos, onde os grupos filogenéticos ancestrais são
1763 mais preservados (Field et al., 2015; Kenrick & Crane, 1997; Ma et al., 2018).

1764 Essas relações permitiram a evolução da morfologia das raízes para uma maior
1765 adaptabilidade de duas formas: uma delas é a associação com fungos micorrízicos, que diminui
1766 à medida que o gradiente edafoclimático entre as ecorregiões se torna sazonalmente inóspito
1767 (temperado, boreal, tundra/desertos), assim como observado nos mapas de traços funcionais
1768 produzidos. A outra via é o investimento em raízes finas, o que melhora a exploração do solo
1769 por unidade de carbono utilizada, maximizando a eficiência na absorção de nutrientes.

1770 Outro fator importante a ser considerado são as regiões onde se observa a coincidência
1771 espacial entre altos valores de colonização micorrízica (RMC) e de respiração específica das
1772 raízes (SRR), dado seu papel central no ciclo do carbono do solo. Um dos principais processos
1773 envolvidos nesse ciclo é o efluxo de CO₂ do solo, considerado o maior fluxo anual de carbono
1774 da biosfera terrestre para a atmosfera (Raich & Schlesinger, 1992). No que se refere às raízes,
1775 esse efluxo é regulado predominantemente por dois processos metabólicos: a respiração
1776 autotrófica, responsável pela oxidação de açúcares gerados na fotossíntese das folhas para
1777 manutenção e crescimento do tecido radicular, e a exsudação radicular, que corresponde à
1778 liberação de compostos orgânicos solúveis no solo (Chapin et al., 2006; Kuzyakov &
1779 Larionova, 2005). Embora a exsudação em si não represente respiração radicular, os compostos
1780 exsudados tornam-se substrato para a atividade microbiana heterotrófica, que, ao mineralizar
1781 esses compostos, contribui indiretamente para o aumento do efluxo de CO₂ (Bais et al., 2006;
1782 de Vries et al., 2019; Hou et al., 2025; Tripathi et al., 2025). A taxa específica de respiração
1783 radicular (SRR), tal como utilizada neste estudo, corresponde ao fluxo de CO₂ liberado pelas
1784 raízes por unidade de massa seca e por unidade de tempo, mensurado geralmente em condições
1785 controladas, sem contabilizar a respiração microbiana posterior à exsudação (Bardgett & Van
1786 Der Putten, 2014; Lak et al., 2020).

1787 As plantas podem alocar quantidades expressivas de carbono fotoassimilado para suas
1788 raízes e simbiontes, e uma grande porção deste carbono é respirada pelas raízes ou liberada no
1789 solo por meio da exsudação da raiz (Jones et al., 2004). A contribuição da respiração das raízes
1790 para o fluxo total de CO₂ do solo atinge valores entre 40 e 50% (Bond-Lamberty et al., 2004;
1791 Hanson et al., 2000), embora varie consideravelmente entre espécies de plantas e com outros
1792 fatores como a disponibilidade de nitrogênio na raiz e no solo (Phillips & Fahey, 2007; Reich
1793 et al., 2008) e a presença e tipo de fungos micorrízicos colonizando as raízes (Martin & Stutz,
1794 2004; Trocha et al., 2010).

1795 Ainda se sabe pouco sobre como a exsudação das raízes afeta a respiração do solo,
1796 embora parte significativa das pesquisas mostre que a liberação de exsudatos pode estimular a
1797 atividade microbiana e a decomposição da matéria orgânica, aumentando assim a perda de
1798 carbono do solo em uma escala de ecossistema (Fontaine et al., 2004).

1799 Adicionalmente, os elevados valores de SRR observados especialmente nos desertos
1800 refletem padrões baseados na temperatura. A respiração do solo, seja autotrófica ou
1801 heterotrófica (Carey et al., 2016; Rustad et al., 2000; Sáez-Sandino et al., 2023; Wang et al.,
1802 2014), responde principalmente à temperatura, padrão este estabelecido pelo modelo ajustado
1803 (Figura 10). Uma vez que esse padrão térmico é estabelecido, observa-se que, à medida que a
1804 temperatura aumenta, a respiração também tende a se intensificar, o que explica as altas
1805 projeções de SRR em regiões mais quentes. No entanto, a respiração nessas áreas pode ser
1806 fortemente limitada pela escassez de biomassa, tanto aérea quanto subterrânea, o que requer
1807 cautela na interpretação dos dados. Por outro lado, a tendência de redução da SRR em altas
1808 latitudes (acima de 45°N), conforme evidenciado nos mapas de predição, pode estar relacionada
1809 à curta duração da estação de crescimento, à menor atividade microbiana e às baixas
1810 temperaturas médias anuais. Esses fatores reduzem a demanda metabólica das raízes e limitam
1811 os fluxos de carbono abaixo do solo (Bardgett et al., 2005; W. Chen et al., 2013).

1812 Considerando esses padrões, ainda que o desempenho dos modelos ajustados não tenha
1813 sido elevado para todos os traços, as métricas foram significativamente superiores às do modelo
1814 nulo, o que justifica sua aplicação. O traço RTR ilustra bem essa condição: embora o RMSE
1815 não tenha diferido do modelo nulo, o MAE foi significativamente menor, revelando um ganho
1816 preditivo médio, apesar da maior dispersão dos erros. Essa diferença metodológica entre
1817 métricas é amplamente reconhecida, sendo o RMSE mais sensível a *outliers*, enquanto o MAE
1818 oferece uma estimativa mais robusta do erro médio (Chai & Draxler, 2014). Além disso, RTR
1819 foi o traço com menor número de amostras no banco de dados ($n = 523$), o que certamente

1820 contribuiu para sua maior incerteza. Dada essa limitação, reforça-se a necessidade de novos
1821 esforços para expandir a cobertura amostral e melhorar a qualidade das predições para traços
1822 com baixa representatividade global.

1823 Nesse sentido, o mapeamento do índice de dissimilaridade (DI) complementou as
1824 análises de predição ao indicar padrões de extrapolação ambiental associados às áreas de maior
1825 incerteza dos modelos. Regiões áridas e de altas latitudes, como o deserto do Saara, a Península
1826 Arábica, a Groenlândia e o norte do Canadá, apresentaram elevados valores de DI, refletindo
1827 baixa representatividade amostral no espaço ambiental de treinamento. De forma ainda mais
1828 preocupante, ecossistemas tropicais altamente biodiversos, como a Amazônia e a bacia do
1829 Congo, também apresentaram altos valores de DI para traços como RMC, o que evidencia a
1830 carência crítica de dados nessas regiões. Essa deficiência limita severamente a capacidade de
1831 gerar informações cientificamente sólidas (Meyer & Pebesma, 2021) para a compreensão dos
1832 processos abaixo do solo e compromete a formulação de políticas públicas efetivas voltadas à
1833 conservação e ao manejo sustentável. A utilização do DI como métrica adicional permitiu
1834 identificar de forma explícita essas lacunas, reforçando a urgência de investimentos em
1835 campanhas de amostragem direcionadas, especialmente na faixa tropical, para assegurar maior
1836 robustez e aplicabilidade dos modelos globais de traços funcionais de raízes.

1837

1838 **4.3. Governança biótica e edáfica dos traços funcionais de raízes**

1839 As análises de dependência parcial e a importância relativa dos preditores reafirmam
1840 que fatores bióticos (como altura do dossel - CHT, produção primária líquida - NPP e
1841 evapotranspiração - ET) e edáficos (como a mineralogia do solo, pH, CTC e umidade do solo)
1842 governam a distribuição dos traços funcionais de raízes estudados globalmente (Figura S5).

1843 Além disso, preditores climáticos que expressam sazonalidade ou condições extremas,
1844 como BIO 13 e BIO 5 , destacaram-se na explicação da distribuição dos atributos funcionais

1845 das raízes. Isso ocorre porque a sazonalidade influencia diretamente o funcionamento do solo,
1846 modulando a disponibilidade de água e nutrientes ao longo do ano. Por exemplo, em regiões
1847 com estações secas bem definidas, a umidade do solo pode ser drasticamente reduzida durante
1848 determinados períodos, afetando a atividade microbiana e a mineralização da matéria orgânica
1849 (Orchard & Cook, 1983). Esse processo, por sua vez, regula a liberação de nutrientes essenciais
1850 às plantas, como nitrogênio e fósforo. Em ambientes onde a sazonalidade térmica é elevada,
1851 como nas zonas temperadas e boreais, o congelamento e o descongelamento do solo impactam
1852 o *turnover* da matéria orgânica e os fluxos biogeoquímicos (Crowther et al., 2016; Knorr et al.,
1853 2005). Assim, a variabilidade climática sazonal cria pulsos de recursos e impõe restrições
1854 periódicas, moldando estratégias radiculares adaptativas como o aumento da taxa de renovação
1855 (RTR) em ambientes de alta instabilidade ou a alocação de carbono nas raízes (RCC) para
1856 maximizar a absorção durante janelas de oportunidade favoráveis (Stuart Chapin et al., 2012).

1857 Além disso, Freschet et al. (2017), ao analisarem um banco de dados de raízes finas,
1858 encontraram uma forte relação entre as condições edafoclimáticas e a distribuição global das
1859 raízes. Ainda sobre a sazonalidade, as raízes com comprimento específico elevado tendem a
1860 investir mais na sua morfologia para explorar o volume do solo, o que implica em estratégias
1861 de retorno mais rápido (Eissenstat et al., 2000).

1862 Isso pode ser útil em ambientes mais frios (como ecossistemas temperados), onde as
1863 taxas de absorção de nutrientes pelas raízes são menores (Chapin, 1974) e onde a forte
1864 sazonalidade e o congelamento do solo podem levar a uma atividade microbiana do solo
1865 heterogênea ou até interrompida por grandes períodos (Bardgett et al., 2005; W. Chen et al.,
1866 2013). Essa tendência é coerente com os elevados valores encontrados de RCC e RMC, por
1867 exemplo, em habitats mais frios.

1868 Essa relação entre sazonalidade, atividade radicular e a distribuição de traços funcionais
1869 de raízes também tem implicações significativas para o ciclo do carbono no solo. Como o

1870 carbono do solo é influenciado em pouco mais de 20%, dependendo da literatura (Brunner &
1871 Godbold, 2007), pelas raízes, especialmente pelas raízes mortas, a RCC e a RTR tornam-se
1872 fundamentais para uma compreensão aprofundada dos compartimentos do carbono do solo
1873 global.

1874 Portanto, a análise desses fatores é fundamental não apenas para entender a distribuição
1875 atual do carbono no solo, mas também para realizar cálculos de orçamento de carbono no
1876 presente e fazer projeções futuras. Por fim, é importante destacar as variáveis relacionadas ao
1877 solo e como foram essenciais para o qrf extrair padrões na modelagem dos traços funcionais de
1878 raízes, especialmente para RMC e SRR.

1879 Embora não tenham sido incluídas variáveis mais dinâmicas, como P, que a literatura já
1880 estabelece com fortes relações com os traços funcionais (Chapin, 1974; Jones et al., 2004;
1881 Martin & Stutz, 2004; Phillips & Fahey, 2007), os dados mineralógicos carregam informações
1882 relevantes sobre a gênese e formação dos solos. Além disso, esses dados foram baseados nas
1883 classes de solos globais (Ito & Wagai, 2017).

1884 Esses dois fatores fornecem informações integradas sobre o ambiente edáfico. A
1885 mineralogia, por exemplo, pode atuar como um indicador indireto da fertilidade química. Solos
1886 com predominância de minerais primários geralmente estão associados a materiais de origem
1887 mais jovens, com maior concentração de bases trocáveis (Creamer et al., 2019; Singh, 2023).
1888 Em contraste, solos com elevada concentração de óxidos de ferro refletem maior grau de
1889 intemperismo, indicando maior desenvolvimento pedogenético e melhor estrutura física
1890 (Oades, 1988). Esse é o caso dos Latossolos, amplamente distribuídos em regiões tropicais
1891 úmidas, como o Brasil Central e a bacia do Congo. Nessas ecorregiões, a baixa fertilidade
1892 natural é compensada por estratégias adaptativas das plantas, como o investimento em raízes
1893 finas, associações micorrízicas intensas e alta taxa de renovação radicular, refletindo a interação
1894 entre condições edáficas restritivas e elevada produtividade primária (Yaffar et al., 2024).

1895 Por outro lado, as classes de solo são categorias que integram solos com características
1896 físicas, químicas, biológicas e morfológicas semelhantes. Essa combinação de fatores é
1897 potencialmente capaz de fornecer pistas sobre a paisagem, como o tipo de vegetação, e,
1898 conseqüentemente, correlações com os traços funcionais de raízes (como RMC e SRR).
1899 Dependendo do solo, ele pode ser altamente dependente das raízes e de seus exsudados para
1900 manter sua estrutura (Czarnes et al., 2000; Whalley et al., 2005).

1901

1902 **4.4. Limitações e perspectivas futuras**

1903 O qrf ajustado aqui, apesar de apresentar desempenho superior ao modelo nulo,
1904 recomenda-se cautela na utilização dos dados. Os valores apresentaram erros grandes,
1905 especialmente nos traços com menor representatividade global, como o RMC e RTR. Como
1906 mencionado anteriormente, os trópicos são heterogêneos e carecem de uma amostragem mais
1907 densa, especialmente para os atributos com menor quantidade de amostras, devido a uma
1908 diversidade de fatores já discutidos neste estudo, como a quantidade de pesquisas e os métodos
1909 onerosos por exemplo. Adicionalmente, os mapas de desvio padrão das predições indicaram
1910 que as maiores incertezas se concentraram nas regiões com menor densidade amostral, como
1911 zonas áridas, áreas tropicais e porções da Ásia Central. Isso foi particularmente evidente nos
1912 traços RMC e RTR, que apresentaram ampla variabilidade preditiva nessas regiões. Por outro
1913 lado, áreas com melhor cobertura de dados, como o norte do Canadá e partes da Europa,
1914 mostraram menor incerteza, sugerindo maior robustez das predições nestas localidades

1915 Aumentar essas amostras pode aumentar a acurácia de modelagens futuras de traços
1916 funcionais de raízes. Além disso, os preditores utilizados carregam incertezas inerentes aos
1917 processos de modelagem dos quais derivam, o que introduz incertezas nos mapas produzidos.
1918 Os padrões reproduzidos pelo modelo para as regiões desérticas referentes ao SRR podem ter
1919 sido impulsionados por isso, somado a falta de amostras nessas regiões e pelos padrões de

1920 temperatura. Além disso, ressalta-se a presença de padrões anômalos identificado ao norte do
1921 Saara nas predições de RCC podem refletir limitações na cobertura de dados e na representação
1922 ambiental, o que reforça a importância de validações regionais mais detalhadas

1923 Além dos traços funcionais utilizados, há traços e métricas adicionais não considerados
1924 aqui que podem capturar aspectos complementares das funções ecossistêmicas do solo. O
1925 avanço futuro com a modelagem desses traços, a partir de um esforço na composição de bancos
1926 de dados, pode auxiliar na compreensão da dinâmica da rizosfera globalmente e avançar na sub-
1927 representação deste compartimento tão importante. Aumentar a representação de tais traços em
1928 bancos de dados é um próximo passo importante e crítico para quantificar a multifuncionalidade
1929 da biodiversidade do solo.

1930

1931 **5. Conclusão**

1932 Este estudo representa um avanço importante na compreensão dos traços funcionais de
1933 raízes em escala global, ao integrar informações que extrapolam a dimensão tradicional da
1934 respiração. A combinação de traços relacionados à dinâmica e decomposição (RTR), associação
1935 microbiana (RMC), composição química (RCC) e fisiologia radicular (SRR) permitiu uma
1936 avaliação mais abrangente dos processos ecossistêmicos associados à rizosfera.

1937 Os resultados evidenciaram a complexa interação entre as raízes e os fatores bióticos e
1938 abióticos, ressaltando seu papel central na ciclagem de nutrientes e no fluxo de carbono no solo.
1939 No entanto, a sub-representação de dados em regiões tropicais e áridas destaca a necessidade
1940 de ampliar o esforço de amostragem e monitoramento nessas áreas, o que é essencial para
1941 reduzir as incertezas e aprimorar a capacidade preditiva dos modelos.

1942 Avanços futuros devem priorizar a expansão e harmonização dos bancos de dados de
1943 traços funcionais, bem como a inclusão de métricas adicionais que capturem aspectos
1944 complementares das funções ecossistêmicas subterrâneas. Tais esforços são fundamentais para

1945 fortalecer a quantificação da multifuncionalidade da biodiversidade do solo em escala global e
1946 apoiar políticas baseadas em evidências para o manejo e conservação dos ecossistemas
1947 terrestres.

1948

1949 Referências

- 1950 Bais, H. P., Weir, T. L., Perry, L. G., Gilroy, S., & Vivanco, J. M. (2006). The role of root
1951 exudates in rhizosphere interactions with plants and other organisms. *Annual Review of*
1952 *Plant Biology*, 57(Volume 57, 2006), 233–266.
1953 <https://doi.org/10.1146/annurev.arplant.57.032905.105159>
- 1954 Bardgett, R. D., Bowman, W. D., Kaufmann, R., & Schmidt, S. K. (2005). A temporal approach
1955 to linking aboveground and belowground ecology. *Trends in Ecology and Evolution*,
1956 20(11), 634–641. <https://doi.org/10.1016/j.tree.2005.08.005>
- 1957 Bardgett, R. D., Mommer, L., & De Vries, F. T. (2014). Going underground: Root traits as
1958 drivers of ecosystem processes. *Trends in Ecology and Evolution*, 29(12), 692–699.
1959 <https://doi.org/10.1016/j.tree.2014.10.006>
- 1960 Bardgett, R. D., & Van Der Putten, W. H. (2014a). Belowground biodiversity and ecosystem
1961 functioning. *Nature*, 515(7528), 505–511. <https://doi.org/10.1038/nature13855>
- 1962 Bardgett, R. D., & Van Der Putten, W. H. (2014b). Belowground biodiversity and ecosystem
1963 functioning. *Nature*, 515(7528), 505–511. <https://doi.org/10.1038/nature13855>
- 1964 Bengtsson, H. (2023). *parallelly: Enhancing the “parallel” Package* (1.36.0). [https://cran.r-](https://cran.r-project.org/package=parallelly)
1965 [project.org/package=parallelly](https://cran.r-project.org/package=parallelly)
- 1966 Bergmann, J., Ryo, M., Prati, D., Hempel, S., & Rillig, M. C. (2017). Root traits are more than
1967 analogues of leaf traits: the case for diaspore mass. *New Phytologist*, 216(4), 1130–1139.
1968 <https://doi.org/10.1111/nph.14748>

- 1969 Bond-Lamberty, B., Wang, C., & Gower, S. T. (2004). A global relationship between the
1970 heterotrophic and autotrophic components of soil respiration? *Global Change Biology*,
1971 *10*(10), 1756–1766. <https://doi.org/10.1111/j.1365-2486.2004.00816.x>
- 1972 Brunner, I., & Godbold, D. L. (2007). Tree roots in a changing world. *Journal of Forest*
1973 *Research*, *12*(2), 78–82. <https://doi.org/10.1007/s10310-006-0261-4>
- 1974 Cadotte, M. W., Cavender-Bares, J., Tilman, D., & Oakley, T. H. (2009). Using phylogenetic,
1975 functional and trait diversity to understand patterns of plant community productivity. *PLoS*
1976 *ONE*, *4*(5), e5695. <https://doi.org/10.1371/journal.pone.0005695>
- 1977 Carey, J. C., Tang, J., Templer, P. H., Kroeger, K. D., Crowther, T. W., Burton, A. J., Dukes, J.
1978 S., Emmett, B., Frey, S. D., Heskell, M. A., Jiang, L., Machmuller, M. B., Mohan, J.,
1979 Panetta, A. M., Reich, P. B., Reinschj, S., Wang, X., Allison, S. D., Bamminger, C., ...
1980 Tietema, A. (2016). Temperature response of soil respiration largely unaltered with
1981 experimental warming. *Proceedings of the National Academy of Sciences of the United*
1982 *States of America*, *113*(48), 13797–13802. <https://doi.org/10.1073/pnas.1605365113>
- 1983 Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error
1984 (MAE)? -Arguments against avoiding RMSE in the literature. *Geoscientific Model*
1985 *Development*, *7*(3), 1247–1250. <https://doi.org/10.5194/gmd-7-1247-2014>
- 1986 Chapin, F. S. (1974). Morphological and Physiological Mechanisms of Temperature
1987 Compensation in Phosphate Absorption along a Latitudinal Gradient. *Ecology*, *55*(6),
1988 1180–1198. <https://doi.org/10.2307/1935449>
- 1989 Chapin, F. S., Woodwell, G. M., Randerson, J. T., Rastetter, E. B., Lovett, G. M., Baldocchi, D.
1990 D., Clark, D. A., Harmon, M. E., Schimel, D. S., Valentini, R., Wirth, C., Aber, J. D., Cole,
1991 J. J., Goulden, M. L., Harden, J. W., Heimann, M., Howarth, R. W., Matson, P. A.,
1992 McGuire, A. D., ... Schulze, E. D. (2006). Reconciling carbon-cycle concepts,

- 1993 terminology, and methods. *Ecosystems*, 9(7), 1041–1050. <https://doi.org/10.1007/s10021->
- 1994 005-0105-7
- 1995 Chen, J. M., Menges, C. H., & Leblanc, S. G. (2005). Global mapping of foliage clumping
- 1996 index using multi-angular satellite data. *Remote Sensing of Environment*, 97(4), 447–457.
- 1997 <https://doi.org/10.1016/j.rse.2005.05.003>
- 1998 Chen, W., Zeng, H., Eissenstat, D. M., & Guo, D. (2013). Variation of first-order root traits
- 1999 across climatic gradients and evolutionary trends in geological time. *Global Ecology and*
- 2000 *Biogeography*, 22(7), 846–856. <https://doi.org/10.1111/geb.12048>
- 2001 Clark, D. A., Clark, D. B., & Oberbauer, S. F. (2013). Field-quantified responses of tropical
- 2002 rainforest aboveground productivity to increasing CO₂ and climatic stress, 1997-2009.
- 2003 *Journal of Geophysical Research: Biogeosciences*, 118(2), 783–794.
- 2004 <https://doi.org/10.1002/jgrg.20067>
- 2005 Comas, L. H., Becker, S. R., Cruz, V. M. V., Byrne, P. F., & Dierig, D. A. (2013). Root traits
- 2006 contributing to plant productivity under drought. *Frontiers in Plant Science*, 4(NOV),
- 2007 62325. <https://doi.org/10.3389/fpls.2013.00442>
- 2008 Creamer, C. A., Foster, A. L., Lawrence, C., McFarland, J., Schulz, M., & Waldrop, M. P.
- 2009 (2019). Mineralogy dictates the initial mechanism of microbial necromass association.
- 2010 *Geochimica et Cosmochimica Acta*, 260, 161–176.
- 2011 <https://doi.org/10.1016/j.gca.2019.06.028>
- 2012 Crowther, T. W., Todd-Brown, K. E. O., Rowe, C. W., Wieder, W. R., Carey, J. C., MacHmuller,
- 2013 M. B., Snoek, B. L., Fang, S., Zhou, G., Allison, S. D., Blair, J. M., Bridgham, S. D.,
- 2014 Burton, A. J., Carrillo, Y., Reich, P. B., Clark, J. S., Classen, A. T., Dijkstra, F. A., Elberling,
- 2015 B., ... Bradford, M. A. (2016). Quantifying global soil carbon losses in response to
- 2016 warming. *Nature* 2016 540:7631, 540(7631), 104–108.
- 2017 <https://doi.org/10.1038/nature20150>

- 2018 Czarnes, S., Hallett, P. D., Bengough, A. G., & Young, I. M. (2000). Influence de mucilages
2019 racinaire et microbiens modeles sur la structure du sol et le transport d'eau. *European*
2020 *Journal of Soil Science*, 51(3), 435–443. [https://doi.org/10.1046/j.1365-](https://doi.org/10.1046/j.1365-2389.2000.00327.x)
2021 2389.2000.00327.x
- 2022 de Vries, F. T., Williams, A., Stringer, F., Willcocks, R., McEwing, R., Langridge, H., &
2023 Straathof, A. L. (2019). Changes in root-exudate-induced respiration reveal a novel
2024 mechanism through which drought affects ecosystem carbon cycling. *New Phytologist*,
2025 224(1), 132–145. <https://doi.org/10.1111/nph.16001>
- 2026 Dechant, B., Kattge, J., Pavlick, R., Schneider, F. D., Sabatini, F. M., Moreno-Martínez, Á.,
2027 Butler, E. E., van Bodegom, P. M., Vallicrosa, H., Kattenborn, T., Boonman, C. C. F.,
2028 Madani, N., Wright, I. J., Dong, N., Feilhauer, H., Peñuelas, J., Sardans, J., Aguirre-
2029 Gutiérrez, J., Reich, P. B., ... Townsend, P. A. (2024). Intercomparison of global foliar trait
2030 maps reveals fundamental differences and limitations of upscaling approaches. *Remote*
2031 *Sensing of Environment*, 311. <https://doi.org/10.1016/j.rse.2024.114276>
- 2032 Díaz, S., Lavorel, S., De Bello, F., Quétier, F., Grigulis, K., & Robson, T. M. (2007).
2033 Incorporating plant functional diversity effects in ecosystem service assessments.
2034 *Proceedings of the National Academy of Sciences of the United States of America*, 104(52),
2035 20684–20689. <https://doi.org/10.1073/pnas.0704716104>
- 2036 Eissenstat, D. M., Wells, C. E., Yanai, R. D., & Whitbeck, J. L. (2000). Building roots in a
2037 changing environment: Implications for root longevity. *New Phytologist*, 147(1), 33–42.
2038 <https://doi.org/10.1046/j.1469-8137.2000.00686.x>
- 2039 Fan, Y., Li, H., & Miguez-Macho, G. (2013). Global patterns of groundwater table depth.
2040 *Science*, 339(6122), 940–943. <https://doi.org/10.1126/science.1229881>

- 2041 Fang, Y., & Xiong, L. (2015). General mechanisms of drought response and their application
2042 in drought resistance improvement in plants. *Cellular and Molecular Life Sciences*, 72(4),
2043 673–689. <https://doi.org/10.1007/s00018-014-1767-0>
- 2044 Fernandes Filho, E. I., Moquedace, C. M., & Veloso, G. V. (2021). *mdsFuncs: Auxiliar functions*
2045 *to pre-process raster data and modelling with machine learning models using caret*
2046 *package* (0.2.10).
- 2047 Field, K. J., Pressel, S., Duckett, J. G., Rimington, W. R., & Bidartondo, M. I. (2015). Symbiotic
2048 options for the conquest of land. *Trends in Ecology and Evolution*, 30(8), 477–486.
2049 <https://doi.org/10.1016/j.tree.2015.05.007>
- 2050 Finzi, A. C., Abramoff, R. Z., Spiller, K. S., Brzostek, E. R., Darby, B. A., Kramer, M. A., &
2051 Phillips, R. P. (2015). Rhizosphere processes are quantitatively important components of
2052 terrestrial carbon and nutrient cycles. *Global Change Biology*, 21(5), 2082–2094.
2053 <https://doi.org/10.1111/gcb.12816>
- 2054 Fisher, J. I., & Mustard, J. F. (2007). Cross-scalar satellite phenology from ground, Landsat,
2055 and MODIS data. *Remote Sensing of Environment*, 109(3), 261–273.
2056 <https://doi.org/10.1016/j.rse.2007.01.004>
- 2057 Fontaine, S., Bardoux, G., Abbadie, L., & Mariotti, A. (2004). Carbon input to soil may decrease
2058 soil carbon content. *Ecology Letters*, 7(4), 314–320. [https://doi.org/10.1111/j.1461-](https://doi.org/10.1111/j.1461-0248.2004.00579.x)
2059 [0248.2004.00579.x](https://doi.org/10.1111/j.1461-0248.2004.00579.x)
- 2060 Freschet, G. T., Pagès, L., Iversen, C. M., Comas, L. H., Rewald, B., Roumet, C., Klimešová,
2061 J., Zadworny, M., Poorter, H., Postma, J. A., Adams, T. S., Bagniewska-Zadworna, A.,
2062 Bengough, A. G., Blancaflor, E. B., Brunner, I., Cornelissen, J. H. C., Garnier, E., Gessler,
2063 A., Hobbie, S. E., ... McCormack, M. L. (2021). A starting guide to root ecology:
2064 strengthening ecological concepts and standardising root classification, sampling,

- 2065 processing and trait measurements. *New Phytologist*, 232(3), 973–1122.
2066 <https://doi.org/10.1111/nph.17572>
- 2067 Freschet, G. T., Roumet, C., Comas, L. H., Weemstra, M., Bengough, A. G., Rewald, B.,
2068 Bardgett, R. D., De Deyn, G. B., Johnson, D., Klimešová, J., Lukac, M., McCormack, M.
2069 L., Meier, I. C., Pagès, L., Poorter, H., Prieto, I., Wurzbürger, N., Zadworny, M.,
2070 Bagniewska-Zadworna, A., ... Stokes, A. (2021). Root traits as drivers of plant and
2071 ecosystem functioning: current understanding, pitfalls and future research needs. *New*
2072 *Phytologist*, 232(3), 1123–1158. <https://doi.org/10.1111/nph.17072>
- 2073 Freschet, G. T., Valverde-Barrantes, O. J., Tucker, C. M., Craine, J. M., McCormack, M. L.,
2074 Violle, C., Fort, F., Blackwood, C. B., Urban-Mead, K. R., Iversen, C. M., Bonis, A.,
2075 Comas, L. H., Cornelissen, J. H. C., Dong, M., Guo, D., Hobbie, S. E., Holdaway, R. J.,
2076 Kembel, S. W., Makita, N., ... Roumet, C. (2017). Climate, soil and plant functional types
2077 as drivers of global fine-root trait variation. *Journal of Ecology*, 105(5), 1182–1196.
2078 <https://doi.org/10.1111/1365-2745.12769>
- 2079 Fujii, K., Shibata, M., Kitajima, K., Ichie, T., Kitayama, K., & Turner, B. L. (2018). Plant–soil
2080 interactions maintain biodiversity and functions of tropical forest ecosystems. *Ecological*
2081 *Research*, 33(1), 149–160. <https://doi.org/10.1007/s11284-017-1511-y>
- 2082 Gadgil, R. L., & Gadgil, P. D. (1971). Mycorrhiza and litter decomposition. *Nature*, 233(5315),
2083 133. <https://doi.org/10.1038/233133a0>
- 2084 Gomes, L. C., Faria, R. M., Souza, E., Veloso, G. V., Schaefer, C. E. G. R., & Fernandes Filho,
2085 E. I. (2019). Modelling and mapping soil organic carbon stocks in Brazil. *Geoderma*,
2086 340(January), 337–350. <https://doi.org/10.1016/j.geoderma.2019.01.007>
- 2087 Grime, J. P., Thompson, K., Hunt, R., Hodgson, J. G., Cornelissen, J. H. C., Rorison, I. H.,
2088 Hendry, G. A. F., Ashenden, T. W., Askew, A. P., Band, S. R., Booth, R. E., Bossard, C. C.,
2089 Campbell, B. D., Cooper, J. E. L., Davison, A. W., Gupta, P. L., Hall, W., Hand, D. W.,

- 2090 Hannah, M. A., ... Whitehouse, J. (1997). Integrated Screening Validates Primary Axes of
2091 Specialisation in Plants. *Oikos*, 79(2), 259. <https://doi.org/10.2307/3546011>
- 2092 Gruber, B. D., Giehl, R. F. H., Friedel, S., & von Wirén, N. (2013). Plasticity of the Arabidopsis
2093 root system under nutrient deficiencies. *Plant Physiology*, 163(1), 161–179.
2094 <https://doi.org/10.1104/pp.113.218453>
- 2095 Guerrero-Ramírez, N. R., Mommer, L., Freschet, G. T., Iversen, C. M., McCormack, M. L.,
2096 Kattge, J., Poorter, H., van der Plas, F., Bergmann, J., Kuyper, T. W., York, L. M.,
2097 Bruelheide, H., Laughlin, D. C., Meier, I. C., Roumet, C., Semchenko, M., Sweeney, C.
2098 J., van Ruijven, J., Valverde-Barrantes, O. J., ... Weigelt, A. (2021). Global root traits
2099 (GRoot) database. *Global Ecology and Biogeography*, 30(1), 25–37.
2100 <https://doi.org/10.1111/geb.13179>
- 2101 Hanson, P. J., Edwards, N. T., Garten, C. T., & Andrews, J. A. (2000). Separating root and soil
2102 microbial contributions to soil respiration: A review of methods and observations.
2103 *Biogeochemistry*, 48(1), 115–146. <https://doi.org/10.1023/A:1006244819642>
- 2104 Hawker, L., Uhe, P., Paulo, L., Sosa, J., Savage, J., Sampson, C., & Neal, J. (2022). A 30 m
2105 global map of elevation with forests and buildings removed. *Environmental Research
2106 Letters*, 17(2), 024016. <https://doi.org/10.1088/1748-9326/ac4d4f>
- 2107 Hengl, T., De Jesus, J. M., Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A.,
2108 Shanguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A.,
2109 Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I.,
2110 Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based
2111 on machine learning. *PLoS ONE*, 12(2), e0169748.
2112 <https://doi.org/10.1371/journal.pone.0169748>
- 2113 Hengl, T., Heuvelink, G. B. M., Kempen, B., Leenaars, J. G. B., Walsh, M. G., Shepherd, K.
2114 D., Sila, A., MacMillan, R. A., De Jesus, J. M., Tamene, L., & Tondoh, J. E. (2015).

- 2115 Mapping soil properties of Africa at 250 m resolution: Random forests significantly
2116 improve current predictions. *PLoS ONE*, *10*(6), e0125814.
2117 <https://doi.org/10.1371/journal.pone.0125814>
- 2118 Hijmans, R. J. (2023). *terra: Spatial Data Analysis* (1.7-39). <https://rspatial.org/terra/>
- 2119 Hou, F., Hinojosa, L., Enderle, E., Jansen, B., Morriën, E., & de Vries, F. T. (2025). Root
2120 exudates from drought-affected plants increase soil respiration across a range of grassland
2121 species. *Soil Biology and Biochemistry*, *203*, 109731.
2122 <https://doi.org/10.1016/j.soilbio.2025.109731>
- 2123 Ito, A., & Wagai, R. (2017). Global distribution of clay-size minerals on land surface for
2124 biogeochemical and climatological studies. *Scientific Data*, *4*(1), 170103.
2125 <https://doi.org/10.1038/sdata.2017.103>
- 2126 Iversen, C. M. (2014). Using root form to improve our understanding of root function. *New*
2127 *Phytologist*, *203*(3), 707–709. <https://doi.org/10.1111/nph.12902>
- 2128 Jian, J., Frissell, M., Hao, D., Tang, X., Berryman, E., & Bond-Lamberty, B. (2022). The global
2129 contribution of roots to total soil respiration. *Global Ecology and Biogeography*, *31*(4),
2130 685–699. <https://doi.org/10.1111/geb.13454>
- 2131 Jones, D. L., Hodge, A., & Kuzyakov, Y. (2004). Plant and mycorrhizal regulation of
2132 rhizodeposition. *New Phytologist*, *163*(3), 459–480. <https://doi.org/10.1111/j.1469-8137.2004.01130.x>
- 2134 Joswig, J. S., Wirth, C., Schuman, M. C., Kattge, J., Reu, B., Wright, I. J., Sippel, S. D., Rüger,
2135 N., Richter, R., Schaepman, M. E., van Bodegom, P. M., Cornelissen, J. H. C., Díaz, S.,
2136 Hattingh, W. N., Kramer, K., Lens, F., Niinemets, Ü., Reich, P. B., Reichstein, M., ...
2137 Mahecha, M. D. (2022). Climatic and soil factors explain the two-dimensional spectrum
2138 of global plant trait variation. *Nature Ecology and Evolution*, *6*(1), 36–50.
2139 <https://doi.org/10.1038/s41559-021-01616-8>

- 2140 Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann,
2141 N. E., Linder, H. P., & Kessler, M. (2017). Climatologies at high resolution for the earth's
2142 land surface areas. *Scientific Data*, 4(1), 170122. <https://doi.org/10.1038/sdata.2017.122>
- 2143 Kenrick, P., & Crane, P. R. (1997). The origin and early evolution of plants on land. *Nature*,
2144 389(6646), 33–39. <https://doi.org/10.1038/37918>
- 2145 Kier, G., Mutke, J., Dinerstein, E., Ricketts, T. H., Küper, W., Kreft, H., & Barthlott, W. (2005).
2146 Global patterns of plant diversity and floristic knowledge. *Journal of Biogeography*, 32(7),
2147 1107–1116. <https://doi.org/10.1111/j.1365-2699.2005.01272.x>
- 2148 Knorr, W., Prentice, I. C., House, J. I., & Holland, E. A. (2005). Long-term sensitivity of soil
2149 carbon turnover to warming. *Nature* 2004 433:7023, 433(7023), 298–301.
2150 <https://doi.org/10.1038/nature03226>
- 2151 Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical*
2152 *Software*, 28(5), 1–26. <https://doi.org/10.18637/jss.v028.i05>
- 2153 Kuzyakov, Y., & Larionova, A. A. (2005). Root and rhizomicrobial respiration: A review of
2154 approaches to estimate respiration by autotrophic and heterotrophic organisms in soil.
2155 *Journal of Plant Nutrition and Soil Science*, 168(4), 503–520.
2156 <https://doi.org/10.1002/jpln.200421703>
- 2157 Lak, Z. A., Sandén, H., Mayer, M., & Rewald, B. (2020). Specific root respiration of three plant
2158 species as influenced by storage time and conditions. *Plant and Soil*, 453(1–2), 615–626.
2159 <https://doi.org/10.1007/S11104-020-04619-9/TABLES/3>
- 2160 Lambers, H., Raven, J. A., Shaver, G. R., & Smith, S. E. (2008). Plant nutrient-acquisition
2161 strategies change with soil age. *Trends in Ecology and Evolution*, 23(2), 95–103.
2162 <https://doi.org/10.1016/j.tree.2007.10.008>
- 2163 Lavorel, S., Díaz, S., Cornelissen, J. H. C., Garnier, E., Harrison, S. P., McIntyre, S., Pausas, J.
2164 G., Pérez-Harguindeguy, N., Roumet, C., & Urcelay, C. (2007). Plant Functional Types:

- 2165 Are We Getting Any Closer to the Holy Grail? *Terrestrial Ecosystems in a Changing*
2166 *World*, 149–164. https://doi.org/10.1007/978-3-540-32730-1_13
- 2167 Lee, M. Y., & Andersen, K. M. (2023). Global root trait underrepresentation in Paleotropical
2168 communities: a qualitative analysis. *Frontiers in Forests and Global Change*, 6, 1206225.
2169 <https://doi.org/10.3389/ffgc.2023.1206225>
- 2170 Lehmann, C. E. R., Anderson, T. M., Sankaran, M., Higgins, S. I., Archibald, S., Hoffmann, W.
2171 A., Hanan, N. P., Williams, R. J., Fensham, R. J., Felfili, J., Hutley, L. B., Ratnam, J., San
2172 Jose, J., Montes, R., Franklin, D., Russell-Smith, J., Ryan, C. M., Durigan, G., Hiernaux,
2173 P., ... Bond, W. J. (2014). Savanna vegetation-fire-climate relationships differ among
2174 continents. *Science*, 343(6170), 548–552. <https://doi.org/10.1126/science.1247355>
- 2175 Li, H., Siri, M., Wang, B., He, Y., Liu, C., Feng, C., & Liu, K. (2022). Global Root Traits
2176 Research during 2000–2021: A Bibliometric Analysis. *Agronomy*, 12(10), 2471.
2177 <https://doi.org/10.3390/agronomy12102471>
- 2178 Liu, Y., Liu, R., Qi, L., Chen, J., Dong, J., & Wei, X. (2024). Global mapping of fractional tree
2179 cover for forest cover change analysis. *ISPRS Journal of Photogrammetry and Remote*
2180 *Sensing*, 211, 67–82. <https://doi.org/10.1016/j.isprsjprs.2024.03.019>
- 2181 Ma, Z., Guo, D., Xu, X., Lu, M., Bardgett, R. D., Eissenstat, D. M., McCormack, M. L., &
2182 Hedin, L. O. (2018). Evolutionary history resolves global organization of root functional
2183 traits. *Nature*, 555(7694), 94–97. <https://doi.org/10.1038/nature25783>
- 2184 Martin, C. A., & Stutz, J. C. (2004). Interactive effects of temperature and arbuscular
2185 mycorrhizal fungi on growth, P uptake and root respiration of *Capsicum annuum* L.
2186 *Mycorrhiza*, 14(4), 241–244. <https://doi.org/10.1007/s00572-003-0261-6>
- 2187 McBratney, A. B., Mendonça Santos, M. L., & Minasny, B. (2003). On digital soil mapping.
2188 *Geoderma*, 117(1–2), 3–52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)

- 2189 McCormack, M. L., Dickie, I. A., Eissenstat, D. M., Fahey, T. J., Fernandez, C. W., Guo, D.,
2190 Helmisaari, H. S., Hobbie, E. A., Iversen, C. M., Jackson, R. B., Leppälammil-Kujansuu,
2191 J., Norby, R. J., Phillips, R. P., Pregitzer, K. S., Pritchard, S. G., Rewald, B., & Zadworny,
2192 M. (2015). Redefining fine roots improves understanding of below-ground contributions
2193 to terrestrial biosphere processes. *New Phytologist*, 207(3), 505–518.
2194 <https://doi.org/10.1111/nph.13363>
- 2195 Meinshausen, N. (2017). *quantregForest: Quantile Regression Forests* (1.3-7). [https://cran.r-](https://cran.r-project.org/package=quantregForest)
2196 [project.org/package=quantregForest](https://cran.r-project.org/package=quantregForest)
- 2197 Mello, D. C., Francelino, M. R., Moquedace, C. M., Baldi, C. G. O., Silva, L. V., Siqueira, R.
2198 G., Veloso, G. V., Fernandes-Filho, E. I., Thomazini, A., Demattê, J. A. M., Ferreira, T. O.,
2199 Gomes, L. C., Senra, E. O., & Schaefer, C. E. G. R. (2025). Global warming may turn ice-
2200 free areas of Maritime and Peninsular Antarctica into potential soil organic carbon sinks.
2201 *Communications Earth and Environment*, 6(1), 1–14. [https://doi.org/10.1038/s43247-024-](https://doi.org/10.1038/s43247-024-01937-z)
2202 [01937-z](https://doi.org/10.1038/s43247-024-01937-z)
- 2203 Melton, J. R., Chan, E., Millard, K., Fortier, M., Winton, R. S., Martín-López, J. M., Cadillo-
2204 Quiroz, H., Kidd, D., & Verchot, L. V. (2022). A map of global peatland extent created
2205 using machine learning (Peat-ML). *Geoscientific Model Development*, 15(12), 4709–
2206 4738. <https://doi.org/10.5194/gmd-15-4709-2022>
- 2207 Meyer, H., Milà, C., Ludwig, M., Linnenbrink, J., & Schumacher, F. (2025). *CAST: “caret”*
2208 *Applications for Spatial-Temporal Models* (1.0.3).
- 2209 Meyer, H., & Pebesma, E. (2021). Predicting into unknown space? Estimating the area of
2210 applicability of spatial prediction models. *Methods in Ecology and Evolution*, 12(9), 1620–
2211 1633. <https://doi.org/10.1111/2041-210X.13650>
- 2212 Microsoft Corporation, & Weston, S. (2022). *doParallel: Foreach Parallel Adaptor for the*
2213 *“parallel” Package* (1.0.17). <https://cran.r-project.org/package=doParallel>

- 2214 Moquedace, C. M., Baldi, C. G. O., Siqueira, R. G., Cardoso, I. M., Souza, E. F. M. de, Fontes,
2215 R. L. F., Francelino, M. R., Gomes, L. C., & Fernandes-Filho, E. I. (2024). High-resolution
2216 mapping of soil carbon stocks in the western Amazon. *Geoderma Regional*, *36*, e00773.
2217 <https://doi.org/10.1016/j.geodrs.2024.e00773>
- 2218 Moreno-Martínez, Á., Camps-Valls, G., Kattge, J., Robinson, N., Reichstein, M., van Bodegom,
2219 P., Kramer, K., Cornelissen, J. H. C., Reich, P., Bahn, M., Niinemets, Ü., Peñuelas, J.,
2220 Craine, J. M., Cerabolini, B. E. L., Minden, V., Laughlin, D. C., Sack, L., Allred, B.,
2221 Baraloto, C., ... Running, S. W. (2018). A methodology to derive global maps of leaf traits
2222 using remote sensing and climate data. *Remote Sensing of Environment*, *218*, 69–88.
2223 <https://doi.org/10.1016/J.RSE.2018.09.006>
- 2224 Muñoz Sabater, J. (2024). ERA5-Land hourly data from 1950 to present. Em *Copernicus*
2225 *Climate Change Service (C3S) Climate Data Store (CDS)* (p. 1–10). ECMWF.
2226 <https://doi.org/10.24381/cds.e2161bac>
- 2227 Myneni, R., Knyazikhin, Y., & Park, T. (2021). MOD15A2H MODIS/Terra Leaf Area
2228 Index/FPAR 8-Day L4 Global 500m SIN Grid V061. Em *NASA EOSDIS Land Processes*
2229 *DAAC*. <https://doi.org/10.5067/MODIS/MCD15A3H.061>
- 2230 Oades, J. M. (1988). The retention of organic matter in soils. *Biogeochemistry*, *5*(1), 35–70.
2231 <https://doi.org/10.1007/BF02180317>
- 2232 Öpik, M., Moora, M., Liira, J., & Zobel, M. (2006). Composition of root-colonizing arbuscular
2233 mycorrhizal fungal communities in different ecosystems around the globe. *Journal of*
2234 *Ecology*, *94*(4), 778–790. <https://doi.org/10.1111/J.1365-2745.2006.01136.X>
- 2235 Orchard, V. A., & Cook, F. J. (1983). Relationship between soil respiration and soil moisture.
2236 *Soil Biology and Biochemistry*, *15*(4), 447–453. [https://doi.org/10.1016/0038-](https://doi.org/10.1016/0038-0717(83)90010-X)
2237 [0717\(83\)90010-X](https://doi.org/10.1016/0038-0717(83)90010-X)

- 2238 Ostle, N. J., Smith, P., Fisher, R., Ian Woodward, F., Fisher, J. B., Smith, J. U., Galbraith, D.,
2239 Levy, P., Meir, P., McNamara, N. P., & Bardgett, R. D. (2009). Integrating plant-soil
2240 interactions into global carbon cycle models. *Journal of Ecology*, *97*(5), 851–863.
2241 <https://doi.org/10.1111/j.1365-2745.2009.01547.x>
- 2242 Pan, Y., Birdsey, R. A., Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. A., Phillips, O. L.,
2243 Shvidenko, A., Lewis, S. L., Canadell, J. G., Ciais, P., Jackson, R. B., Pacala, S. W.,
2244 McGuire, A. D., Piao, S., Rautiainen, A., Sitch, S., & Hayes, D. (2011). A large and
2245 persistent carbon sink in the world's forests. *Science*, *333*(6045), 988–993.
2246 <https://doi.org/10.1126/science.1201609>
- 2247 Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The*
2248 *R Journal*, *10*(1), 439. <https://doi.org/10.32614/RJ-2018-009>
- 2249 Phillips, R. P., & Fahey, T. J. (2007). Fertilization effects on fineroot biomass, rhizosphere
2250 microbes and respiratory fluxes in hardwood forest soils. *New Phytologist*, *176*(3), 655–
2251 664. <https://doi.org/10.1111/j.1469-8137.2007.02204.x>
- 2252 Poggio, L., De Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., &
2253 Rossiter, D. (2021). SoilGrids 2.0: Producing soil information for the globe with quantified
2254 spatial uncertainty. *Soil*, *7*(1), 217–240. <https://doi.org/10.5194/soil-7-217-2021>
- 2255 Pregitzer, K. S., DeForest, J. L., Burton, A. J., Allen, M. F., Ruess, R. W., & Hendrick, R. L.
2256 (2002). Fine root architecture of nine North American trees. *Ecological Monographs*,
2257 *72*(2), 293–309. [https://doi.org/10.1890/0012-9615\(2002\)072\[0293:FRAONN\]2.0.CO;2](https://doi.org/10.1890/0012-9615(2002)072[0293:FRAONN]2.0.CO;2)
- 2258 Quesada, C. A., Lloyd, J., Schwarz, M., Patiño, S., Baker, T. R., Czimeczik, C., Fyllas, N. M.,
2259 Martinelli, L., Nardoto, G. B., Schmerler, J., Santos, A. J. B., Hodnett, M. G., Herrera, R.,
2260 Luizão, F. J., Arneith, A., Lloyd, G., Dezzeo, N., Hilke, I., Kuhlmann, I., ... Paiva, R.
2261 (2010). Variations in chemical and physical properties of Amazon forest soils in relation

- 2262 to their genesis. *Biogeosciences*, 7(5), 1515–1541. <https://doi.org/10.5194/bg-7-1515->
2263 2010
- 2264 Quesada, C. A., Phillips, O. L., Schwarz, M., Czimczik, C. I., Baker, T. R., Patiño, S., Fyllas,
2265 N. M., Hodnett, M. G., Herrera, R., Almeida, S., Alvarez Dávila, E., Arneeth, A., Arroyo,
2266 L., Chao, K. J., Dezzeo, N., Erwin, T., Di Fiore, A., Higuchi, N., Honorio Coronado, E.,
2267 ... Lloyd, J. (2012). Basin-wide variations in Amazon forest structure and function are
2268 mediated by both soils and climate. *Biogeosciences*, 9(6), 2203–2246.
2269 <https://doi.org/10.5194/bg-9-2203-2012>
- 2270 Raich, J. W., & Schlesinger, W. H. (1992). The global carbon dioxide flux in soil respiration
2271 and its relationship to vegetation and climate. *Tellus B*, 44(2), 81–99.
2272 <https://doi.org/10.1034/j.1600-0889.1992.t01-1-00001.x>
- 2273 Read, D. J. (1991). Mycorrhizas in ecosystems. *Experientia*, 47(4), 376–391.
2274 <https://doi.org/10.1007/BF01972080>
- 2275 Reich, P. B., Tjoelker, M. G., Pregitzer, K. S., Wright, I. J., Oleksyn, J., & Machado, J. L. (2008).
2276 Scaling of respiration to nitrogen in leaves, stems and roots of higher land plants. *Ecology*
2277 *Letters*, 11(8), 793–801. <https://doi.org/10.1111/j.1461-0248.2008.01185.x>
- 2278 Roscher, C., Schumacher, J., Gubsch, M., Lipowsky, A., Weigelt, A., Buchmann, N., Schmid,
2279 B., & Schulze, E. D. (2012). Using plant functional traits to explain diversity-productivity
2280 relationships. *PLoS ONE*, 7(5), e36760. <https://doi.org/10.1371/journal.pone.0036760>
- 2281 Running, S., Mu, Q., & Zhao, M. (2021). MODIS/Terra Gross Primary Productivity 8-Day L4
2282 Global 500m SIN Grid V061, NASA EOSDIS Land Processes DAAC [data set]. Em *NASA*
2283 *EOSDIS Land Processes Distributed Active Archive Center*.
2284 <https://doi.org/10.5067/MODIS/MOD17A2H.061>

- 2285 Rustad, L. E., Huntington, T. G., & Boone, R. D. (2000). Controls on soil respiration:
2286 Implications for climate change. *Biogeochemistry*, 48(1), 1–6.
2287 <https://doi.org/10.1023/A:1006255431298>
- 2288 Ryan, C. M., Williams, M., & Grace, J. (2011). Above- and belowground carbon stocks in a
2289 miombo woodland landscape of mozambique. *Biotropica*, 43(4), 423–432.
2290 <https://doi.org/10.1111/j.1744-7429.2010.00713.x>
- 2291 Sáez-Sandino, T., García-Palacios, P., Maestre, F. T., Plaza, C., Guirado, E., Singh, B. K., Wang,
2292 J., Cano-Díaz, C., Eisenhauer, N., Gallardo, A., & Delgado-Baquerizo, M. (2023). The soil
2293 microbiome governs the response of microbial respiration to warming across the globe.
2294 *Nature Climate Change*, 13(12), 1382–1387. [https://doi.org/10.1038/s41558-023-01868-](https://doi.org/10.1038/s41558-023-01868-1)
2295 1
- 2296 Serna-Chavez, H. M., Fierer, N., & Van Bodegom, P. M. (2013). Global drivers and patterns of
2297 microbial abundance in soil. *Global Ecology and Biogeography*, 22(10), 1162–1172.
2298 <https://doi.org/10.1111/geb.12070>
- 2299 Singh, B. (2023). Primary minerals. *Encyclopedia of Soils in the Environment, Second Edition*,
2300 V4-91-V4-108. <https://doi.org/10.1016/B978-0-12-822974-3.00132-4>
- 2301 Siqueira, R. G., Moquedace, C. M., Fernandes-Filho, E. I., Schaefer, C. E. G. R., Francelino,
2302 M. R., Sacramento, I. F., & Michel, R. F. M. (2024). Modelling and prediction of major
2303 soil chemical properties with Random Forest: Machine learning as tool to understand soil-
2304 environment relationships in Antarctica. *Catena*, 235, 107677.
2305 <https://doi.org/10.1016/j.catena.2023.107677>
- 2306 Siqueira, R. G., Moquedace, C. M., Francelino, M. R., Schaefer, C. E. G. R., & Fernandes-
2307 Filho, E. I. (2023). Machine learning applied for Antarctic soil mapping: Spatial prediction
2308 of soil texture for Maritime Antarctica and Northern Antarctic Peninsula. *Geoderma*, 432,
2309 116405. <https://doi.org/10.1016/j.geoderma.2023.116405>

- 2310 Soudzilovskaia, N. A., Douma, J. C., Akhmetzhanova, A. A., van Bodegom, P. M., Cornwell,
2311 W. K., Moens, E. J., Treseder, K. K., Tibbett, M., Wang, Y. P., & Cornelissen, J. H. C.
2312 (2015). Global patterns of plant root colonization intensity by mycorrhizal fungi explained
2313 by climate and soil chemistry. *Global Ecology and Biogeography*, 24(3), 371–382.
2314 <https://doi.org/10.1111/geb.12272>
- 2315 Sousa, L. M., Poggio, L., & Kempen, B. (2019). Comparison of FOSS4G Supported Equal-
2316 Area Projections Using Discrete Distortion Indicatrices. *ISPRS International Journal of*
2317 *Geo-Information*, 8(8), 351. <https://doi.org/10.3390/ijgi8080351>
- 2318 Steve Running, & Qiaozhen Mu. (2015). *MOD16A2 - MODIS/Terra Net Evapotranspiration*
2319 *8-Day L4 Global 500m SIN Grid*. NASA EOSDIS Land Processes Distributed Active
2320 Archive Center. <https://doi.org/10.5067/MODIS/MOD16A2GF.061>
- 2321 Stuart Chapin, F., Matson, P. A., & Vitousek, P. M. (2012). Principles of terrestrial ecosystem
2322 ecology. *Principles of Terrestrial Ecosystem Ecology*, 1–529. [https://doi.org/10.1007/978-](https://doi.org/10.1007/978-1-4419-9504-9)
2323 [1-4419-9504-9](https://doi.org/10.1007/978-1-4419-9504-9)
- 2324 Taylor, M. K., Lankau, R. A., & Wurzbarger, N. (2016). Mycorrhizal associations of trees have
2325 different indirect effects on organic matter decomposition. *Journal of Ecology*, 104(6),
2326 1576–1584. <https://doi.org/10.1111/1365-2745.12629>
- 2327 Tennekes, M. (2018). tmap: Thematic Maps in R. *Journal of Statistical Software*, 84(6).
2328 <https://doi.org/10.18637/jss.v084.i06>
- 2329 Tolan, J., Yang, H. I., Nosarzewski, B., Couairon, G., Vo, H. V., Brandt, J., Spore, J., Majumdar,
2330 S., Haziza, D., Vamaraju, J., Moutakanni, T., Bojanowski, P., Johns, T., White, B., Tiecke,
2331 T., & Couprie, C. (2024). Very high resolution canopy height maps from RGB imagery
2332 using self-supervised vision transformer and convolutional decoder trained on aerial lidar.
2333 *Remote Sensing of Environment*, 300, 113888. <https://doi.org/10.1016/j.rse.2023.113888>

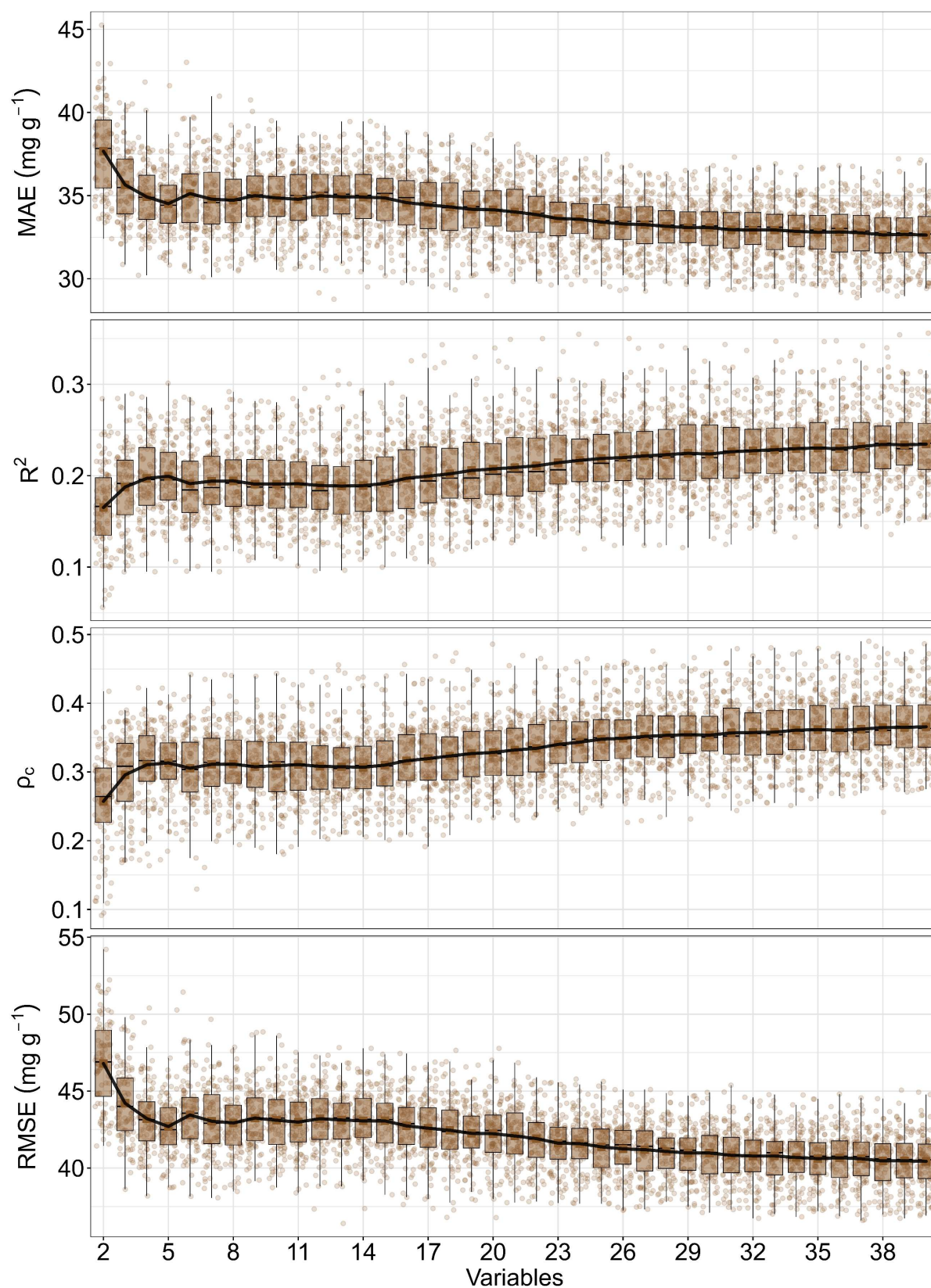
- 2334 Tomlinson, K. W., Sterck, F. J., Bongers, F., da Silva, D. A., Barbosa, E. R. M., Ward, D.,
2335 Bakker, F. T., van Kaauwen, M., Prins, H. H. T., de Bie, S., & van Langevelde, F. (2012).
2336 Biomass partitioning and root morphology of savanna trees across a water gradient.
2337 *Journal of Ecology*, *100*(5), 1113–1121. [https://doi.org/10.1111/j.1365-](https://doi.org/10.1111/j.1365-2745.2012.01975.x)
2338 [2745.2012.01975.x](https://doi.org/10.1111/j.1365-2745.2012.01975.x)
- 2339 Tripathi, B. M., Piñeiro, J., Dang, C., Brzostek, E., & Morrissey, E. M. (2025). Mycorrhiza—
2340 Saprotroph Interactions and Carbon Cycling in the Rhizosphere. *Global Change Biology*,
2341 *31*(4), e70173. <https://doi.org/10.1111/gcb.70173>
- 2342 Trocha, L. K., Mucha, J., Eissenstat, D. M., Reich, P. B., & Oleksyn, J. (2010). Ectomycorrhizal
2343 identity determines respiration and concentrations of nitrogen and non-structural
2344 carbohydrates in root tips: A test using *Pinus sylvestris* and *Quercus robur* saplings. *Tree*
2345 *Physiology*, *30*(5), 648–654. <https://doi.org/10.1093/treephys/tpq014>
- 2346 Tumber-Dávila, S. J., Schenk, H. J., Du, E., & Jackson, R. B. (2022). Plant sizes and shapes
2347 above and belowground and their interactions with climate. *New Phytologist*, *235*(3),
2348 1032–1056. <https://doi.org/10.1111/nph.18031>
- 2349 van den Hoogen, J., Geisen, S., Routh, D., Ferris, H., Traunspurger, W., Wardle, D. A., de
2350 Goede, R. G. M., Adams, B. J., Ahmad, W., Andriuzzi, W. S., Bardgett, R. D., Bonkowski,
2351 M., Campos-Herrera, R., Cares, J. E., Caruso, T., de Brito Caixeta, L., Chen, X., Costa, S.
2352 R., Creamer, R., ... Crowther, T. W. (2019). Soil nematode abundance and functional
2353 group composition at a global scale. *Nature*, *572*(7768), 194–198.
2354 <https://doi.org/10.1038/s41586-019-1418-6>
- 2355 Vermote, E. (2021). MOD09GA, MODIS/Terra Surface Reflectance Daily L2G Global 1km
2356 and 500m Dataset. Em *Nasa Eosdis Lp Daac*.
2357 <https://doi.org/10.5067/MODIS/MOD09GA.061>

- 2358 Violle, C., Navas, M. L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., & Garnier, E. (2007).
2359 Let the concept of trait be functional! *Oikos*, *116*(5), 882–892.
2360 <https://doi.org/10.1111/j.0030-1299.2007.15559.x>
- 2361 Wang, X., Liu, L., Piao, S., Janssens, I. A., Tang, J., Liu, W., Chi, Y., Wang, J., & Xu, S. (2014).
2362 Soil respiration under climate warming: Differential response of heterotrophic and
2363 autotrophic respiration. *Global Change Biology*, *20*(10), 3229–3237.
2364 <https://doi.org/10.1111/gcb.12620>
- 2365 Whalley, W. R., Riseley, B., Leeds-Harrison, P. B., Bird, N. R. A., Leech, P. K., & Adderley, W.
2366 P. (2005). Structural differences between bulk and rhizosphere soil. *European Journal of*
2367 *Soil Science*, *56*(3), 353–360. <https://doi.org/10.1111/j.1365-2389.2004.00670.x>
- 2368 Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Em *Journal of Statistical*
2369 *Software* (Vol. 35, Número Book Review 1). Springer-Verlag New York.
- 2370 Williamson, G. J., Tng, D. Y., & Bowman, D. M. (2024). Climate, fire, and anthropogenic
2371 disturbance determine the current global distribution of tropical forest and savanna.
2372 *Environmental Research Letters*, *19*(2), 024032. [https://doi.org/10.1088/1748-](https://doi.org/10.1088/1748-9326/ad20ac)
2373 [9326/ad20ac](https://doi.org/10.1088/1748-9326/ad20ac)
- 2374 Wolf, S., Mahecha, M. D., Sabatini, F. M., Wirth, C., Bruelheide, H., Kattge, J., Moreno
2375 Martínez, Á., Mora, K., & Kattenborn, T. (2022). Citizen science plant observations
2376 encode global trait patterns. *Nature Ecology and Evolution*, *6*(12), 1850–1859.
2377 <https://doi.org/10.1038/s41559-022-01904-x>
- 2378 Woodward, F. I., & Osborne, C. P. (2000). The representation of root processes in models
2379 addressing the responses of vegetation to global change. *New Phytologist*, *147*(1), 223–
2380 232. <https://doi.org/10.1046/j.1469-8137.2000.00691.x>
- 2381 Yaffar, D., Lugli, L. F., Wong, M. Y., Norby, R. J., Addo-Danso, S. D., Arnaud, M., Cordeiro,
2382 A. L., Dietterich, L. H., Diaz-Toribio, M. H., Lee, M. Y., Ghimire, O. P., Smith-Martin, C.

- 2383 M., Toro, L., Andersen, K., McCulloch, L. A., Meier, I. C., Powers, J. S., Sanchez-Julia,
2384 M., Soper, F. M., & Cusack, D. F. (2024). Tropical root responses to global changes: A
2385 synthesis. *Global Change Biology*, 30(7), e17420. <https://doi.org/10.1111/gcb.17420>
- 2386 Zhang, Y., Schaap, M. G., & Zha, Y. (2018). A High-Resolution Global Map of Soil Hydraulic
2387 Properties Produced by a Hierarchical Parameterization of a Physically Based Water
2388 Retention Model. *Water Resources Research*, 54(12), 9774–9790.
2389 <https://doi.org/10.1029/2018WR023539>
- 2390 Zhao, X., Chen, J. M., Zhang, Y., Jiao, Z., Liu, L., Qiu, F., Zang, J., & Cao, R. (2024). Global
2391 mapping of forest clumping index based on GEDI canopy height and complementary data.
2392 *ISPRS Journal of Photogrammetry and Remote Sensing*, 209, 1–16.
2393 <https://doi.org/10.1016/j.isprsjprs.2024.01.014>

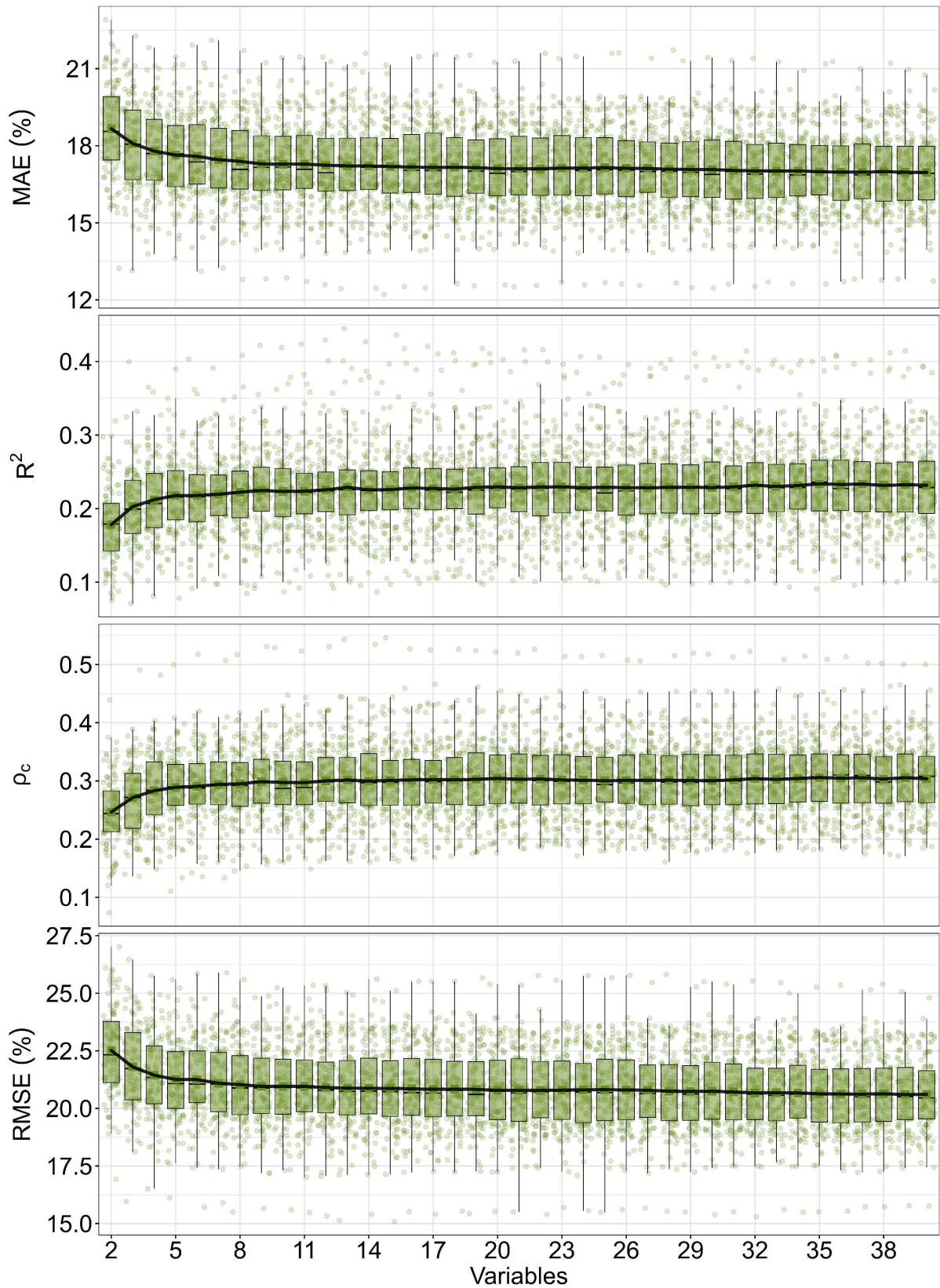
2394

Material suplementar



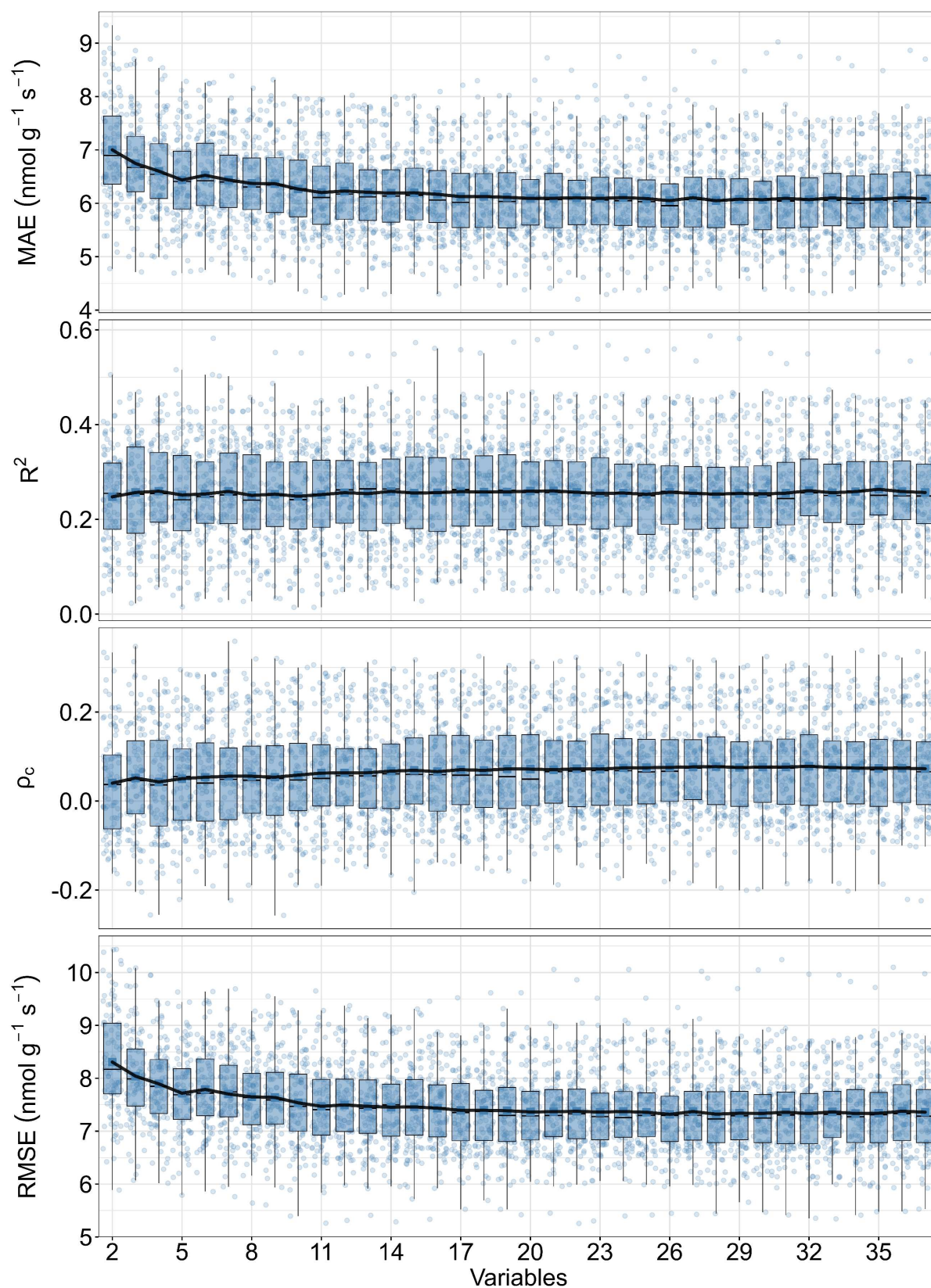
2395
 2396
 2397
 2398
 2399

Figura S1. Desempenho RFE para o quantile random forest nos diferentes subconjuntos de preditores da concentração de carbono nas raízes. ρ_c = coeficiente de correlção de concordância; RMSE = raiz do erro quadrático médio; MAE = erro absoluto médio.



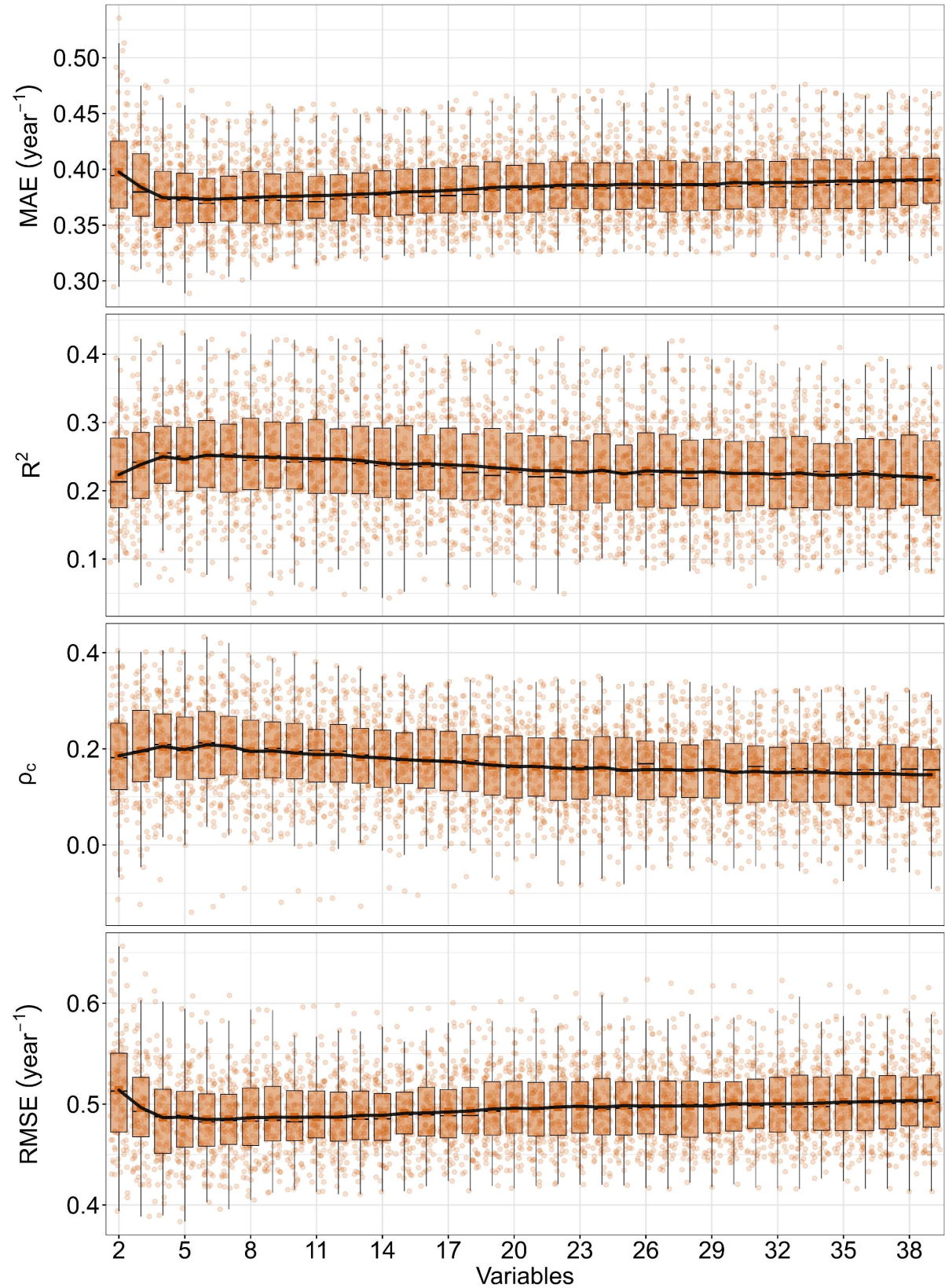
2400
 2401
 2402
 2403
 2404

Figura S2. Desempenho RFE para o quantile random forest nos diferentes subconjuntos de preditores da colonização micorrízica das raízes. ρ_c = coeficiente de correlção de concordância; RMSE = raiz do erro quadrático médio; MAE = erro absoluto médio.



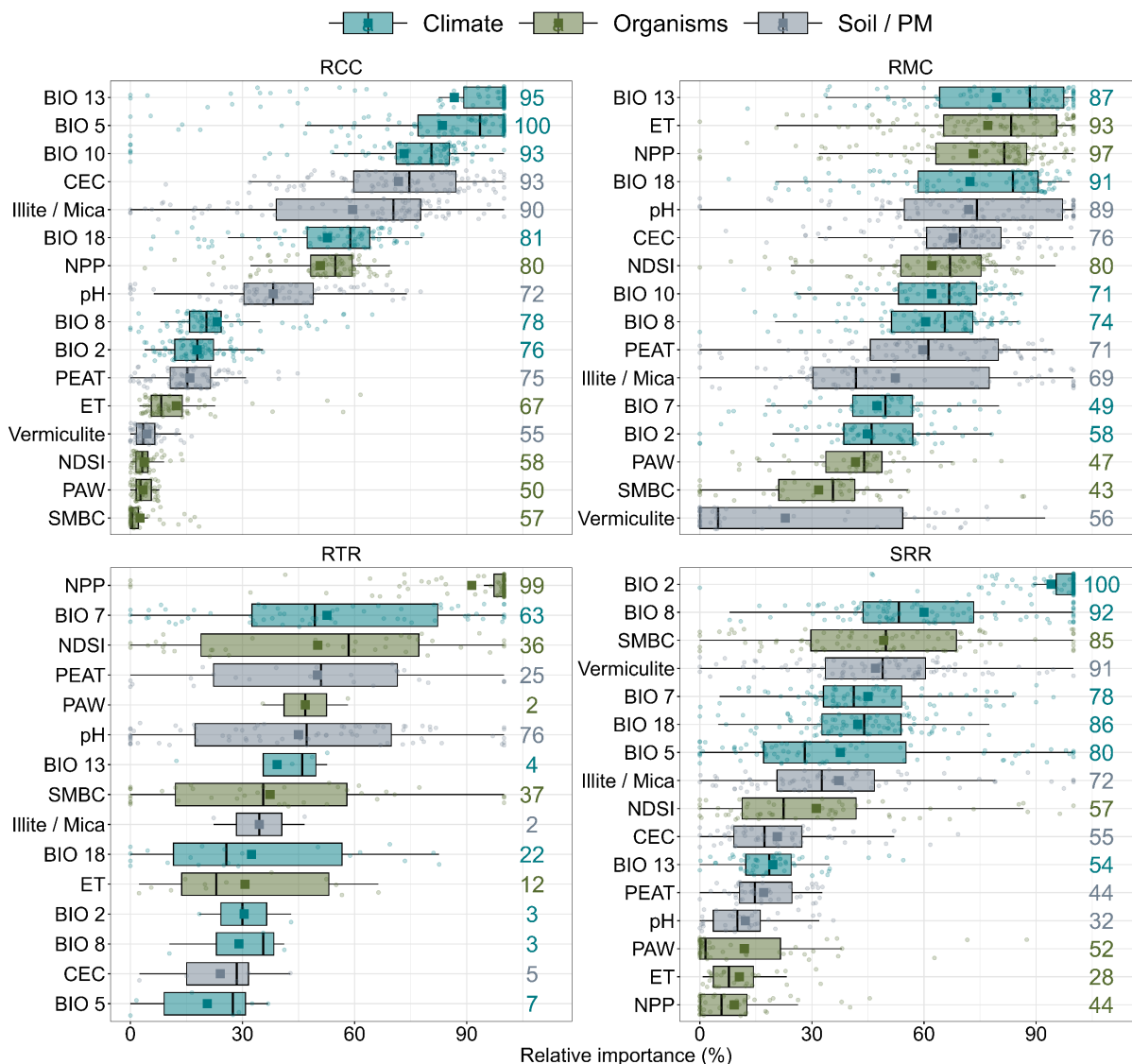
2405
 2406
 2407
 2408
 2409

Figura S3. Desempenho RFE para o quantile random forest nos diferentes subconjuntos de preditores da respiração específica das raízes. ρ_c = coeficiente de correlação de concordância; RMSE = raiz do erro quadrático médio; MAE = erro absoluto médio.



2410
 2411
 2412
 2413
 2414

Figura S4. Desempenho RFE para o quantile random forest nos diferentes subconjuntos de preditores da taxa de rotatividade das raízes. ρ_c = coeficiente de correlção de concordância; RMSE = raiz do erro quadrático médio; MAE = erro absoluto médio.



2415
2416
2417
2418
2419
2420
2421

Figura S5. Importância relativa e frequência de seleção dos preditores mais importantes para o modelo quantile random forest na modelagem da concentração de carbono nas raízes (RCC), colonização micorrízica das raízes (RMC), respiração específica das raízes (SRR) e taxa de rotatividade das raízes (RTR) globalmente. Os números à direita indicam a frequência de seleção do preditor, e os quadrados coloridos representam a média da importância relativa. PM = Material de origem.

2422 **Artigo 3: Aprendizado de máquina e dados pedológicos: perspectivas mais realistas para**
2423 **a modelagem da respiração do solo e feedbacks carbono-clima**

2424

2425 Cássio Marques Moquedace^{1,3}; Lucas Carvalho Gomes^{1,2}; Ivan Francisco Souza³; Lars
2426 Vesterdal⁴; Marcio Rocha Francelino^{1,3}; Carlos Ernesto Gonçalves Reynaud Schaefer^{1,3};
2427 Kendalynn A Morris⁵; Rodrigo Vargas⁶; Ben Bond-Lamberty⁵; Elpídio Inácio Fernandes
2428 Filho^{1,3}

2429

2430 ¹Laboratório de Pedometria e Geoprocessamento - LabGeo, Departamento de Solos,
2431 Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brasil

2432 ²Department of Agroecology, Aarhus University, Tjele, Denmark

2433 ³Departamento de Solos, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brasil

2434 ⁴Department of Geosciences and Natural Resource Management, University of Copenhagen,
2435 Copenhagen, Denmark

2436 ⁵Pacific Northwest National Laboratory, Joint Global Change Research Institute: College Park,
2437 MD, USA

2438 ⁶Department of Plant and Soil Sciences, University of Delaware, Newark, USA

2439

2440 Autor correspondente: Lucas Gomes (lucas.gomes@agro.au.dk)

2441 1. Introdução

2442 Mudanças climáticas futuras são esperadas e vão impactar os estoques de carbono
2443 orgânico do solo (SOC) globalmente, reduzindo-os na maioria das regiões (Crowther et al.,
2444 2016; García-Palacios et al., 2021; Knorr et al., 2005). No entanto, a magnitude do *feedback*
2445 entre carbono do solo e clima permanece amplamente incerta (Bond-Lamberty & Thomson,
2446 2010; Friedlingstein et al., 2006; Guo et al., 2020; Shukla et al., 2007).

2447 Essa incerteza é relevante, pois as emissões de carbono do solo podem constituir um dos
2448 principais impulsionadores das mudanças climáticas globais ao longo do século XXI
2449 (Friedlingstein et al., 2022; Varney et al., 2020). Geralmente, as perdas de SOC são
2450 principalmente impulsionadas pela atividade microbiana (respiração heterotrófica – Rh), que,
2451 combinada com a respiração radicular (respiração autotrófica – Ra), representa o fluxo global
2452 de carbono do solo para a atmosfera (respiração total do solo - Rs) (Beer et al., 2010).

2453 A Rs está fortemente associada à temperatura (Bond-Lamberty & Thomson, 2010), e o
2454 aquecimento do solo pode aumentar as perdas de SOC (Melillo et al., 2017; Soong et al., 2021).
2455 No entanto, as respostas projetadas a longo prazo de Rs e Rh ao aquecimento ainda são incertas
2456 e variáveis entre as ecorregiões globais (Carey et al., 2016; Haaf et al., 2021; Leon et al., 2014).

2457 Entre as principais incertezas na modelagem dos *feedbacks* entre SOC e clima estão os
2458 mecanismos complexos, processos e fatores que controlam as taxas de Rs e Rh em escala local
2459 e global. Geralmente, Rs e Rh resultam da interação entre vegetação, clima, atributos do solo e
2460 microrganismos (Sáez-Sandino et al., 2023; Tang et al., 2020; Yao et al., 2021), todos os quais
2461 podem ser impactados pelas mudanças climáticas.

2462 Embora muitos fatores possam influenciar a respiração do solo, a distribuição espacial
2463 global existente de Rs e Rh é frequentemente baseada apenas em covariáveis de clima e
2464 vegetação (Lu et al., 2021; Warner et al., 2019). Atualmente, os *feedbacks* entre SOC e clima
2465 não consideram satisfatoriamente as covariáveis que informam sobre as condições ambientais

2466 do solo e processos, como a estrutura do solo, em parte devido à sua alta variabilidade espacial
2467 e à disponibilidade restrita de dados.

2468 Um estudo recente combinou processos em escala de poros, temperatura e teor de
2469 umidade em sistemas simulados de espaço poroso de solo homogêneo para projetar tendências
2470 globais em Rh do solo (Nissan et al., 2023). No entanto, o solo é um sistema complexo e
2471 heterogêneo, e para melhorar a compreensão de seu controle sobre a dinâmica do C terrestre,
2472 deve-se reduzir as lacunas de conhecimento nas conexões entre propriedades do solo e
2473 respiração do solo sob restrições específicas impostas por forças externas (por exemplo,
2474 vegetação e clima).

2475 Portanto, pode-se esperar que as covariáveis relacionadas ao solo melhorem a
2476 capacidade de identificar padrões na variação espaço-temporal de Rs global, bem como de Rh
2477 e Ra. As propriedades do solo têm um efeito importante na regulação da persistência e
2478 renovação das reservas de SOC (Doetterl et al., 2015; Luo et al., 2016; Sulman et al., 2014),
2479 com uma relação consistente entre C associado a minerais e conteúdo de argila + silte (Georgiou
2480 et al., 2022).

2481 No entanto, replicar esses efeitos em modelos preditivos ainda precisa ser melhorado. A
2482 textura e a porosidade do solo estão diretamente relacionadas à estrutura do solo e influenciam
2483 o fluxo preferencial de água nos solos (Franklin et al., 2021), bem como a atividade e o acesso
2484 das comunidades microbianas ao SOC (Cable et al., 2008; Patel et al., 2021). As proporções
2485 entre SOC e componentes de textura do solo poderiam, portanto, fornecer insights importantes
2486 sobre a vulnerabilidade do C à mineralização e possíveis mudanças futuras nas taxas de Rs e
2487 Rh.

2488 Por exemplo, partículas de tamanho de areia têm baixa capacidade de interagir com a
2489 matéria orgânica, e menores proporções de SOC / Areia devem se traduzir em maior
2490 vulnerabilidade à mineralização, implicando altas taxas de Rh associadas a solos arenosos.

2491 Estimativas de aumentos nas taxas de Rs até 2100 têm sido baseadas principalmente na relação
2492 entre Rs e a temperatura do ar (Jian et al., 2018). Em contraste, os efeitos associados da
2493 precipitação, umidade do solo e entradas de biomassa em diferentes classes de textura do solo
2494 ainda precisam de atenção. Recentemente, outro estudo também correlacionou um modelo de
2495 Rh baseado em processos e cenários futuros projetados que consideraram apenas mudanças na
2496 precipitação e temperaturas (Nissan et al., 2023).

2497 Assim, considerar a proteção do SOC por mecanismos de estabilização física e química
2498 deve melhorar as projeções das futuras taxas de Rs e Rh sob condições de aquecimento. Embora
2499 a proteção do SOC esteja intrinsecamente relacionada às propriedades do solo, forças externas
2500 como a dinâmica da vegetação e as condições climáticas podem atuar como reguladores desses
2501 processos (Bond-Lamberty et al., 2004, 2018). Isso traz um alto grau de complexidade para as
2502 previsões de mudanças futuras e a análise dos *feedbacks* entre carbono do solo e clima. Por
2503 exemplo, altos níveis atmosféricos de CO₂ e condições de aquecimento podem aumentar a
2504 produção de biomassa e impulsionar a entrada de C nos solos, ao mesmo tempo em que
2505 aumentam as taxas de Rh, o que pode diminuir os SOC (Ruehr et al., 2023).

2506 Por um lado, temperaturas mais altas podem favorecer o crescimento da vegetação e
2507 aumentar as entradas de C no solo mais rapidamente do que as perdas de CO₂ do solo, mesmo
2508 com o aumento nas taxas de respiração do solo. Isto resultaria em um cenário de *feedback*
2509 climático negativo. Alternativamente, condições em que as taxas de Rs e Rh aumentam mais
2510 rapidamente do que as novas entradas de C nos solos a partir de matéria orgânica superficial
2511 e/ou subterrânea constituiriam um *feedback* positivo entre carbono do solo e clima, implicando
2512 perdas adicionais de C do ecossistema (Bond-Lamberty & Thomson, 2010).

2513 *Feedbacks* positivos ou negativos entre carbono do solo e clima são ambos possíveis,
2514 mas podem variar substancialmente entre ecossistemas em todo o mundo, e continua sendo
2515 desafiador prever suas zonas de ocorrência mais prováveis. Curiosamente, todas essas

2516 dinâmicas dependem das propriedades locais do solo e do desenvolvimento futuro do solo que
2517 afetarão a permanência do SOC.

2518 Além disso, uma análise temporal da relação Rh/Rs, considerando as propriedades do
2519 solo e fatores externos (por exemplo, vegetação e clima), deve fornecer importantes *insights*
2520 sobre a direção e magnitude das mudanças na dinâmica do SOC com o aquecimento (Bond-
2521 Lamberty et al., 2004, 2018) e, portanto, sobre o sumidouro de carbono terrestre geral (Ruehr
2522 et al., 2023).

2523 Algoritmos de aprendizado de máquina - ML como por exemplo, Random Forest,
2524 proposto por Breiman (2001) são ferramentas poderosas para identificar relações lineares e não
2525 lineares entre variáveis-resposta e múltiplas variáveis explicativas e têm sido usados para
2526 prever atributos nas áreas de ecologia e geociências em escalas globais (Lu et al., 2021;
2527 Moreno-Martínez et al., 2018; van den Hoogen et al., 2019). Assim, o ML tem sido amplamente
2528 utilizado para projetar a distribuição global passada e atual de Rs e sua relação com o clima e a
2529 cobertura do solo (Lu et al., 2021; Stell et al., 2021; Warner et al., 2019; Zhao et al., 2017). No
2530 entanto, é fundamental considerar que a distribuição espacial das medições de Rs e Rh é
2531 variável no tempo e no espaço (Stell et al., 2021), o que pode influenciar significativamente as
2532 previsões e o desempenho dos modelos.

2533 O banco de dados global mais recente sobre respiração do solo, SRDB-V5, contém cerca
2534 de 10.000 observações de Rs (Jian et al., 2021). Embora o SRDB-V5 ofereça menos dados
2535 sobre Rh do que sobre Rs, algumas tentativas foram realizadas para projetar a distribuição
2536 espacial de Rh em nível global (Hashimoto et al., 2015; Lu et al., 2021; Warner et al., 2019) e
2537 suas dinâmicas espaço-temporais (Tang et al., 2020). Estes autores alcançaram modelos com
2538 coeficientes de determinação variando de 0,6 a 0,8. Embora esses estudos ofereçam *insights*
2539 relevantes sobre a dinâmica global da respiração do solo, os valores reportados parecem
2540 inflacionados, possivelmente em função do método adotado: considerar amostras temporais

2541 com as mesmas coordenadas tanto para o treinamento quanto para a avaliação pode
2542 superestimar a real capacidade de generalização do modelo. Ainda assim, o banco de dados
2543 evidenciou o potencial de abordagens empíricas orientadas por dados.

2544 Objetivou-se com este estudo explorar a influência da textura do solo na respiração do
2545 solo global futura, R_s e R_h , sob cenários de mudanças climáticas até o ano 2100, utilizando três
2546 cenários de Caminhos Socioeconômicos Compartilhados (SSP1-2.6, SSP3-7.0 e SSP5-8.5). A
2547 variável textura do solo foi escolhida devido à sua importância para a resposta da respiração do
2548 solo ao aquecimento (Haaf et al., 2021). Investigou-se o impacto da inclusão ou exclusão de
2549 variáveis de textura do solo, como a quantidade de argila, sobre as estimativas de R_s e R_h .
2550 Assim, os efeitos combinados das alterações climáticas e da textura do solo nas projeções de
2551 R_s e R_h até o ano de 2100 foram analisados.

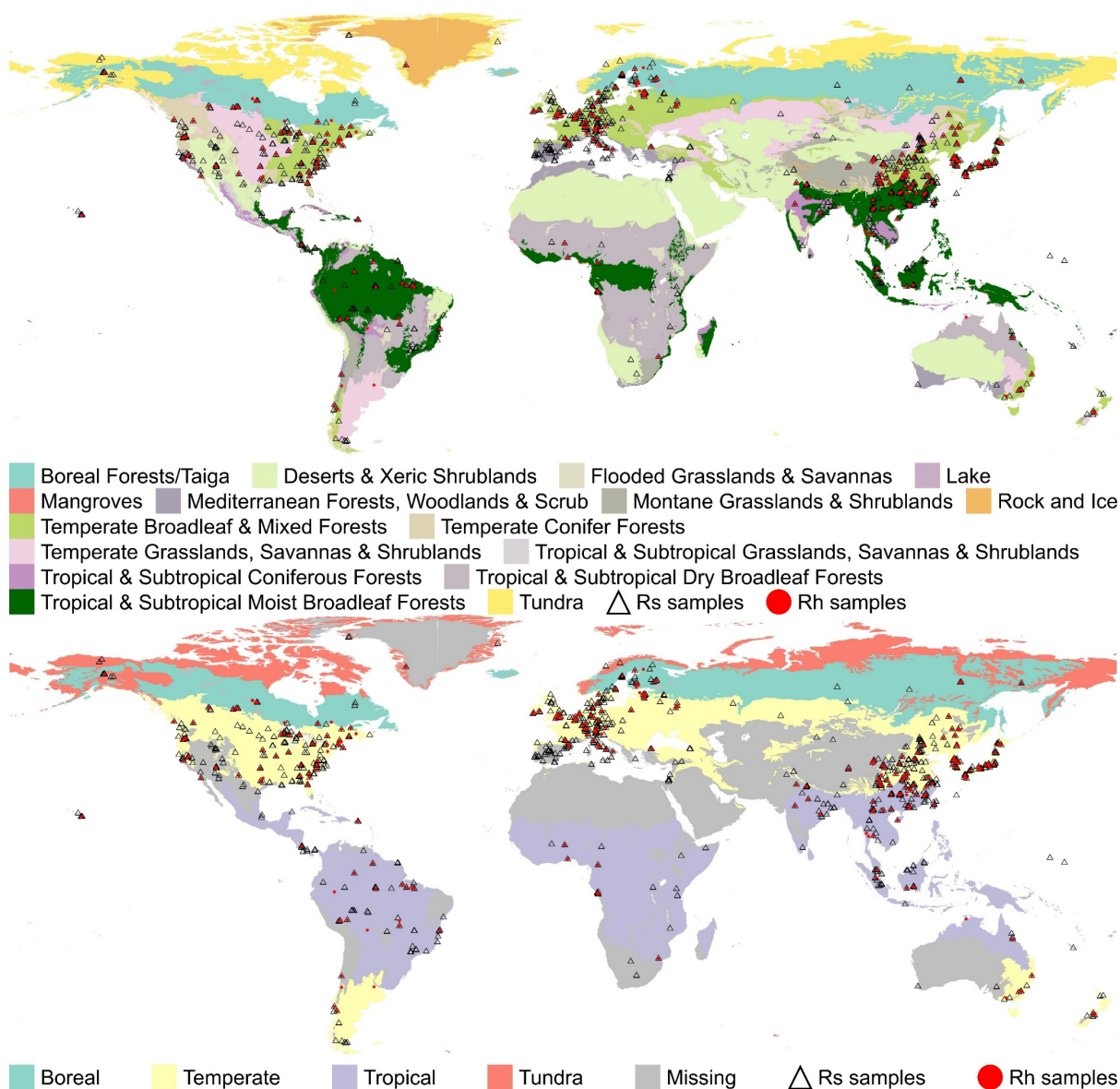
2552

2553 **2. Material e métodos**

2554 **2.1. Dados de respiração do solo e variáveis explicativas**

2555 Para explorar a distribuição espacial de R_s e R_h , utilizou-se o conjunto de dados SRDB-
2556 V5 (Jian et al., 2021). O banco de dados foi filtrado para utilizar apenas medidas não derivadas
2557 de abordagens experimentais, como aquecimento do solo, mudanças na precipitação e áreas
2558 altamente perturbadas. Como resultado, restaram 3.288 observações para R_s e 832 para R_h
2559 (Figura 1).

2560



2561 **Figura 1.** Distribuição das amostras de respiração total do solo (Rs) e respiração heterotrófica
 2562 (Rh) em diferentes ecorregiões terrestres globais (a; Olson et al. (2001)) e um grupo de quatro
 2563 ecorregiões (b; Ecossistemas Tropicais, Ecossistemas Temperados, Tundra e Boreal).
 2564
 2565

2566 As bases de dados foram unificadas, e harmonizadas, compreendendo variáveis
 2567 climáticas (19 bioclimáticas), vegetação (NPP) e solo, incluindo textura (argila e areia), carbono
 2568 orgânico do solo, minerais de argila (ilita, mica, vermiculita, óxidos de Fe), capacidade de troca
 2569 de cátions (CTC) e pH (Tabela 1).

2570

2571

2572 **Tabela 1.** Banco de dados de variáveis explicativas inicial usado para modelar a respiração
 2573 global total do solo e a respiração heterotrófica.

Clima		
BIO 1 - Temperatura média anual	BIO 8 - Temperatura média do trimestre mais chuvoso	BIO 15 - Sazonalidade da Precipitação
BIO 2 - Amplitude média diária	BIO 9 - Temperatura média do trimestre mais seco	BIO 16 - Precipitação do trimestre mais chuvoso
BIO 3 - Isotermalidade	BIO 10 - Temperatura média do trimestre mais quente	BIO 17 - Precipitação do trimestre mais seco
BIO 4 - Sazonalidade da temperatura	BIO 11 - Temperatura média do trimestre mais frio	BIO 18 - Precipitação do trimestre mais quente
BIO 5 - Temperatura Máxima do Mês Mais Quente	BIO 12 - Precipitação Anual	BIO 19 - Precipitação do trimestre mais frio
BIO 6 - Temperatura mínima do mês mais frio	BIO 13 - Precipitação do mês mais chuvoso	-
BIO 7 - Amplitude anual de temperatura	BIO 14 - Precipitação do mês mais seco	-
Organismos		
NPP – Produtividade Primária Líquida		
Solo / MO		
Argila - Concentração de argila	SOC – Concentração de carbono orgânico do solo	SOCS / Argila
Areia - Concentração de areia	SOCD – Densidade de carbono orgânico do solo	SOC / Areia
SOCS – Estoque de carbono orgânico do solo	SOCS / Areia	SOC / Argila
Illite / mica / vermiculita / Óxido Fe	CTC – Capacidade de troca de cátions	pH

2574

2575 Foram utilizadas variáveis de clima e NPP do presente (1981-2010) e de diferentes
 2576 períodos futuros (2011-2040, 2041-2070, 2071-2100), obtidas do banco de dados CHELSA 2.1
 2577 (Karger et al., 2017). As variáveis bioclimáticas e NPP estão disponíveis para três diferentes
 2578 cenários de Caminhos Socioeconômicos Compartilhados (SSP) selecionados (SSP1-2.6, SSP3-
 2579 7.0 e SSP5-8.5), representando trajetórias socioeconômicas globais futuras contrastantes. O
 2580 SSP1 (2.6) prevê um mundo sustentável com baixas emissões e uso de energia limpa; o SSP3
 2581 (7.0) representa um mundo fragmentado com baixa cooperação internacional e emissões

2582 moderadas; e o SSP5 (8.5) retrata uma projeção na qual o mundo é focado no crescimento
2583 econômico rápido com alta dependência de combustíveis fósseis, resultando em altas emissões
2584 de gases de efeito estufa (O'Neill et al., 2014).

2585 Visando descrever melhor a incerteza das projeções, foram utilizados dados climáticos
2586 de cinco Modelos de Sistema Terrestre (GFDL-ESM4, IPSL-CM6A-LR, MPI-ESM1-2-HR,
2587 MRI-ESM2-0, UKESM1-0-LL) do banco de dados CHELSA 2.1. Esses modelos representam
2588 a distinta sensibilidade climática de equilíbrio entre os modelos disponíveis no CMIP6
2589 (Warszawski et al., 2014).

2590 Os tipos de argila do solo foram derivados de Ito & Wagai (2017), selecionando-se os
2591 mapas globais de illita/mica, vermiculita e óxidos de Fe. Os mapas globais de conteúdo e
2592 estoques de carbono orgânico do solo, textura do solo, CTC e pH a 0-30 cm de profundidade
2593 foram derivados do SoilGrids 2.0 (Poggio et al., 2021). Essas variáveis foram selecionadas
2594 devido à importância da distribuição do tamanho das partículas do solo para a respiração do
2595 solo, conteúdo e estoques de carbono (Doetterl et al., 2015).

2596 Os mapas de estoques ou densidade atuais de SOC têm alta correlação com a respiração
2597 do solo e podem ser uma variável-chave. No entanto, não existem dados de SOC disponíveis
2598 para cenários futuros, o que limita o uso desses mapas para prever taxas futuras de respiração
2599 do solo. Foram calculadas novas variáveis para compreender as relações, a partir de razões entre
2600 os mapas. As divisões consistiram nas relações entre a concentração de SOC (dg kg^{-1}) e areia
2601 ou argila (SOC / Areia; SOC / Argila), as relações entre densidade de SOC (hg m^{-3} ; SOCD) e
2602 areia ou argila (SOCD / Areia; SOCD / Argila), e o estoque de SOC (t ha^{-1} ; SOCS) com areia e
2603 argila (SOCS / Areia; SOCS / Argila). SOC / Areia foi considerada como um *proxie* da
2604 vulnerabilidade do carbono do solo à mineralização, enquanto SOC / Argila foi considerada um
2605 indicador do potencial dos solos para proteger o carbono orgânico contra
2606 decomposição/mineralização.

2607 Consequentemente, valores elevados de SOC / Areia ou SOC / Argila podem estar
2608 associados a áreas com alto teor de SOC juntamente com baixo teor de areia ou alto teor de
2609 argila e solos orgânicos em áreas de solos alagados. Assim, as razões SOC / Areia ou SOCS /
2610 Areia poderiam ser altas em solos onde a mineralização do SOC seria prejudicada pela proteção
2611 dos minerais argilosos ou pela falta de oxigênio, respectivamente.

2612

2613 **2.2. Predição e avaliação do modelo**

2614 O modelo quantile random forest - qrf (Meinshausen, 2017) foi utilizado para mapear o
2615 Rs e Rh globais presentes e futuros. A vantagem desse algoritmo é que ele permite a
2616 quantificação de quantis e desvios padrão dos valores médios/medianos preditos. Para isso,
2617 aplicou-se uma abordagem em três etapas. Primeiro, foram mapeados os Rs e Rh presentes
2618 utilizando todas as covariáveis estáticas e dinâmicas.

2619 Antes de obter as predições do modelo, o banco de dados inicial foi filtrado para remover
2620 uma de cada duas variáveis explicativas que apresentavam alta correlação (correlação de
2621 Spearman $> |0,95|$), e foi aplicado o *recursive feature elimination* (RFE) para selecionar apenas
2622 as variáveis mais importantes. Os resultados do RFE são puramente matemáticos, nesse
2623 contexto, aplicou-se um critério de perda máxima de desempenho de 4% (erro absoluto médio
2624 – MAE) (Moquedace et al., 2024; Siqueira et al., 2023, 2024). Esse limite permitiu descartar
2625 variáveis que contribuíam pouco para o desempenho do modelo.

2626 Para fins de comparação do impacto dessas variáveis na distribuição espacial, foram
2627 construídos mapas do presente utilizando tanto as variáveis selecionadas pelo RFE e o critério
2628 de desempenho, quanto todas as covariáveis (dinâmicas e estáticas). Infelizmente, o uso de
2629 todos os mapas de propriedades do solo para prever Rs e Rh futuros foi limitado, pois variáveis
2630 de solo mais dinâmicas, como CTC e pH, não estão disponíveis para cenários futuros. Portanto,
2631 o modelo utilizado para prever dados futuros não incluiu essas variáveis no treinamento.

2632 O processo de seleção de variáveis foi aplicado somente para Rs, já que ele contém mais
2633 amostras, e utilizou-se as mesmas variáveis para a predição de Rh, o que permitiu comparar as
2634 dinâmicas futuras de Rs e Rh. Para predições futuras de Rs e Rh, foi aplicado o modelo
2635 construído para o presente, mas direcionado por covariáveis ambientais projetadas para
2636 cenários futuros. Assim, foram produzidos dados espacialmente explícitos de Rs e Rh futuros
2637 para três períodos (2011-2040, 2041-2070, 2071-2100), sob os três cenários SSP (SSP1-2.6,
2638 SSP3-7.0, SSP5-8.5) e cinco modelos climáticos (GFDL-ESM4, IPSL-CM6A-LR, MPI-
2639 ESM1-2-HR, MRI-ESM2-0, UKESM1-0-LL).

2640 Para examinar a influência das propriedades do solo nas predições de Rs e Rh, todo o
2641 processo foi repetido, incluindo (+Soil) e excluindo variáveis do solo (-Soil) tanto para o
2642 presente quanto para o período de 2071-2100 no cenário SSP5-8.5 e nos cinco modelos
2643 climáticos descritos acima. Essa última etapa permitiu testar a influência das propriedades
2644 pedológicas nas estimativas de Rs e Rh. A análise de dependência parcial também foi aplicada
2645 para identificar o efeito individual de cada variável selecionada nos modelos de Rs e Rh (Warner
2646 et al., 2019).

2647 Para testar o desempenho dos modelos, os dados foram divididos em dois conjuntos:
2648 um para treinamento, com validação cruzada repetida 10 vezes em 10 *folds* (80%), e outro para
2649 teste independente (20%). O processo de modelagem foi conduzido de forma a garantir que
2650 amostras com as mesmas coordenadas, ou seja, dados temporais de respiração do solo em um
2651 mesmo ponto, não se misturassem entre os *folds* (validação cruzada) e entre a separação de
2652 treino e teste. Adicionalmente, o desempenho do modelo foi analisado sem essa condição
2653 aplicada.

2654 Além disso, calculou-se um modelo nulo para comparações com o desempenho geral
2655 do modelo (Moquedace et al., 2024; Siqueira et al., 2023). Para contornar o possível problema
2656 de resultados otimistas ou pessimistas devido à aleatoriedade na separação de treinamento e

2657 teste, o processo foi repetido 100 vezes, variando os conjuntos e gerando assim a média e o
2658 desvio padrão associados às métricas de desempenho: erro absoluto médio (MAE), raiz do erro
2659 quadrático médio (RMSE), coeficiente de determinação (R^2) e coeficiente de correlação
2660 concordante (ρ_c).

2661 Para obter as predições e criar os mapas, um modelo final foi ajustado para Rs e outro
2662 para Rh com todas as amostras para o presente, sendo então aplicado para projetar os cenários
2663 futuros. Dessa forma, utilizou-se a validação cruzada repetida (10 vezes) com 10 *folds* e com
2664 condicionamento espacial, impedindo que locais com mais de uma amostra fossem divididos
2665 entre os *folds* de treinamento. Para as predições de Rs e Rh, foram calculados a média, o
2666 coeficiente de variação (CV), a mediana (quantil 0,50, $q_{0.5}$), o quantil 0,25 ($q_{0.25}$) e o quantil
2667 0,75 ($q_{0.75}$), bem como os limites inferior e superior de um intervalo de predição de 50%.

2668 Complementarmente, a diferença da respiração heterotrófica do solo (ΔRh), definida
2669 como a subtração entre os mapas futuros (2071–2100) e os mapas do presente ($\Delta Rh = \text{futuro} -$
2670 presente), foi estratificada em classes de variação da temperatura média anual ($\Delta BIO1$) e do
2671 teor de argila. Para cada classe, foram calculadas as médias de ΔRh e os respectivos intervalos
2672 de confiança de 99%, estimados por meio do método bootstrap percentílico não-paramétrico.
2673 Esses intervalos foram definidos pelos percentis 0,5% e 99,5% da distribuição empírica das
2674 médias reamostradas.

2675 Toda a computação foi realizada usando R (R Core Team, 2025), com auxílio dos
2676 pacotes caret (Kuhn, 2022), DescTools (Signorell, 2023), doParallel (Microsoft Corporation &
2677 Weston, 2022), exactextractr (Bastin, 2022), ggplot2 (Wickham, 2016), ggpubr (Kassambara,
2678 2023), parallelly (Bengtsson, 2023), quantregForest (Meinshausen, 2017), rfUtilities (Evans &
2679 Murphy, 2015), sf (Pebesma, 2018), terra (Hijmans, 2023), tmap (Tennekes, 2018) e mdsFuncs
2680 (Fernandes Filho et al., 2021). Todo o código analítico está disponível em [url - GitHub].

2681 Tanto os resultados presentes quanto futuros de R_s , R_h e a razão R_h/R_s foram
2682 estratificados nas 14 ecorregiões terrestres globais (Olson et al., 2001), mas a dinâmica temporal
2683 dessas variáveis foi discutida apenas com as ecorregiões agrupadas (Ecorregiões Tropicais,
2684 Ecorregiões Temperados, Tundra e Boreal; Figura 1).

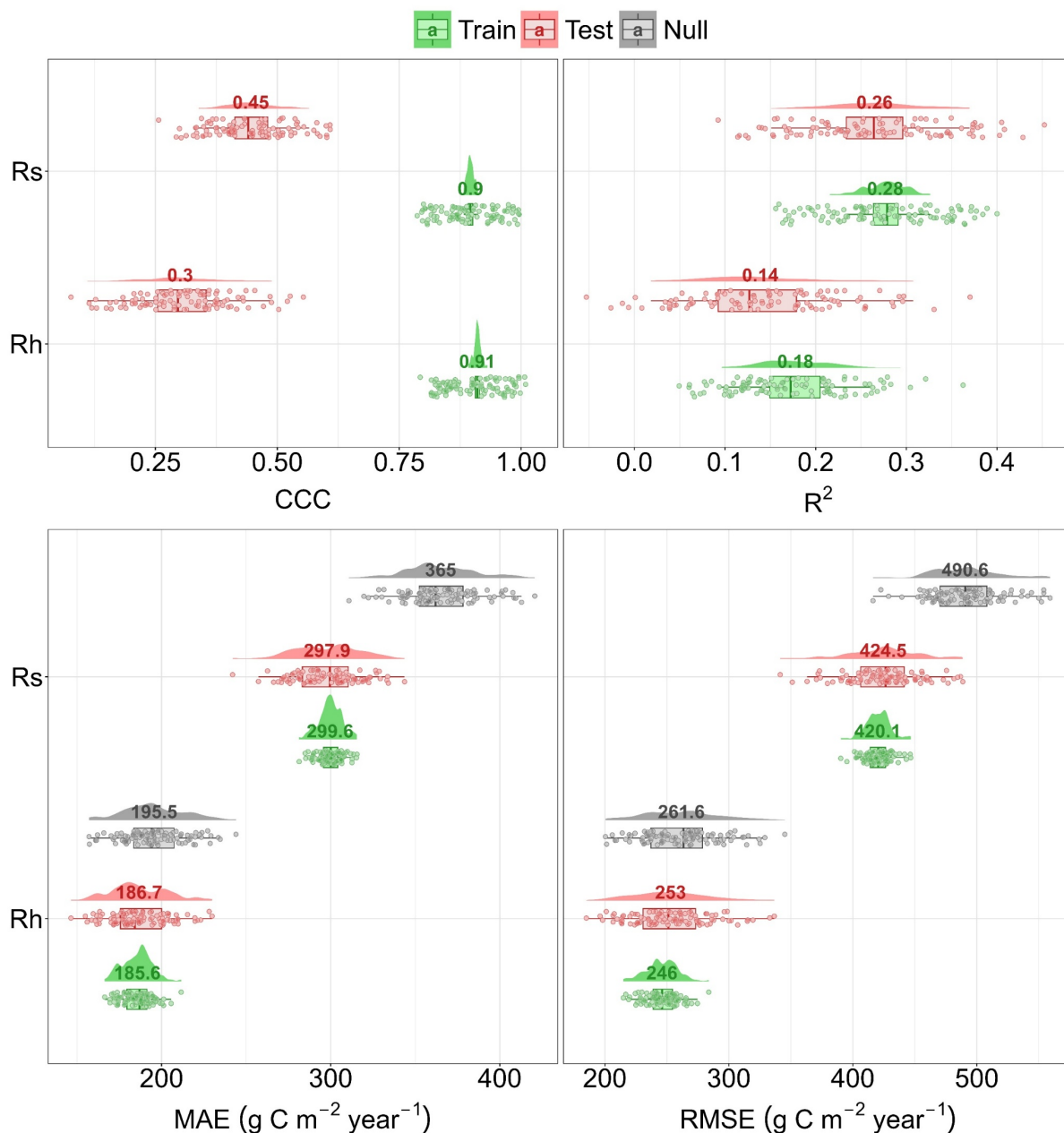
2685

2686 3. Resultados

2687 3.1. Efeitos de overfitting em projeções de R_s e R_h

2688 As projeções com restrições de geolocalização das observações de R_s e R_h no processo
2689 de validação cruzada (seção 2.2) exibiram desempenho relativamente baixo para R_s ($\rho_c = 0,49$;
2690 $R^2 = 0,28$; $RMSE = 420,07 \text{ gC m}^{-2} \text{ ano}^{-1}$; $MAE = 299,7 \text{ gC m}^{-2} \text{ ano}^{-1}$; $RMSE_{\text{nulo}} = 490,6 \text{ gC}$
2691 $\text{m}^{-2} \text{ ano}^{-1}$; $MAE_{\text{nulo}} = 364,9 \text{ gC m}^{-2} \text{ ano}^{-1}$) e R_h ($\rho_c = 0,49$; $R^2 = 0,18$; $RMSE = 246,0 \text{ gC m}^{-2}$
2692 ano^{-1} ; $MAE = 185,6 \text{ gC m}^{-2} \text{ ano}^{-1}$; $RMSE_{\text{nulo}} = 261,6 \text{ gC m}^{-2} \text{ ano}^{-1}$; $MAE_{\text{nulo}} = 195,5 \text{ gC}$
2693 $\text{m}^{-2} \text{ ano}^{-1}$) (Figura 2).

2694



2695
2696
2697
2698
2699
2700

Figura 2. Desempenho do quantile random forest para modelar a respiração global total do solo (Rs) e a respiração heterotrófica (Rh). Esses resultados são de 100 execuções do modelo, nas quais foram realizadas a separação aleatória das amostras em treinamento (80%) e teste (20%). Quadrados coloridos indicam valor médio.

2701
2702
2703
2704

Comparativamente, as projeções com validação cruzada padrão (ignorando a restrição de geolocalização) resultaram em desempenho aparentemente inflado dos modelos na predição de Rs ($\rho_c = 0,90$; $R^2 = 0,65$; $RMSE = 292,0$ gC m⁻² ano⁻¹; $MAE = 190,0$ gC m⁻² ano⁻¹; $RMSE_{nulo} = 490,6$ gC m⁻² ano⁻¹; $MAE_{nulo} = 365,8$ gC m⁻² ano⁻¹) e Rh ($\rho_c = 0,91$; $R^2 = 0,61$;

2705 RMSE = 162,7 gC m⁻² ano⁻¹; MAE = 111,4 gC m⁻² ano⁻¹; RMSE_nulo = 261,5 gC m⁻² ano⁻¹;
2706 MAE_nulo = 193,6 gC m⁻² ano⁻¹).

2707 Estes resultados sugerem que os processos de modelagem anteriores de respiração do
2708 solo, que utilizaram a base de dados do SRDB, tanto a versão 5 quanto as anteriores (Hursh et
2709 al., 2017; Lu et al., 2021; Stell et al., 2021; Tang et al., 2020; Warner et al., 2019), alcançaram
2710 resultados otimistas devido ao vazamento de dados de mesma coordenada, produzindo um
2711 efeito de *overfitting* e mascarando incertezas e erros relativos à capacidade real de generalização
2712 do modelo. Detalhes adicionais sobre o desempenho da modelagem são apresentados na Figura
2713 2.

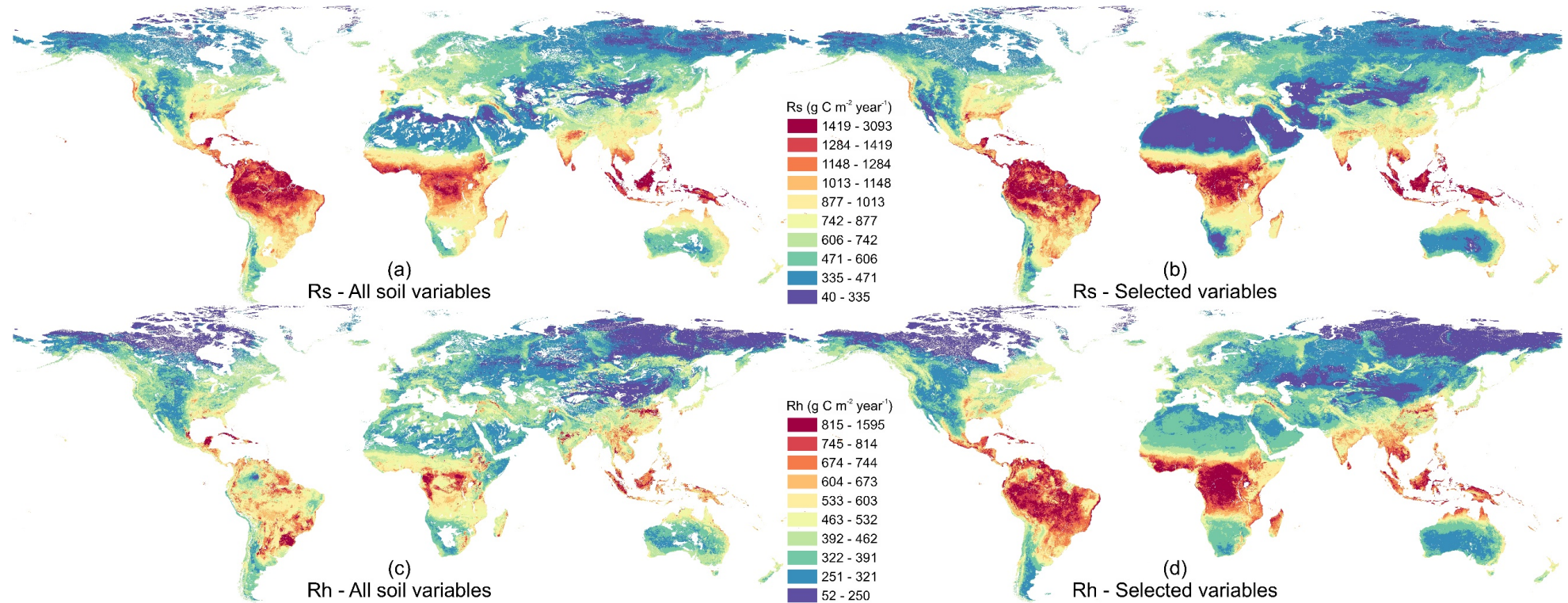
2714 O resultado do RFE apontou 26 variáveis para explicar Rs (Figura S1); no entanto, após
2715 aplicar o critério de aceitação baseado na perda de desempenho, as variáveis mais importantes
2716 selecionadas, em ordem de importância, foram NPP, BIO 12, BIO 4, BIO 3, BIO 6, BIO 13,
2717 BIO 19, BIO 1, Argila e CTC (Tabela 1). Embora os modelos de aprendizado de máquina
2718 possam identificar padrões relevantes, as covariáveis selecionadas podem ser insignificantes se
2719 não forem parâmetros interpretáveis relacionados aos fatores de formação do solo (Wadoux et
2720 al., 2020).

2721 Baseando-se no conhecimento tácito, decidiu-se manter a variável SOC / Areia, expressa
2722 em hectogramas de C/kg de areia (hgC kg⁻¹ areia; 1 hg = 0,1 kg), os tipos de argila e pH que
2723 estavam entre as 26 primeiras variáveis selecionadas pelo RFE. É importante destacar que o
2724 tipo de argila, pH e a relação SOC / Areia não atingiram o limite de 4%. No entanto, é bem
2725 estabelecido que todas essas variáveis estão diretamente ou indiretamente ligadas à proteção
2726 mineral da matéria orgânica do solo contra a decomposição (Dexter et al., 2008; Georgiou et
2727 al., 2022).

2728

2729 **3.2. Rs e Rh presentes e principais fatores**

2730 As previsões de Rs e Rh para o presente são apresentadas na Figura 2, ilustrando o uso
2731 de todas as covariáveis (Figura 2a e Figura 2c) e do modelo com filtro de variáveis projetado
2732 para simular cenários futuros (Figura 2b e Figura 2d). De modo geral, as previsões para Rs e
2733 Rh mostraram semelhanças entre os dois métodos, com taxas mais altas de respiração do solo
2734 nos trópicos e taxas mais baixas em altas latitudes.



2735
2736
2737
2738

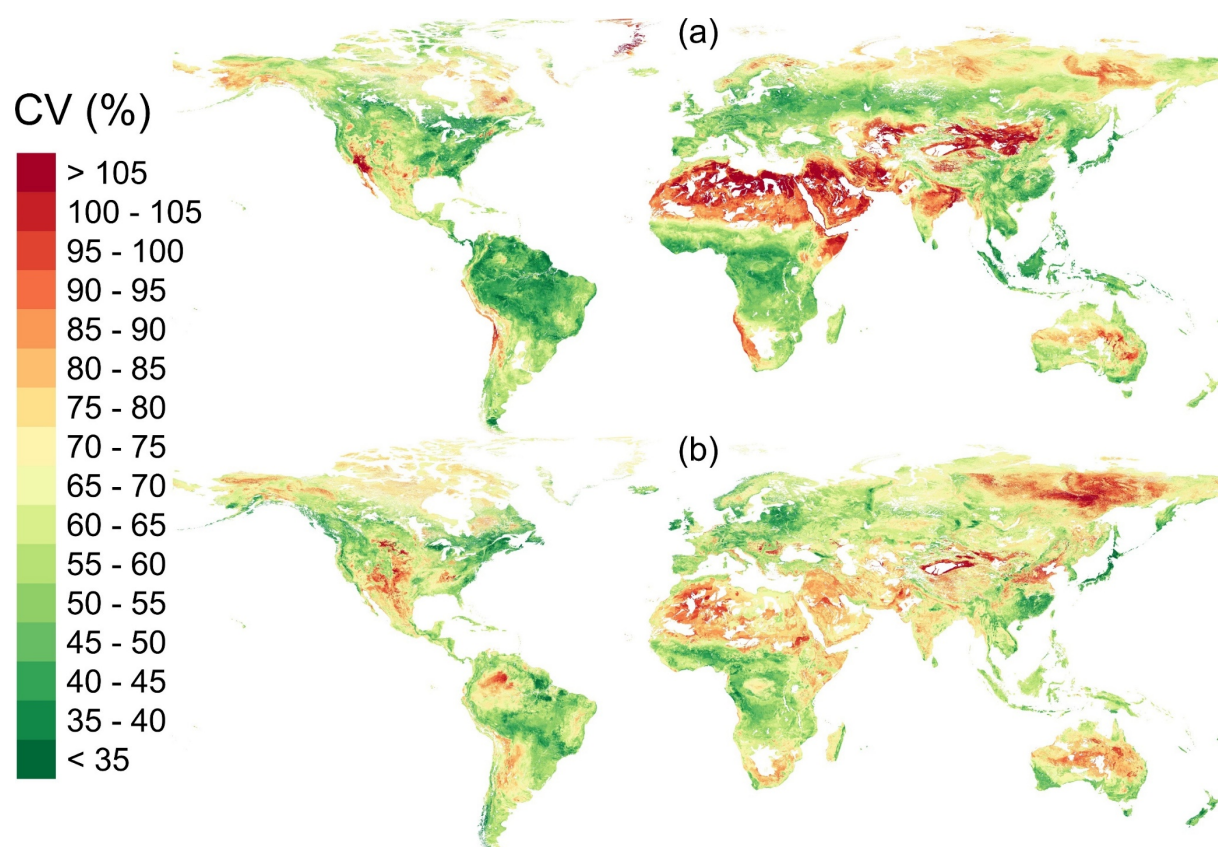
Figura 2. Distribuição global da respiração do solo - Rs (a, b) e da respiração heterotrófica - Rh (c, d) com base em dados históricos/presentes. Os mapas do lado esquerdo (a, c) representam os Rs e Rh usando todas as variáveis de solo disponíveis (26 variáveis) e os mapas do lado direito mostram os Rs e Rh obtidos usando o subconjunto selecionado de covariáveis (8 variáveis).

2739 Os valores específicos para várias ecorregiões estão detalhadas na Tabela S1. Os valores
2740 globais de Rs e Rh para o presente, utilizando todas as variáveis, foram de 106 PgC ano⁻¹ e 60
2741 PgC ano⁻¹, respectivamente. O modelo final resultou em 91,9 PgC ano⁻¹ para Rs (q0.25 = 54,8
2742 e q0.75 = 120,2 PgC ano⁻¹) e 58,0 PgC ano⁻¹ para Rh (q0.25 = 34,8 e q0.75 = 73,5 PgC ano⁻¹).

2743 No entanto, o modelo que incorporou todas as covariáveis estimou valores elevados de
2744 Rs em algumas regiões áridas, que exibiram padrões semelhantes aos das regiões Tropicais.
2745 Além disso, esse mapa mostrou valores mais baixos de Rh e uma distribuição espacial distinta,
2746 apresentando alguns *hotspots* com valores mais altos de Rh. Por exemplo, regiões no sul do
2747 Brasil, África e América Central apresentaram localmente valores mais altos de Rh.

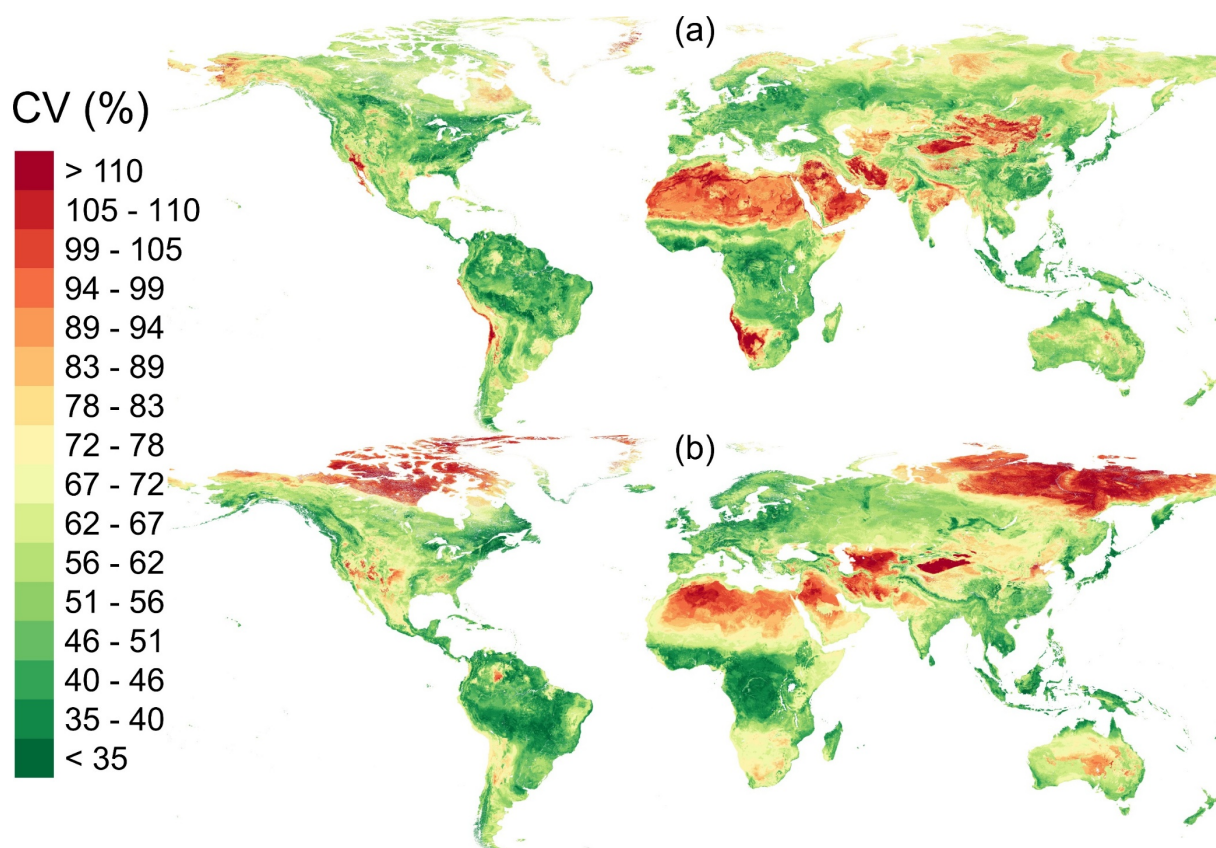
2748 Também foram observados aumentos consideráveis na incerteza para altas latitudes e
2749 regiões áridas em ambos os modelos (Figura 3 e Figura 4). A média global da razão Rh/Rs ficou
2750 em 0,62 para o modelo final, com o valor mais baixo observado na Tundra (0,54) e o mais alto
2751 em desertos (0,87; Tabela S1).

2752



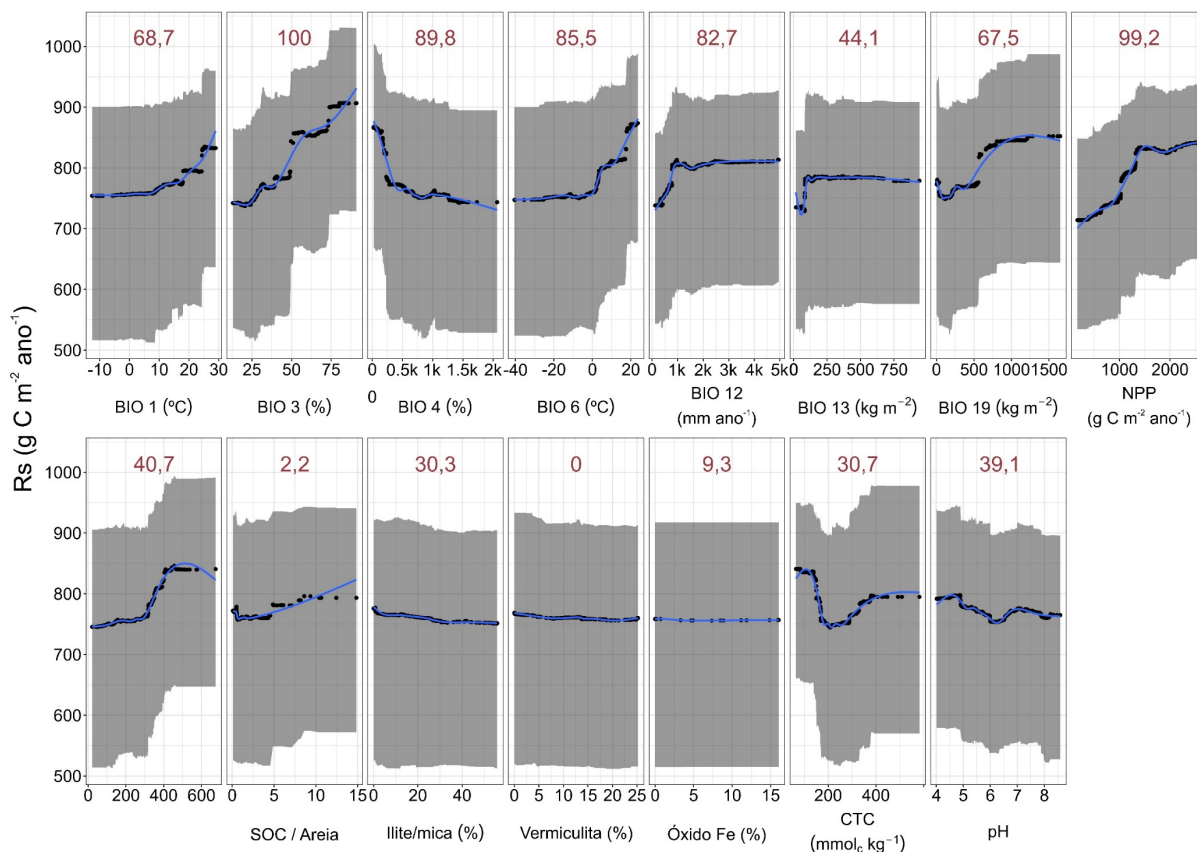
2753
2754
2755
2756
2757

Figura 3. O coeficiente de variação (CV%) representa a incerteza na predição da respiração total anual do solo (a) e da respiração heterotrófica do solo (b) a partir do modelo usando todas as 26 covariáveis.

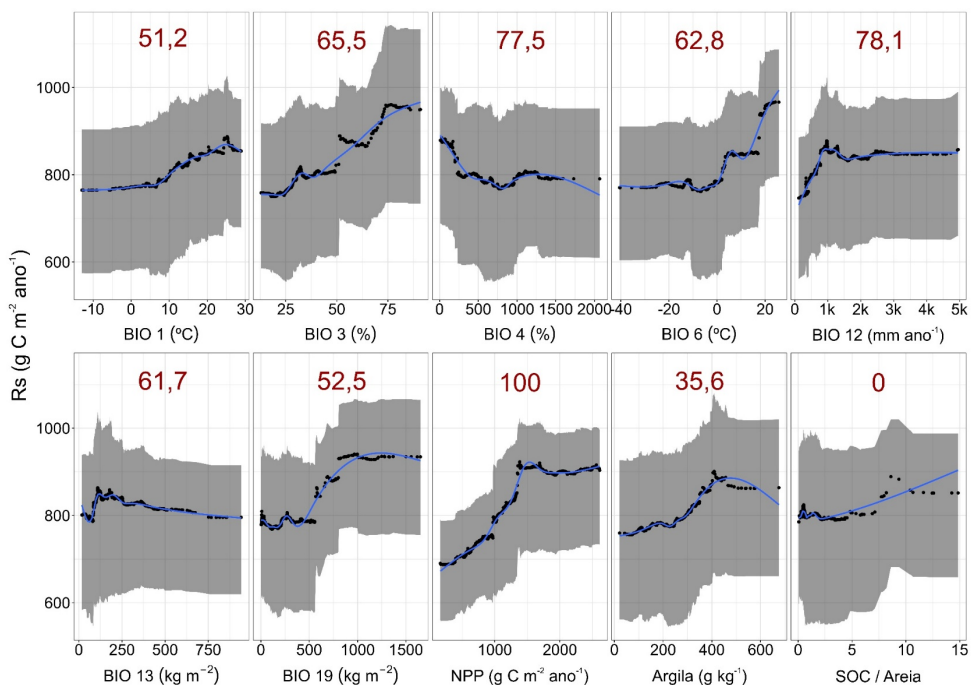


2758
 2759 **Figura 4.** O coeficiente de variação (CV%) representa a incerteza na predição da respiração
 2760 total anual do solo (a) e da respiração heterotrófica do solo (b) a partir do modelo final usado
 2761 para prever cenários futuros usando 8 covariáveis.
 2762

2763 O modelo que utilizou todas as covariáveis indicou que temperatura, precipitação, NPP
 2764 e propriedades do solo, como argila, CEC e pH, tiveram uma influência substancial em Rs e Rh
 2765 (Figura 5 e Figura 6). temperatura, precipitação, NPP e conteúdo de argila foram positivamente
 2766 correlacionados com Rs, enquanto illita/mica, CEC e pH exibiram relações inversamente
 2767 proporcionais. Para Rh, os resultados indicaram uma associação mais fraca com temperatura,
 2768 sendo que precipitação, argila, CEC e pH foram as variáveis mais expressivas.
 2769



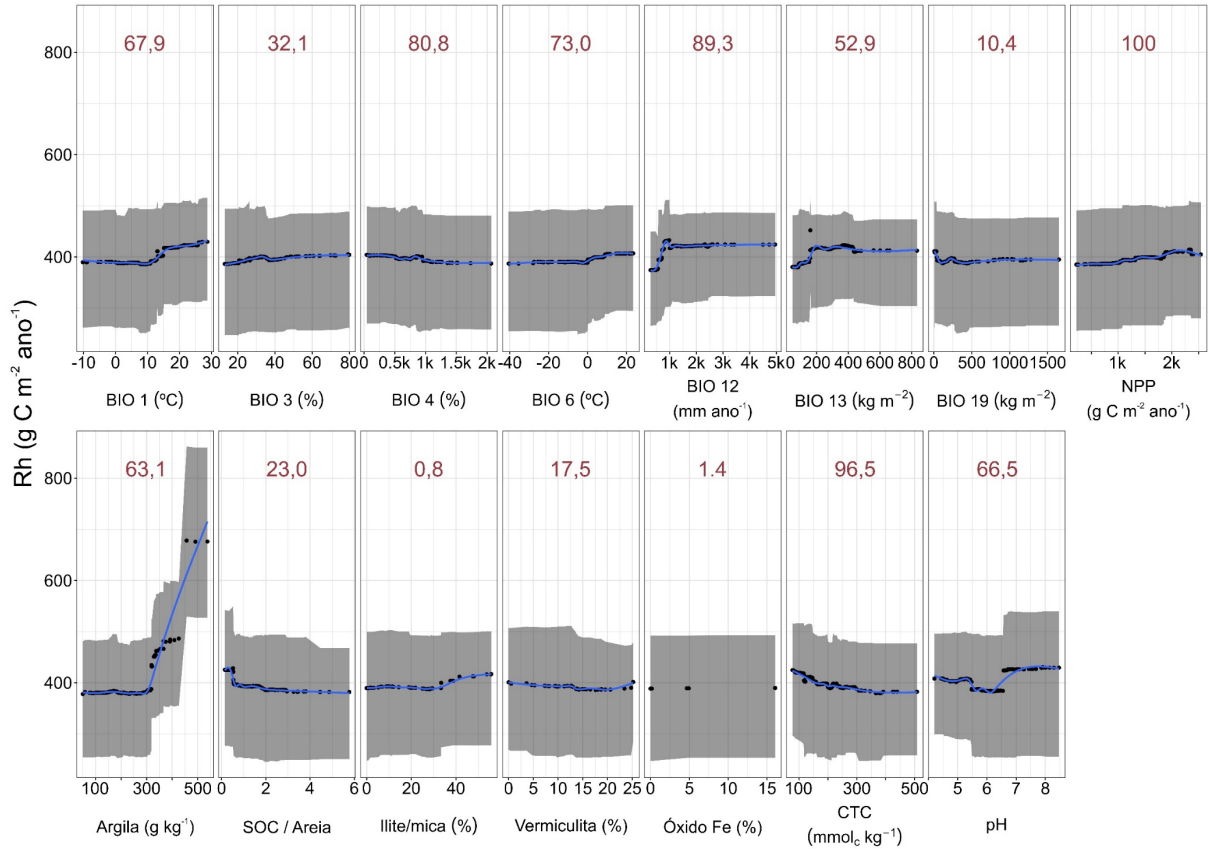
(a)



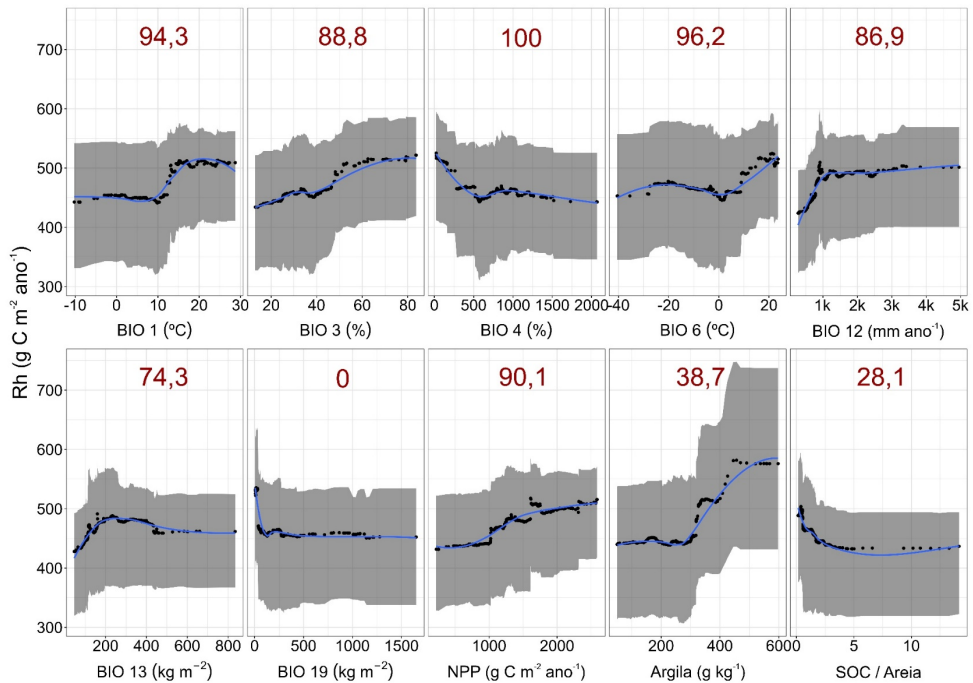
(b)

2770
2771
2772
2773
2774
2775

Figura 5. Dependência parcial das variáveis utilizadas para a distribuição espacial da respiração total do solo (R_s) utilizando todas as covariáveis (a) e o subconjunto selecionado (b). Os números em vermelho representam a importância relativa da variável para o modelo final, para mais detalhes sobre as variáveis consultar a Tabela 1.



(a)



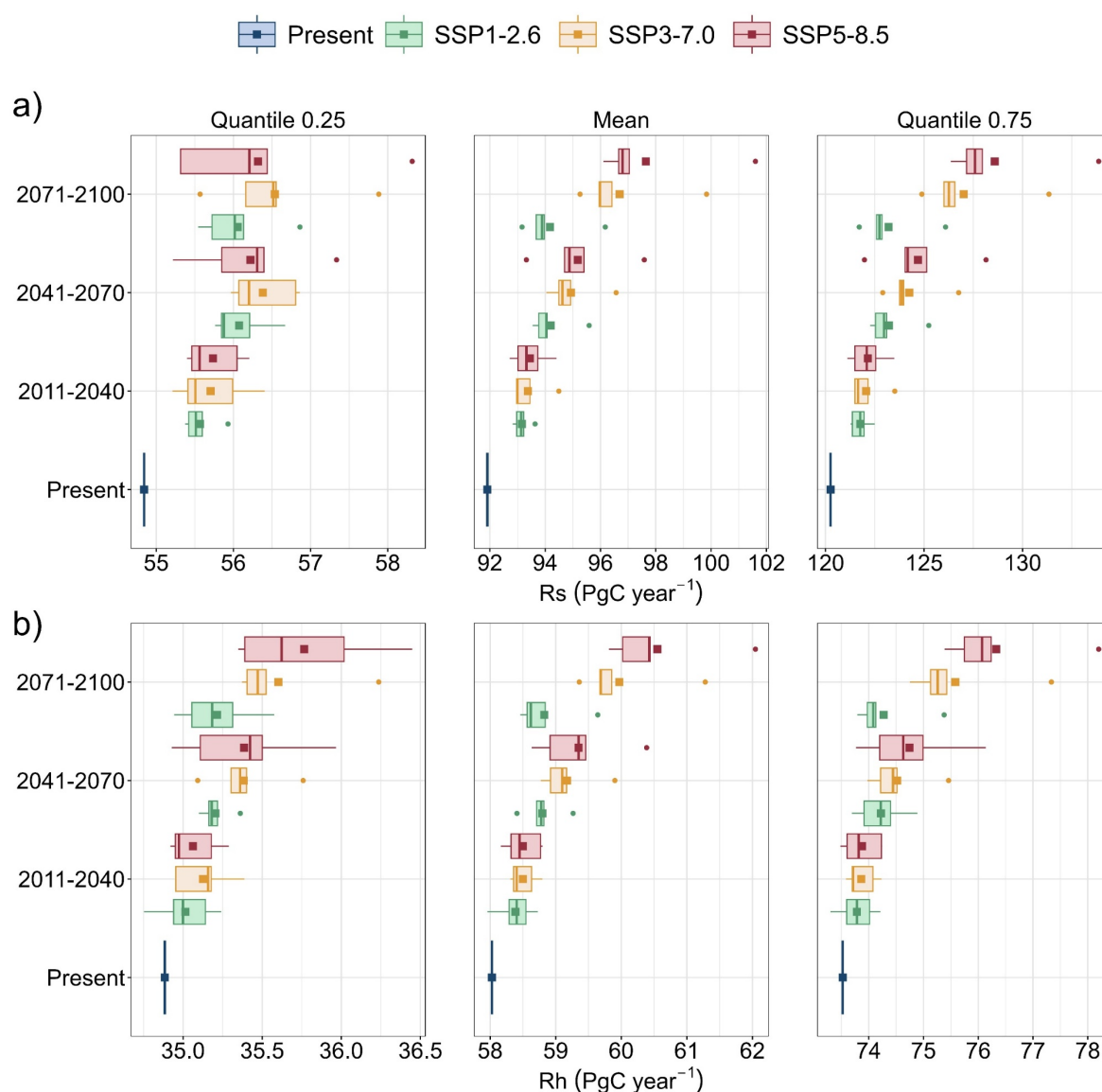
(b)

Figura 6. Dependência parcial das variáveis utilizadas para a distribuição espacial da respiração heterotrófica (R_h) utilizando todas as covariáveis (a) e o subconjunto selecionado (b). Os números em vermelho representam a importância relativa da variável para o modelo final, para mais detalhes sobre as variáveis consultar a Tabela 1.

2776
2777
2778
2779
2780
2781

2782 3.3. Taxas futuras de respiração do solo

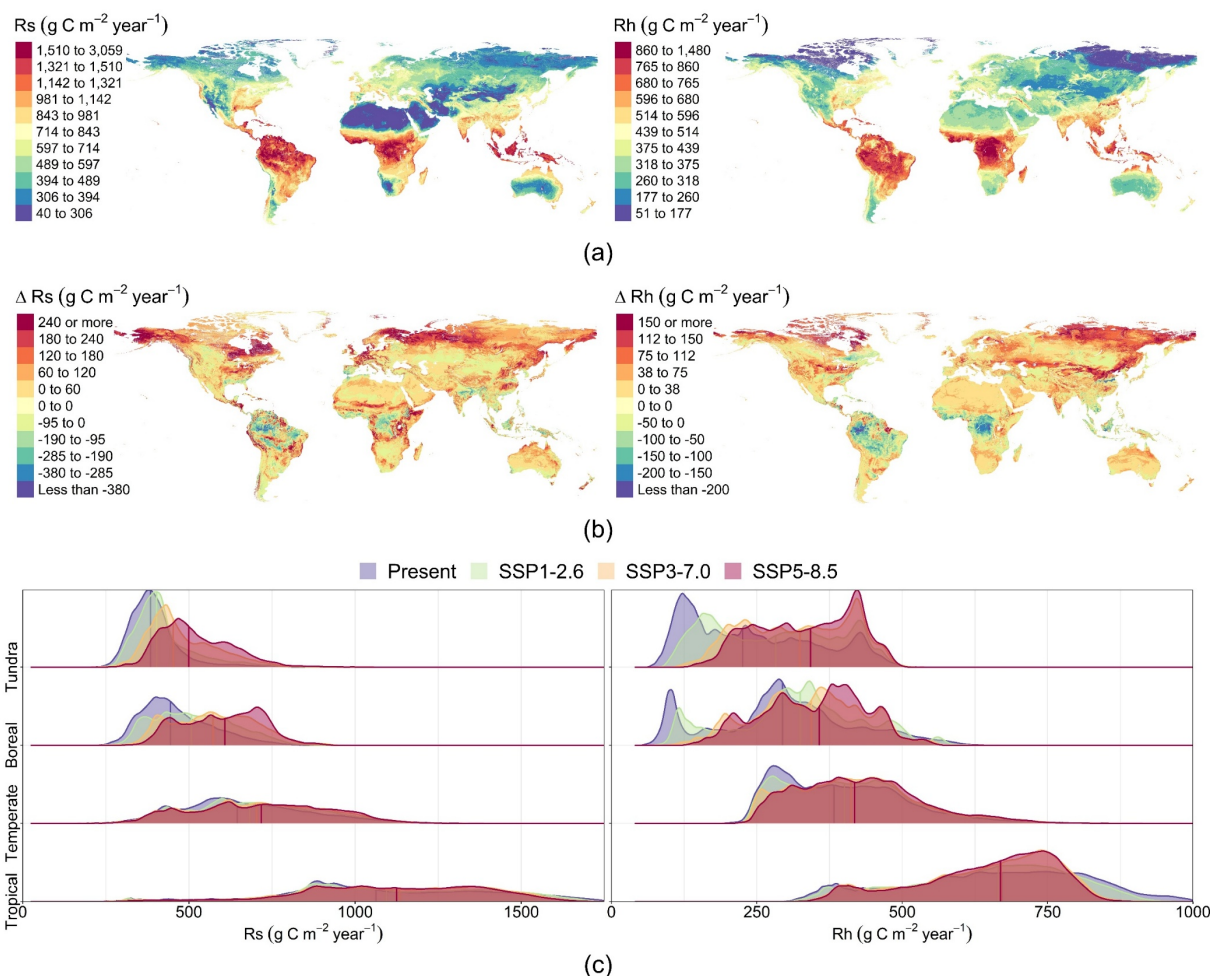
2783 Foi encontrado que o R_s global até 2100 pode variar de 56,0 (quantil 0.25) a 128,6 PgC
2784 ano^{-1} (quantil 0.75; Figura 7), com uma média de 96,2 PgC ano^{-1} , e um R_h futuro variando de
2785 35,2 (quantil 0.25) a 76,3 PgC ano^{-1} (quantil 0.75), com uma média de 59,8 PgC ano^{-1} . Sob o
2786 cenário SSP5-8.5 (forçante radiativa de 8,5 W m^{-2} até o ano 2100), o R_s e R_h podem atingir até
2787 97,6 e 60,6 PgC ano^{-1} , respectivamente, enquanto no cenário SSP1-2.6 (forçante radiativa de
2788 2,6 W m^{-2} até o ano 2100), espera-se que o R_s aumente para 94,1 e R_h para 58,2 PgC ano^{-1} em
2789 2100.
2790



2791
 2792 **Figura 7.** Valores médios e quantis anuais atuais e futuros (q0.25 e q0.75) da respiração total
 2793 do solo (Rs; a) e respiração heterotrófica do solo (Rh; b) de 2011-2040 a 2071-2100 nos cenários
 2794 de Caminhos Socioeconômicos Compartilhados (126, 370, 585). Os valores médios e quartis
 2795 representam os valores médios de cinco modelos climáticos (GFDL-ESM4, IPSL-CM6A-LR,
 2796 MPIESM1-2-HR, MRI-ESM2-0, UKESM1-0-LL). Os quadrados coloridos indicam os valores
 2797 médios.
 2798

2799 Globalmente, tanto Rs quanto Rh exibiram aumentos acentuados em direção às altas
 2800 latitudes e tenderam a diminuir nos trópicos úmidos (Figura 9). As projeções do SSP5-8.5
 2801 indicaram que Rh deve aumentar na Tundra (+50,2%) e na região Boreal (+19,3%), enquanto
 2802 taxas inalteradas ou ligeiramente decrescentes de Rs e Rh (-1,7%) podem ser esperadas para os
 2803 trópicos até 2100 (Figura 9b).

2804



2805

2806

2807

2808

2809

2810

2811

2812

2813

Figura 9. Distribuição global da respiração do solo (Rs) e da respiração heterotrófica (Rh) para o presente (a) e as mudanças projetadas (ΔRs e ΔRh) do presente para o período 2071-2100 (b). As mudanças espaciais em Rs e Rh representam os valores médios dos cenários SSP1-2.6 e SSP5-8.5 para o ano 2100 e de cinco modelos climáticos (GFDL-ESM4, IPSL-CM6A-LR, MPI-ESM1-2-HR, MRI-ESM2-0, UKESM1-0-LL). Os gráficos de densidade (linha vertical = valores médios) mostram a distribuição de Rs e Rh para os cenários presente e futuro SSP1-2.6, SSP3-7.0 e SSP5-8.5 (c).

2814

2815

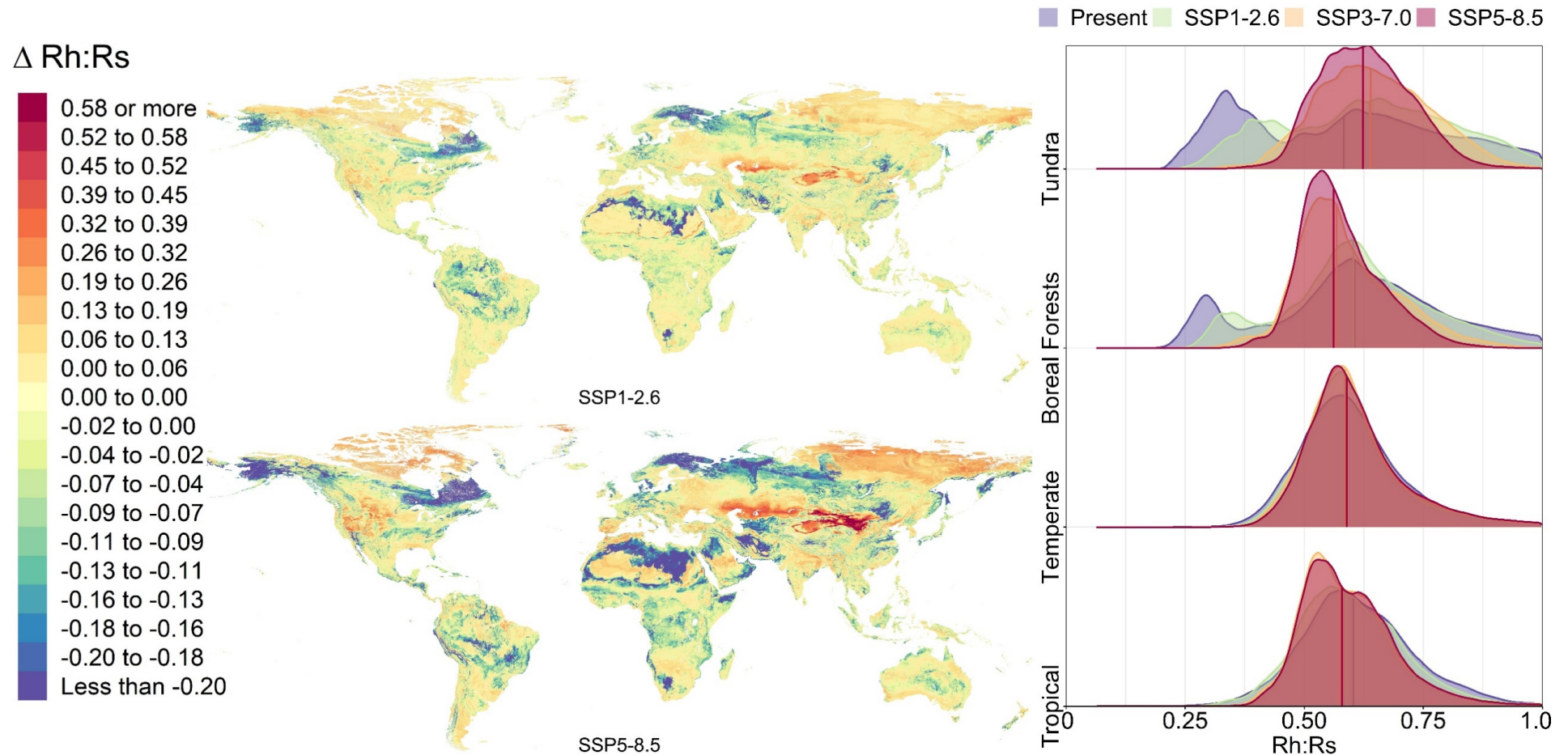
2816

2817

2818

Com base nas razões Rh/Rs , evidenciou-se que Rh aumentará mais do que Rs em cenários de alta emissão, com aumentos marcados (média de 0,54 para 0,62 globalmente) nas razões Rh/Rs em regiões secas da Ásia e, particularmente, na Tundra (Figura 10). No entanto, as razões Rh/Rs devem apresentar uma redução nas regiões boreais e Tropicais, sem mudanças generalizadas evidentes esperadas para as regiões Temperadas. Essas mudanças na razão de

2819 fluxo são diagnósticas de um *feedback* climático acelerado, à medida que o carbono do solo
2820 armazenado há muito tempo é mineralizado para a atmosfera (Bond-Lamberty et al., 2018).



2821
2822
2823
2824

Figura 10. Mudanças espaciais na relação Rh/Rs (Δ) para o cenário SSP1-2.6 e o cenário SSP5-8.5 desde o presente até 2071-2100. Todos os mapas representam os valores médios de cinco modelos climáticos (GFDL-ESM4, IPSL-CM6A-LR, MPI-ESM1-2-HR, MRI-ESM2-0, UKESM1-0-LL). Os gráficos de densidade (linha = valores médios) mostram a distribuição da relação Rh/Rs para o presente e os cenários SSP1-2.6, SSP3-7.0 e SSP5-8.5.

2825 3.4. Principais fatores que impulsionam as mudanças futuras em Rs e Rh

2826 A produtividade primária líquida (NPP), a precipitação, a temperatura e o teor de argila
2827 mostraram o efeito mais forte tanto em Rs quanto em Rh (Figura 5 e Figura 6). No entanto,
2828 esses fatores exerceram efeitos específicos em cada componente da respiração do solo. Assim,
2829 Rs tende a diminuir quando as temperaturas são superiores a 26 °C, enquanto Rh tende a
2830 aumentar.

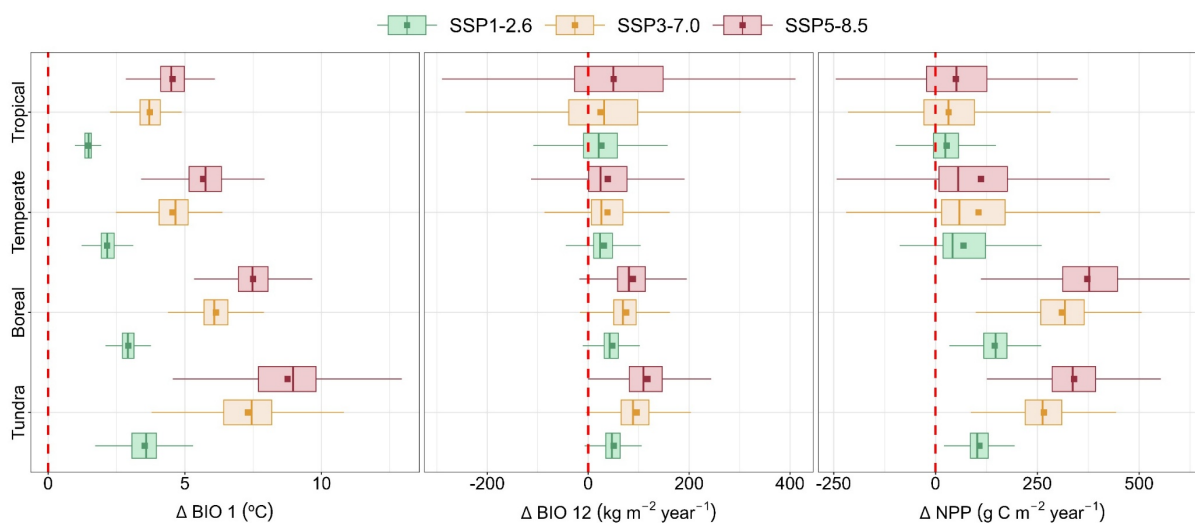
2831 Além disso, a precipitação foi positivamente correlacionada com ambos Rs e Rh até
2832 1.000 mm, com uma subsequente diminuição até alcançar 2.000 mm, onde tendeu a se
2833 estabilizar. A NPP teve uma relação positiva constante com ambos Rs e Rh, mas essas variáveis
2834 não responderam a níveis de produtividade acima de 1.700 gC m⁻² ano⁻¹ (Figura 5 e Figura 6).

2835 O teor de argila exibiu uma influência ligeiramente mais forte em Rh em relação a Rs,
2836 enquanto as razões SOC / Areia não tiveram influência significativa em Rs, mas ajudaram a
2837 explicar a variação em Rh (Figura 6). Para o período considerado até 2100, projeta-se que a
2838 temperatura (BIO 1) aumente em todos os cenários (Figura 11), com valores mais altos na
2839 Tundra (+8,9; 5,6 a 11,6 °C) e nas regiões Boreais (+7,4; 6 a 8,8 °C), Temperadas (+5,7; 3,7 a
2840 6,9 °C) e Tropicais (+4,4; 3,6 a 5,4 °C).

2841 Além disso, as mudanças na precipitação (BIO 12) têm maior probabilidade de serem
2842 positivas na Tundra (112; 55 a 260 mm) e na região Boreal (82; 27 a 192 mm). Por outro lado,
2843 as projeções indicam uma maior variabilidade para outras regiões, com diminuições ou
2844 aumentos na precipitação nas regiões Tropicais (37; -315 a 371 mm) e Temperadas (40; -50 a
2845 196 mm).

2846 Em termos de crescimento das plantas, os dados sugerem variações predominantemente
2847 positivas na NPP (gC m⁻² ano⁻¹), na Tundra (336; 193 a 492,2) e na região Boreal (347; 174 a
2848 517,6), enquanto mudanças tanto negativas quanto positivas podem ocorrer nas regiões
2849 Tropicais (40,9; -200 a 275) e Temperados (143,1; -32,2 a 385).

2850



2851

2852

2853 **Figura 11.** Mudanças futuras na temperatura média anual ($\Delta \text{BIO 1}$), precipitação média anual
 2854 ($\Delta \text{BIO 12}$) e produtividade primária líquida (ΔNPP) do presente até o período 2071-2100 para
 2855 os cenários SPP1-2.6, SPP3-7.0 e SSP5-8.5. A linha em vermelho indica limite entre aumento
 2856 ou redução para as variáveis e os quadrados coloridos representam valores médios.

2856

2857

2858 Geralmente, o teor de argila e as razões SOC / Areia tiveram uma influência mais forte
 2859 nas mudanças projetadas para Rh do que para Rs (Figura 5 e Figura 6). Estimou-se um aumento
 2860 de cerca de $110 \text{ gC m}^{-2} \text{ ano}^{-1}$ em Rs para teores de argila variando entre aproximadamente 200
 2861 até 400 g kg^{-1} solo, após o que Rs diminuiu (Figura 5). Para Rh, projetou-se um aumento médio
 2862 de cerca de $150 \text{ gC m}^{-2} \text{ ano}^{-1}$ à medida que o teor de argila aumentou de cerca de 300 até 500 g
 2863 kg^{-1} solo (Figura 6).

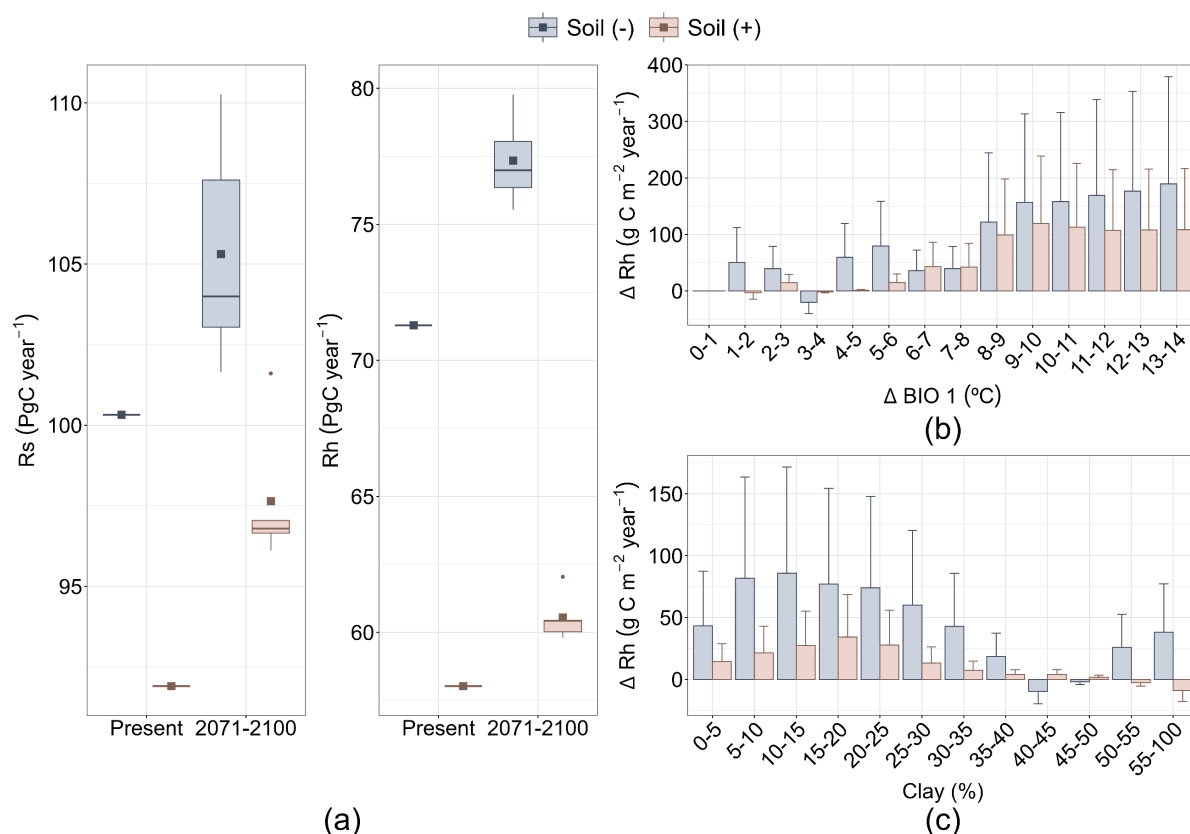
2863

2864 Além disso, a maior variação em Rh como função das razões SOC / Areia seria
 2865 observada para valores menores que 5 hgC kg^{-1} areia, representando áreas com solos arenosos
 2866 e/ou uma combinação de alto teor de areia e baixos estoques de SOC. Abaixo desse limite, uma
 2867 diminuição de 1 unidade nas razões SOC / Areia levaria a um aumento de até 2 vezes em Rh
 2868 (Figura 6).

2868

2869 Sob as condições climáticas atuais, estimaram-se valores menores de Rs ($91,9 \text{ PgC ano}^{-1}$)
 2870 e Rh ($58,0 \text{ PgC ano}^{-1}$) ao considerar variáveis de solo em comparação com as mesmas
 2871 projeções obtidas sem elas ($R_s = 100,3 \text{ PgC ano}^{-1}$; $R_h = 71,3 \text{ PgC ano}^{-1}$), conforme mostrado

2871 na Figura 12. Da mesma forma, observou-se estimativas menores para Rs ($97,6 \text{ PgC ano}^{-1}$) e
 2872 Rh ($60,5 \text{ PgC ano}^{-1}$) ao incluir variáveis de solo sob o clima futuro (2071-2100), enquanto essas
 2873 estimativas eram cerca de $105,3 \text{ PgC ano}^{-1}$ para Rs e $77,3 \text{ PgC ano}^{-1}$ para Rh quando não
 2874 limitadas por propriedades do solo (Figura 12).
 2875



2876 **Figura 12.** Respiração global total do solo (Rs) e respiração heterotrófica (Rh) do presente e
 2877 para o período 2071-2100 modelada na ausência (-solo) e na presença de propriedades do solo
 2878 (+solo) (a). Os valores para 2071-2100 representam os valores médios de Rs e Rh do cenário
 2879 SSP5-8.5 e cinco modelos climáticos (GFDL-ESM4, IPSL-CM6A-LR, MPI-ESM1-2-HR,
 2880 MRI-ESM2-0, UKESM1-0 -LL); mudanças projetadas na respiração heterotrófica (ΔRh) em
 2882 relação à variação da temperatura média anual ($\Delta \text{BIO 1}$) do presente até o período 2071-2100
 2883 (b), e aos intervalos de conteúdo de argila (c). As barras verticais representam os limites do
 2884 intervalo de confiança de 99% estimado por bootstrap percentílico. Os quadrados coloridos
 2885 representam os valores médios.
 2886

2887 Além disso, incluir variáveis de solo reduziu a variabilidade entre os modelos (Figura
 2888 12a), e as variáveis de solo podem moderar o efeito previsto do aquecimento em Rh,
 2889 especialmente onde se espera que as temperaturas médias aumentem mais de $8,0 \text{ }^{\circ}\text{C}$ (Figura

2890 12b). As projeções indicam uma redução consistente em Rh com o aumento do teor de argila,
2891 especialmente em solos com teores acima de 200 g kg⁻¹ (Figura 12c). Por outro lado, a
2892 modelagem sem propriedades do solo superestimou sistematicamente Rh para solos com teor
2893 de argila abaixo de 400 g kg⁻¹ solo ou acima de 500 g kg⁻¹ solo (Figura 12c). Esse padrão é
2894 consistente com a relação entre argila e Rh (Figura 5 e Figura 6).

2895

2896 4. Discussão

2897 4.1. Restrições e fatores desencadeantes da respiração do solo em ecorregiões no futuro

2898 Estimativas anteriores de respiração do solo (Rs) global, baseadas em modelos
2899 orientados por dados, variaram de 68,0 a 101 PgC ano⁻¹ (Hashimoto et al., 2023), e um dos
2900 estudos mais recentes sobre respiração heterotrófica (Rh) global relatou um fluxo de 53,4 PgC
2901 ano⁻¹ (Stell et al., 2021). As estimativas globais encontradas para Rs (91,6 PgC ano⁻¹) e Rh (57,9
2902 PgC ano⁻¹) estão dentro dos intervalos reportados na literatura.

2903 Além disso, os resultados corroboram estudos anteriores que indicam uma diminuição
2904 significativa de Rs em ecorregiões Tropicais e Subtropicais e um aumento significativo em
2905 regiões Boreais e árticas entre 1987 e 2016 (Lei et al., 2021). A disponibilidade de água é um
2906 fator fundamental na regulação do *feedback* entre carbono e clima em condições de
2907 aquecimento (Quan et al., 2019), e uma redução geral em Rs e Rh nos trópicos é provavelmente
2908 relacionada à projeção de declínio da precipitação em algumas áreas (Figura 9 e Figura 11).

2909 Na verdade, a precipitação média anual (BIO 12) foi um dos principais fatores
2910 explicativos da distribuição espacial de Rs em nosso conjunto de dados (Figura 5), e resultados
2911 experimentais anteriores mostraram reduções de Rs de até 45% sob condições de seca (Singh
2912 et al., 2021; Zheng et al., 2021). Contudo, é importante destacar isto pode variar, pois as regiões
2913 Tropicais abrangem ecossistemas diversos com diferentes tipos de solo e microclimas.

2914 Observou-se que Rh também pode aumentar em certas áreas, como as terras baixas e as
2915 maiores altitudes da Amazônia brasileira (Figura 9b). Isso corrobora resultados in situ que
2916 indicaram aumentos significativos em Rh em resposta ao aquecimento experimental do solo em
2917 uma floresta Tropical de planície no Panamá (Nottingham et al., 2020).

2918 Assim, os níveis geralmente reduzidos de Rs em áreas Tropicais (Figura 9b) podem ser
2919 mais problemáticos do que parecem (Hartley et al., 2021): aumentos de temperatura e reduções
2920 de precipitação espacialmente acoplados devem reduzir Rs não apenas devido à menor
2921 mineralização microbiana de SOC, mas também devido à menor disponibilidade de água que
2922 prejudica o crescimento da vegetação (Doughty et al., 2015).

2923 Análises históricas indicaram uma diminuição na resiliência de florestas Tropicais,
2924 áridas e Temperadas associada à limitação de água e variabilidade climática (Forzieri et al.,
2925 2022). Isso é preocupante, pois a biomassa viva armazena proporcionalmente mais carbono
2926 com a diminuição da latitude, e é muito improvável que a mortalidade de árvores sob seca
2927 favoreça o armazenamento de SOC (McDowell et al., 2018; Yu et al., 2022).

2928 Contrariamente à tendência geral, os resultados encontrados também sugerem uma
2929 redução de Rs e Rh em áreas úmidas Tropicais da África Central, onde se espera um aumento
2930 na precipitação e temperatura médias anuais. As áreas úmidas da África Central abrigam uma
2931 das maiores áreas de turfeiras Tropicais do mundo, que armazenam grandes quantidades de
2932 carbono (Crezee et al., 2022), e aumentos de temperatura associados ao rebaixamento do nível
2933 freático podem promover a aeração do solo e facilitar a decomposição de SOC nessas áreas,
2934 resultando em aumentos nas emissões de CO₂ (Dommain et al., 2018; Huang et al., 2021).

2935 Embora nossos modelos não considerem o nível de água sazonal, que afeta o balanço
2936 líquido de CO₂ em turfeiras (Helbig et al., 2022), valores mais altos na precipitação média do
2937 mês mais úmido (BIO 13) estão associados a reduções de Rs e Rh (Figura 5 e Figura 6), o que
2938 pode ajudar a explicar as descobertas deste estudo. No entanto, as dinâmicas futuras desses

2939 ambientes complexos estão longe de serem elucidadas, e estudos locais que considerem as
2940 tendências sazonais de troca de CO₂ em turfeiras Tropicais são urgentemente necessários para
2941 melhorar nossa compreensão das dinâmicas de fonte/sumidouro de carbono dessas áreas únicas
2942 e apoiar modelos climáticos na predição do ciclo global de carbono.

2943 Nas altas latitudes do norte, o aumento projetado na temperatura associado aos altos
2944 estoques de SOC deve favorecer futuros aumentos em Rs e Rh (Figura 9). Em conformidade,
2945 Wieder et al. (2019) sugeriram um papel importante da região ártica no futuro ciclo global de
2946 carbono sob aquecimento. Parte dessas projeções pode ser explicada pelo ambiente criogênico
2947 sob o qual os solos são formados na região circumpolar do Norte, onde os estoques de SOC
2948 podem chegar a 50 kg m⁻² em até 100 cm de profundidade (Tamocai et al., 2009).

2949 Nessa área, os solos da ordem dos Gelisols (Histels, Turbels e Orthels) armazenam
2950 aproximadamente 70% do estoque total de SOC (Hugelius et al., 2014). Portanto, taxas de
2951 degelo aumentadas podem levar a um aumento dramático em Rh, em linha com a resposta
2952 acentuada de Rh na faixa de temperatura entre 0 e 20 °C (Figura 12). Além disso, a área de
2953 *permafrost* de turfeiras do norte atualmente cobre cerca de 1,7 milhão km², mas é projetado que
2954 reduza para apenas 1 milhão km² em resposta a um aquecimento global de 2°C (Hugelius et al.,
2955 2020).

2956 Nas regiões de Tundra e Boreal, a NPP deve aumentar em até 250 g m⁻² ano⁻¹ sob o
2957 cenário SSP5-8.5 (Figura 11), em linha com observações de satélite e de campo que mostram
2958 aumento na cobertura vegetal no Ártico (Aguirre et al., 2021; Heijmans et al., 2022).
2959 Possivelmente, à medida que as plantas respondem ao aquecimento na região circumpolar do
2960 Norte, os aportes de carbono para os solos também devem aumentar.

2961 No entanto, nossas estimativas indicam que Rs e Ra podem ser mais sensíveis à NPP do
2962 que Rh sob as restrições climáticas atuais (Figura 5 e Figura 6). Portanto, o aumento da
2963 respiração radicular à medida que a NPP responde ao aquecimento pode compensar parte dos

2964 possíveis aportes de carbono aos solos. Isso pode ser particularmente relevante para a região
2965 Boreal, onde Ra deve aumentar consideravelmente sob aquecimento futuro (Figura 9).

2966 No entanto, neste estudo projetou-se um *feedback* negativo do carbono do solo e clima
2967 na região Boreal (ou seja, diminuição da razão Rh/Rs), apoiando descobertas anteriores de
2968 experimentos de respiração do ecossistema que sugerem que a região Boreal pode ser um
2969 sumidouro de CO₂ sob um clima mais quente até o final do século (Zhang et al., 2017). Chen
2970 et al. (2023) sugerem uma adaptação térmica da respiração do ecossistema em ecorregiões
2971 globais com o aumento da temperatura do ar, mas o papel das respostas diferenciais de Ra e Rh
2972 nesse processo ainda é desconhecido.

2973 As projeções futuras aqui realizadas são ainda apoiadas por um estudo recente que
2974 identificou uma resiliência crescente das florestas Boreais para se recuperar de perturbações
2975 naturais e antropogênicas entre 2000 e 2020, provavelmente devido aos benefícios de
2976 temperaturas mais altas e “fertilização” por CO₂ (Forzieri et al., 2022). Além disso, a expansão
2977 de arbustos decíduos em direção à Tundra com o aquecimento poderia aumentar as taxas de Rh
2978 (Mekonnen et al., 2021; Phillips & Wurzbarger, 2019).

2979 O *feedback* positivo do carbono do solo e clima aqui projetado para a Tundra, baseado
2980 no aumento da razão Rh/Rs, poderia indicar grandes perdas de SOC. Embora essas perdas de
2981 SOC possam ser menos severas sob o cenário (otimista) SSP1-2.6, é necessário considerar que
2982 múltiplos fatores interagem para afetar Rs. Assim, a incerteza quanto à magnitude da mudança
2983 nos fluxos futuros de Rs no Ártico não deve ser ignorada.

2984 A alta incerteza nas previsões de Rs e Rh no Ártico pode estar relacionada à baixa
2985 densidade de observações de respiração do solo nessa região, o que é conhecido por afetar as
2986 previsões em mapeamentos digitais de solo (Gomes et al., 2019; Moquedace et al., 2024). Por
2987 exemplo, devido a uma maior densidade de medições nesta região, o mapa de Rs tinha menor
2988 incerteza do que o mapa de Rh para a região do Ártico.

2989

2990 **4.2. Importância das propriedades do solo para a respiração do solo**

2991 Os resultados sugerem que os modelos de respiração do solo se beneficiam da inclusão
2992 de um parâmetro chave do solo, como a textura. Além disso, a falta de consideração da
2993 correlação temporal das observações de R_s ou R_h durante a validação cruzada pode subestimar
2994 o nível real de incerteza nas projeções obtidas. Assim, um esforço global é importante para
2995 aumentar o número de pontos de amostragem para R_s e R_h , especialmente nas regiões sub-
2996 representadas.

2997 Com mais dados disponíveis para R_s e R_h , o desempenho dos modelos existentes
2998 poderia ser melhorado. Além disso, os modelos atuais parecem depender excessivamente de
2999 variáveis climáticas, potencialmente superestimando R_h e subestimando R_s , especialmente em
3000 solos de textura fina (Figura 11b).

3001 Estudos recentes revelam que muitos Modelos de Sistemas da Terra (ESMs) têm
3002 dificuldades em representar com precisão a distribuição espacial de R_h , frequentemente
3003 superestimando R_h em áreas com maior precipitação (Guenet et al., 2024). Um estudo que
3004 utilizou um ESM considerando a textura do solo não conseguiu representar o efeito da
3005 temperatura sobre os estoques de C em solos de textura grosseira e superestimou seu impacto
3006 em solos de textura fina em regiões quentes (Hartley et al., 2021).

3007 A inclusão do conteúdo de argila e das razões SOC / Areia na modelagem reduziu as
3008 taxas projetadas de respiração do solo. Taxas baixas de R_h podem refletir altas razões SOC /
3009 Areia e estoques de SOC em solos argilosos, onde a complexação com minerais de tamanho
3010 argila ou a proteção dentro de agregados desacelera a mineralização do SOC (Figura 5 e Figura
3011 6). Em contraste, solos arenosos, frequentemente encontrados em ambientes secos ou frios com
3012 pedogênese limitada (por exemplo, na Europa setentrional), podem ter baixas razões SOC /

3013 Areia, indicando proteção limitada do SOC e maior sensibilidade a aumentos de temperatura
3014 (Hartley et al., 2021).

3015 O aumento projetado de Ra nas regiões Boreais é um ponto fundamental para a dinâmica
3016 do carbono (C) terrestre e das propriedades do solo. O aumento nas taxas de Ra poderia levar a
3017 maiores entradas de C no solo, mas os Podzóis com textura arenosa, que cobrem até 30% dos
3018 ecossistemas Boreais, podem limitar o armazenamento de C, apesar do aumento das entradas
3019 de C (DeLuca & Boisvenue, 2012).

3020 Nos Podzóis, o SOC tende a acumular-se e envelhecer em camadas mais profundas do
3021 solo, com a argila e os óxidos aumentando com a profundidade (Hensgens et al., 2021). Altas
3022 razões SOC / Areia também podem indicar a acumulação de C em ambientes hidromórficos e
3023 gelados, onde a mineralização do SOC é limitada pela disponibilidade de oxigênio e água, como
3024 observado no Ártico (Tamocai et al., 2009).

3025 Nessas regiões, espera-se que a Rh aumente substancialmente com o aumento das
3026 temperaturas, especialmente em áreas com baixo conteúdo de argila e pedogênese limitada
3027 (Shangguan et al., 2014). Assim, as perdas de SOC podem compensar quaisquer ganhos de C
3028 resultantes do aumento da NPP (Naidu & Bagchi, 2021). Melhorias na representação dos
3029 mecanismos que controlam a proteção do SOC e nos processos biogeoquímicos associados são
3030 essenciais (Lehmann et al., 2020; Xu et al., 2016).

3031 O desenvolvimento e a intemperização futura dos solos também são fatores
3032 fundamentais que podem afetar a respiração do solo. No entanto, a aplicação desses fatores em
3033 modelos existentes é limitada pela disponibilidade de dados. Embora não tenhamos incluído
3034 tipos de argila, capacidade de troca de cátions (CTC) e pH nas projeções futuras, a análise
3035 parcial oferece insights sobre seus efeitos globais na Rs e Rh.

3036 O conteúdo de illita/mica mostrou uma inversamente proporcional consistente com as
3037 taxas de respiração do solo (Figura 5 e Figura 6), possivelmente devido a menores taxas de

3038 intemperismo do solo em regiões áridas e de alta latitude (Ito & Wagai, 2017). As taxas de
3039 intemperismo do solo influenciam a abundância de minerais primários e o armazenamento de
3040 carbono orgânico do solo (SOC).

3041 O aumento das taxas de intemperismo pode promover a pedogênese de minerais reativos
3042 e a permanência do SOC no curto prazo (Kögel-Knabner & Amelung, 2021). No entanto, a
3043 longo prazo, a acumulação de minerais menos reativos pode reduzir o potencial do solo para
3044 armazenar SOC, especialmente em regiões úmidas com altos estoques de SOC. Esse cenário
3045 pode contribuir para a mineralização do SOC em resposta às mudanças climáticas.

3046 A CTC e o pH expressaram-se como covariáveis significativas, negativamente
3047 correlacionadas com a R_s e a R_h . A CTC, especialmente a de minerais pouco cristalinos, deve
3048 mudar com o desenvolvimento e intemperismo futuros dos solos. O pH afeta a atividade
3049 microbiana e as taxas de decomposição do SOC, com ambientes ácidos ou alcalinos
3050 influenciando a decomposição. O pH do solo tem sido identificado como um fator importante
3051 na sensibilidade global da respiração do solo à temperatura (Sáez-Sandino et al., 2023).

3052 O aumento da produtividade vegetal devido ao aquecimento provavelmente aumentará
3053 a formação do solo, as taxas de intemperismo e a adição de SOC, influenciando a respiração do
3054 solo. Estudos mostram que o armazenamento de carbono em ambientes frios é
3055 predominantemente nas frações de matéria orgânica associada aos minerais, impulsionado pela
3056 textura do solo e menos vulnerável ao aquecimento (Plaza et al., 2022).

3057 Em contraste, o SOC nas camadas minerais superiores do Ártico, dominado por carbono
3058 orgânico particulado, foi negativamente afetado pelo aquecimento experimental (García-
3059 Palacios et al., 2024). Esses achados destacam as incertezas associadas ao desenvolvimento do
3060 solo e ao *feedback* climático sobre a dinâmica futura do SOC.

3061

3062 **4.3. Desafios atuais e futuras direções na modelagem da respiração do solo**

3063 É importante considerar vários fatores relevantes para as inferências realizadas.
3064 Primeiro, um modelo empírico de espaço-para-tempo foi utilizado para projetar cenários
3065 futuros. Este modelo também foi aplicado em nível regional para explorar a influência da
3066 temperatura nas frações de carbono orgânico do solo (SOC) (Lugato et al., 2021) e globalmente
3067 para analisar as relações entre SOC, argila e temperatura (Hartley et al., 2021).

3068 Embora este modelo não considere as relações históricas entre R_s e R_h com as variáveis
3069 explicativas, modelos espaço-temporais podem indicar essas relações. Melhorar as observações
3070 de R_s e R_h e os dados das variáveis explicativas ao longo do espaço e do tempo (Roberts et al.,
3071 2017) ajudaria no desenvolvimento de modelos e métodos para explorar os efeitos das
3072 mudanças climáticas futuras nas dinâmicas do SOC.

3073 Em segundo lugar, o modelo futuro deve considerar o potencial da deposição de
3074 nitrogênio e outras variáveis que se espera que mudem, impactando a produção de biomassa
3075 futura e, conseqüentemente, R_s e R_h . Além disso, há estimativas mais precisas para a produção
3076 primária líquida (NPP) atual (por exemplo, MODIS) e projeções futuras, mas não foram
3077 utilizadas devido à resolução temporal requerida para este estudo.

3078 Por fim, os modelos para R_h são baseados em um número reduzido de medições e,
3079 embora os resultados para as latitudes mais altas sejam consistentes com as tendências
3080 observadas na literatura, os cenários futuros devem ser analisados com cautela. A alta incerteza
3081 em áreas de alta latitude está relacionada à baixa representatividade das observações de R_s e
3082 R_h (Stell et al., 2021; Xu et al., 2016), limitando a capacidade de fazer previsões espaciais
3083 quando os dados são escassos (Meyer & Pebesma, 2022).

3084 É essencial destacar que os dados espaciais (ou seja, dados de solo e clima) também
3085 resultam de abordagens de modelagem com incerteza associada, o que pode intensificar a
3086 propagação de erros. Portanto, essas regiões devem ser priorizadas em futuras pesquisas, a fim

3087 de aumentar a disponibilidade de dados, especialmente de experimentos de longo prazo, e,
3088 consequentemente, melhorar a capacidade de projetar sua sensibilidade às mudanças climáticas.

3089 Em resumo, forneceu-se evidências de que a melhoria das projeções de *feedbacks* clima-
3090 respiração do solo depende fundamentalmente de: i) considerar propriedades do solo (por
3091 exemplo, textura); ii) superar os desafios para representar os mecanismos físicos e químicos
3092 que regulam a proteção do SOC nos modelos; e iii) melhorar a base de dados para medições
3093 combinadas de R_s e R_h , apesar dos desafios para determinar R_h em condições de campo.
3094 Combinadas, essas iniciativas devem aumentar a capacidade de modelar e restringir os
3095 *feedbacks* clima-respiração do solo, reduzir incertezas conhecidas e desconhecidas, e elucidar
3096 a natureza e o tempo do *feedback* carbono-clima no sistema terrestre.

3097

3098 **5. Conclusão**

3099 O estudo confirmou que a textura do solo, especialmente o teor de argila, desempenha
3100 um papel fundamental na modulação das respostas de R_s e R_h ao aquecimento global. A adição
3101 de variáveis relacionadas à textura do solo nos modelos melhorou a precisão das predições de
3102 R_s e R_h . A análise revelou que os efeitos das mudanças climáticas sobre R_s e R_h variam
3103 significativamente dependendo do cenário.

3104 Cenários com maior aquecimento (SSP3-7.0 e SSP5-8.5) mostraram maior impacto na
3105 respiração do solo, especialmente em regiões de alta latitude e em solos com menor teor de
3106 argila. Os resultados destacaram que a resposta da respiração do solo às mudanças climáticas é
3107 altamente variável entre as regiões.

3108 Em áreas com solos arenosos, a sensibilidade ao aquecimento é maior, enquanto solos
3109 argilosos, que tendem a proteger o carbono orgânico do solo, mostram menor sensibilidade. O
3110 estudo evidenciou a necessidade de ampliar a base de dados de observações de R_s e R_h ,
3111 especialmente em regiões sub-representadas, para refinar ainda mais os modelos. Além disso,

3112 futuras pesquisas devem focar em entender melhor como diferentes tipos de solo respondem às
3113 mudanças climáticas e como isso afeta a dinâmica de carbono global.

3114

3115 **Referências**

3116 Aguirre, D., Benhumea, A. E., & McLaren, J. R. (2021). Shrub encroachment affects tundra
3117 ecosystem properties through their living canopy rather than increased litter inputs. *Soil*
3118 *Biology and Biochemistry*, 153, 108121. <https://doi.org/10.1016/j.soilbio.2020.108121>

3119 Baston, D. (2022). *exactextractr: Fast Extraction from Raster Datasets using Polygons* (0.9.1).
3120 <https://cran.r-project.org/package=exactextractr>

3121 Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rödenbeck, C.,
3122 Arain, M. A., Baldocchi, D., Bonan, G. B., Bondeau, A., Cescatti, A., Lasslop, G.,
3123 Lindroth, A., Lomas, M., Luysaert, S., Margolis, H., Oleson, K. W., Rouspard, O., ...
3124 Papale, D. (2010). Terrestrial gross carbon dioxide uptake: Global distribution and
3125 covariation with climate. *Science*, 329(5993), 834–838.
3126 <https://doi.org/10.1126/science.1184984>

3127 Bengtsson, H. (2023). *parallelly: Enhancing the “parallel” Package* (1.36.0). [https://cran.r-](https://cran.r-project.org/package=parallelly)
3128 [project.org/package=parallelly](https://cran.r-project.org/package=parallelly)

3129 Bond-Lamberty, B., Bailey, V. L., Chen, M., Gough, C. M., & Vargas, R. (2018). Globally rising
3130 soil heterotrophic respiration over recent decades. *Nature*, 560(7716), 80–83.
3131 <https://doi.org/10.1038/s41586-018-0358-x>

3132 Bond-Lamberty, B., & Thomson, A. (2010). Temperature-associated increases in the global soil
3133 respiration record. *Nature*, 464(7288), 579–582. <https://doi.org/10.1038/nature08930>

3134 Bond-Lamberty, B., Wang, C., & Gower, S. T. (2004). A global relationship between the
3135 heterotrophic and autotrophic components of soil respiration? *Global Change Biology*,
3136 10(10), 1756–1766. <https://doi.org/10.1111/j.1365-2486.2004.00816.x>

- 3137 Breiman, L. (2001). Random forests. *Random Forests*, 1–122. *Machine Learning*, 45(45), 5–
3138 32. <https://doi.org/10.1023/A:1010933404324/METRICS>
- 3139 Cable, J. M., Ogle, K., Williams, D. G., Weltzin, J. F., & Huxman, T. E. (2008). Soil texture
3140 drives responses of soil respiration to precipitation pulses in the sonoran desert:
3141 Implications for climate change. *Ecosystems*, 11(6), 961–979.
3142 <https://doi.org/10.1007/s10021-008-9172-x>
- 3143 Carey, J. C., Tang, J., Templer, P. H., Kroeger, K. D., Crowther, T. W., Burton, A. J., Dukes, J.
3144 S., Emmett, B., Frey, S. D., Heskell, M. A., Jiang, L., Machmuller, M. B., Mohan, J.,
3145 Panetta, A. M., Reich, P. B., Reinschj, S., Wang, X., Allison, S. D., Bamminger, C., ...
3146 Tietema, A. (2016). Temperature response of soil respiration largely unaltered with
3147 experimental warming. *Proceedings of the National Academy of Sciences of the United*
3148 *States of America*, 113(48), 13797–13802. <https://doi.org/10.1073/pnas.1605365113>
- 3149 Chen, W., Wang, S., Wang, J., Xia, J., Luo, Y., Yu, G., & Niu, S. (2023). Evidence for
3150 widespread thermal optimality of ecosystem respiration. *Nature Ecology and Evolution*,
3151 7(9), 1379–1387. <https://doi.org/10.1038/s41559-023-02121-w>
- 3152 Crezee, B., Dargie, G. C., Ewango, C. E. N., Mitchard, E. T. A., Emba B, O., Kanyama T, J.,
3153 Bola, P., Ndjango, J. B. N., Girkin, N. T., Bocko, Y. E., Ifo, S. A., Hubau, W., Seidensticker,
3154 D., Batumike, R., Imani, G., Cuní-Sanchez, A., Kiahtipes, C. A., Lebamba, J., Wotzka, H.
3155 P., ... Lewis, S. L. (2022). Mapping peat thickness and carbon stocks of the central Congo
3156 Basin using field data. *Nature Geoscience*, 15(8), 639–644.
3157 <https://doi.org/10.1038/s41561-022-00966-7>
- 3158 Crowther, T. W., Todd-Brown, K. E. O., Rowe, C. W., Wieder, W. R., Carey, J. C., MacHmuller,
3159 M. B., Snoek, B. L., Fang, S., Zhou, G., Allison, S. D., Blair, J. M., Bridgham, S. D.,
3160 Burton, A. J., Carrillo, Y., Reich, P. B., Clark, J. S., Classen, A. T., Dijkstra, F. A., Elberling,

- 3161 B., ... Bradford, M. A. (2016). Quantifying global soil carbon losses in response to
3162 warming. *Nature*, 540(7631), 104–108. <https://doi.org/10.1038/nature20150>
- 3163 DeLuca, T. H., & Boisvenue, C. (2012). Boreal forest soil carbon: Distribution, function and
3164 modelling. *Forestry*, 85(2), 161–184. <https://doi.org/10.1093/forestry/cps003>
- 3165 Dexter, A. R., Richard, G., Arrouays, D., Czyz, E. A., Jolivet, C., & Duval, O. (2008).
3166 Complexed organic matter controls soil physical properties. *Geoderma*, 144(3–4), 620–
3167 627. <https://doi.org/10.1016/j.geoderma.2008.01.022>
- 3168 Doetterl, S., Stevens, A., Six, J., Merckx, R., Van Oost, K., Casanova Pinto, M., Casanova-
3169 Katny, A., Muñoz, C., Boudin, M., Zagal Venegas, E., & Boeckx, P. (2015). Soil carbon
3170 storage controlled by interactions between geochemistry and climate. *Nature Geoscience*,
3171 8(10), 780–783. <https://doi.org/10.1038/ngeo2516>
- 3172 Dommain, R., Frohling, S., Jeltsch-Thömmes, A., Joos, F., Couwenberg, J., & Glaser, P. H.
3173 (2018). A radiative forcing analysis of tropical peatlands before and after their conversion
3174 to agricultural plantations. *Global Change Biology*, 24(11), 5518–5533.
3175 <https://doi.org/10.1111/gcb.14400>
- 3176 Doughty, C. E., Metcalfe, D. B., Girardin, C. A. J., Amézquita, F. F., Cabrera, D. G., Huasco,
3177 W. H., Silva-Espejo, J. E., Araujo-Murakami, A., Da Costa, M. C., Rocha, W., Feldpausch,
3178 T. R., Mendoza, A. L. M., Da Costa, A. C. L., Meir, P., Phillips, O. L., & Malhi, Y. (2015).
3179 Drought impact on forest carbon dynamics and fluxes in Amazonia. *Nature*, 519(7541),
3180 78–82. <https://doi.org/10.1038/nature14213>
- 3181 Evans, J. S., & Murphy, M. A. (2015). *rfUtilities: Random Forests Model Selection and*
3182 *Performance Evaluation* (2.1-5). [https://cran.r-](https://cran.r-project.org/web/packages/rfUtilities/index.html)
3183 [project.org/web/packages/rfUtilities/index.html](https://cran.r-project.org/web/packages/rfUtilities/index.html)

- 3184 Fernandes Filho, E. I., Moquedace, C. M., & Veloso, G. V. (2021). *mdsFuncs: Auxiliar functions*
3185 *to pre-process raster data and modelling with machine learning models using caret*
3186 *package* (0.2.10).
- 3187 Forzieri, G., Dakos, V., McDowell, N. G., Ramdane, A., & Cescatti, A. (2022). Emerging
3188 signals of declining forest resilience under climate change. *Nature*, *608*(7923), 534–539.
3189 <https://doi.org/10.1038/s41586-022-04959-9>
- 3190 Franklin, S. M., Kravchenko, A. N., Vargas, R., Vasilas, B., Fuhrmann, J. J., & Jin, Y. (2021).
3191 The unexplored role of preferential flow in soil carbon dynamics. *Soil Biology and*
3192 *Biochemistry*, *161*, 108398. <https://doi.org/10.1016/j.soilbio.2021.108398>
- 3193 Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., Cadule, P., Doney, S.,
3194 Eby, M., Fung, I., Bala, G., John, J., Jones, C., Joos, F., Kato, T., Kawamiya, M., Knorr,
3195 W., Lindsay, K., Matthews, H. D., ... Zeng, N. (2006). Climate–Carbon Cycle Feedback
3196 Analysis: Results from the C4MIP Model Intercomparison. *Journal of Climate*, *19*(14),
3197 3337–3353. <https://doi.org/10.1175/JCLI3800.1>
- 3198 Friedlingstein, P., Jones, M. W., O’Sullivan, M., Andrew, R. M., Bakker, D. C. E., Hauck, J.,
3199 Le Quéré, C., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Canadell, J. G., Ciais, P.,
3200 Jackson, R. B., Alin, S. R., Anthoni, P., Bates, N. R., Becker, M., Bellouin, N., ... Zeng, J.
3201 (2022). Global Carbon Budget 2021. *Earth System Science Data*, *14*(4), 1917–2005.
3202 <https://doi.org/10.5194/essd-14-1917-2022>
- 3203 García-Palacios, P., Bradford, M. A., Benavente-Ferraces, I., de Celis, M., Delgado-Baquerizo,
3204 M., García-Gil, J. C., Gaitán, J. J., Goñi-Urtiaga, A., Mueller, C. W., Panettieri, M., Rey,
3205 A., Sáez-Sandino, T., Schuur, E. A. G., Sokol, N. W., Tedersoo, L., & Plaza, C. (2024).
3206 Dominance of particulate organic carbon in top mineral soils in cold regions. *Nature*
3207 *Geoscience*, *17*(2), 145–150. <https://doi.org/10.1038/s41561-023-01354-5>

- 3208 García-Palacios, P., Crowther, T. W., Dacal, M., Hartley, I. P., Reinsch, S., Rinnan, R., Rousk,
3209 J., van den Hoogen, J., Ye, J. S., & Bradford, M. A. (2021). Evidence for large microbial-
3210 mediated losses of soil carbon under anthropogenic warming. *Nature Reviews Earth and*
3211 *Environment*, *2*(7), 507–517. <https://doi.org/10.1038/s43017-021-00178-4>
- 3212 Georgiou, K., Jackson, R. B., Vindušková, O., Abramoff, R. Z., Ahlström, A., Feng, W., Harden,
3213 J. W., Pellegrini, A. F. A., Polley, H. W., Soong, J. L., Riley, W. J., & Torn, M. S. (2022).
3214 Global stocks and capacity of mineral-associated soil organic carbon. *Nature*
3215 *Communications*, *13*(1), 1–12. <https://doi.org/10.1038/s41467-022-31540-9>
- 3216 Gomes, L. C., Faria, R. M., de Souza, E., Veloso, G. V., Schaefer, C. E. G. R., & Filho, E. I. F.
3217 (2019). Modelling and mapping soil organic carbon stocks in Brazil. *Geoderma*,
3218 *340*(January), 337–350. <https://doi.org/10.1016/j.geoderma.2019.01.007>
- 3219 Guenet, B., Orliac, J., Cécillon, L., Torres, O., Sereni, L., Martin, P. A., Barré, P., & Bopp, L.
3220 (2024). Spatial biases reduce the ability of Earth system models to simulate soil
3221 heterotrophic respiration fluxes. *Biogeosciences*, *21*(2), 657–669.
3222 <https://doi.org/10.5194/bg-21-657-2024>
- 3223 Guo, X., Gao, Q., Yuan, M., Wang, G., Zhou, X., Feng, J., Shi, Z., Hale, L., Wu, L., Zhou, A.,
3224 Tian, R., Liu, F., Wu, B., Chen, L., Jung, C. G., Niu, S., Li, D., Xu, X., Jiang, L., ... Zhou,
3225 J. (2020). Gene-informed decomposition model predicts lower soil carbon loss due to
3226 persistent microbial adaptation to warming. *Nature Communications*, *11*(1), 4897.
3227 <https://doi.org/10.1038/s41467-020-18706-z>
- 3228 Haaf, D., Six, J., & Doetterl, S. (2021). Global patterns of geo-ecological controls on the
3229 response of soil respiration to warming. *Nature Climate Change*, *11*(7), 623–627.
3230 <https://doi.org/10.1038/s41558-021-01068-9>

- 3231 Hartley, I. P., Hill, T. C., Chadburn, S. E., & Hugelius, G. (2021). Temperature effects on carbon
3232 storage are controlled by soil stabilisation capacities. *Nature Communications*, *12*(1),
3233 6713. <https://doi.org/10.1038/s41467-021-27101-1>
- 3234 Hashimoto, S., Carvalhais, N., Ito, A., Migliavacca, M., Nishina, K., & Reichstein, M. (2015).
3235 Global spatiotemporal distribution of soil respiration modeled using a global database.
3236 *Biogeosciences*, *12*(13), 4121–4132. <https://doi.org/10.5194/bg-12-4121-2015>
- 3237 Hashimoto, S., Ito, A., & Nishina, K. (2023). Divergent data-driven estimates of global soil
3238 respiration. *Communications Earth and Environment*, *4*(1), 460.
3239 <https://doi.org/10.1038/s43247-023-01136-2>
- 3240 Heijmans, M. M. P. D., Magnússon, R., Lara, M. J., Frost, G. V., Myers-Smith, I. H., van
3241 Huissteden, J., Jorgenson, M. T., Fedorov, A. N., Epstein, H. E., Lawrence, D. M., &
3242 Limpens, J. (2022). Tundra vegetation change and impacts on permafrost. *Nature Reviews*
3243 *Earth and Environment*, *3*(1), 68–84. <https://doi.org/10.1038/s43017-021-00233-0>
- 3244 Helbig, M., Živković, T., Alekseychik, P., Aurela, M., El-Madany, T. S., Euskirchen, E. S.,
3245 Flanagan, L. B., Griffis, T. J., Hanson, P. J., Hattakka, J., Helfter, C., Hirano, T.,
3246 Humphreys, E. R., Kiely, G., Kolka, R. K., Laurila, T., Leahy, P. G., Lohila, A.,
3247 Mammarella, I., ... Zaehle, S. (2022). Warming response of peatland CO₂ sink is sensitive
3248 to seasonality in warming trends. *Nature Climate Change*, *12*(8), 743–749.
3249 <https://doi.org/10.1038/s41558-022-01428-z>
- 3250 Hensgens, G., Laudon, H., Johnson, M. S., & Berggren, M. (2021). The undetected loss of aged
3251 carbon from boreal mineral soils. *Scientific Reports*, *11*(1), 6202.
3252 <https://doi.org/10.1038/s41598-021-85506-w>
- 3253 Hijmans, R. J. (2023). *terra: Spatial Data Analysis* (1.7-39). <https://rspatial.org/terra/>
- 3254 Huang, Y., Ciais, P., Luo, Y., Zhu, D., Wang, Y., Qiu, C., Goll, D. S., Guenet, B., Makowski, D.,
3255 De Graaf, I., Leifeld, J., Kwon, M. J., Hu, J., & Qu, L. (2021). Tradeoff of CO₂ and CH₄

- 3256 emissions from global peatlands under water-table drawdown. *Nature Climate Change*,
3257 *11*(7), 618–622. <https://doi.org/10.1038/s41558-021-01059-w>
- 3258 Hugelius, G., Loisel, J., Chadburn, S., Jackson, R. B., Jones, M., MacDonald, G., Marushchak,
3259 M., Olefeldt, D., Packalen, M., Siewert, M. B., Treat, C., Turetsky, M., Voigt, C., & Yu, Z.
3260 (2020). Large stocks of peatland carbon and nitrogen are vulnerable to permafrost thaw.
3261 *Proceedings of the National Academy of Sciences of the United States of America*, *117*(34),
3262 20438–20446. <https://doi.org/10.1073/pnas.1916387117>
- 3263 Hugelius, G., Strauss, J., Zubrzycki, S., Harden, J. W., Schuur, E. A. G., Ping, C. L.,
3264 Schirrmeister, L., Grosse, G., Michaelson, G. J., Koven, C. D., O'Donnell, J. A., Elberling,
3265 B., Mishra, U., Camill, P., Yu, Z., Palmtag, J., & Kuhry, P. (2014). Estimated stocks of
3266 circumpolar permafrost carbon with quantified uncertainty ranges and identified data gaps.
3267 *Biogeosciences*, *11*(23), 6573–6593. <https://doi.org/10.5194/bg-11-6573-2014>
- 3268 Hursh, A., Ballantyne, A., Cooper, L., Maneta, M., Kimball, J., & Watts, J. (2017). The
3269 sensitivity of soil respiration to soil temperature, moisture, and carbon supply at the global
3270 scale. *Global Change Biology*, *23*(5), 2090–2103. <https://doi.org/10.1111/gcb.13489>
- 3271 Ito, A., & Wagai, R. (2017). Global distribution of clay-size minerals on land surface for
3272 biogeochemical and climatological studies. *Scientific Data*, *4*(1), 170103.
3273 <https://doi.org/10.1038/sdata.2017.103>
- 3274 Jian, J., Steele, M. K., Day, S. D., & Thomas, R. Q. (2018). Future Global Soil Respiration
3275 Rates Will Swell Despite Regional Decreases in Temperature Sensitivity Caused by Rising
3276 Temperature. *Earth's Future*, *6*(11), 1539–1554. <https://doi.org/10.1029/2018EF000937>
- 3277 Jian, J., Vargas, R., Anderson-Teixeira, K., Stell, E., Herrmann, V., Horn, M., Kholod, N.,
3278 Manzon, J., Marchesi, R., Paredes, D., & Bond-Lamberty, B. (2021). A restructured and
3279 updated global soil respiration database (SRDB-V5). *Earth System Science Data*, *13*(2),
3280 255–267. <https://doi.org/10.5194/essd-13-255-2021>

- 3281 Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann,
3282 N. E., Linder, H. P., & Kessler, M. (2017). Climatologies at high resolution for the earth's
3283 land surface areas. *Scientific Data*, 4(1), 170122. <https://doi.org/10.1038/sdata.2017.122>
- 3284 Kassambara, A. (2023). *ggpubr: "ggplot2" Based Publication Ready Plots* (0.6.0).
3285 <https://cran.r-project.org/package=ggpubr>
- 3286 Knorr, W., Prentice, I. C., House, J. I., & Holland, E. A. (2005). Long-term sensitivity of soil
3287 carbon turnover to warming. *Nature*, 433(7023), 298–301.
3288 <https://doi.org/10.1038/nature03226>
- 3289 Kögel-Knabner, I., & Amelung, W. (2021). Soil organic matter in major pedogenic soil groups.
3290 *Geoderma*, 384, 114785. <https://doi.org/10.1016/j.geoderma.2020.114785>
- 3291 Kuhn, M. (2022). *caret: Classification and Regression Training* (6.0-91).
3292 <https://doi.org/10.1887/0750303123/b365c43>
- 3293 Lehmann, A., Zheng, W., Ryo, M., Soutschek, K., Roy, J., Rongstock, R., Maaß, S., & Rillig,
3294 M. C. (2020). Fungal Traits Important for Soil Aggregation. *Frontiers in Microbiology*,
3295 10, 495825. <https://doi.org/10.3389/fmicb.2019.02904>
- 3296 Lei, J., Guo, X., Zeng, Y., Zhou, J., Gao, Q., & Yang, Y. (2021). Temporal changes in global
3297 soil respiration since 1987. *Nature Communications*, 12(1), 403.
3298 <https://doi.org/10.1038/s41467-020-20616-z>
- 3299 Leon, E., Vargas, R., Bullock, S., Lopez, E., Panosso, A. R., & La Scala, N. (2014). Hot spots,
3300 hot moments, and spatio-temporal controls on soil CO₂ efflux in a water-limited
3301 ecosystem. *Soil Biology and Biochemistry*, 77, 12–21.
3302 <https://doi.org/10.1016/j.soilbio.2014.05.029>
- 3303 Lu, H., Li, S., Ma, M., Bastrikov, V., Chen, X., Ciais, P., Dai, Y., Ito, A., Ju, W., Lienert, S.,
3304 Lombardozzi, D., Lu, X., Maignan, F., Nakhavali, M., Quine, T., Schindlbacher, A., Wang,
3305 J., Wang, Y., Wrlind, D., ... Yuan, W. (2021). Comparing machine learning-derived global

- 3306 estimates of soil respiration and its components with those from terrestrial ecosystem
3307 models. *Environmental Research Letters*, 16(5), 54048. [https://doi.org/10.1088/1748-](https://doi.org/10.1088/1748-9326/abf526)
3308 [9326/abf526](https://doi.org/10.1088/1748-9326/abf526)
- 3309 Lugato, E., Lavallee, J. M., Haddix, M. L., Panagos, P., & Cotrufo, M. F. (2021). Different
3310 climate sensitivity of particulate and mineral-associated soil organic matter. *Nature*
3311 *Geoscience*, 14(5), 295–300. <https://doi.org/10.1038/s41561-021-00744-x>
- 3312 Luo, Y., Ahlström, A., Allison, S. D., Batjes, N. H., Brovkin, V., Carvalhais, N., Chappell, A.,
3313 Ciais, P., Davidson, E. A., Finzi, A., Georgiou, K., Guenet, B., Hararuk, O., Harden, J. W.,
3314 He, Y., Hopkins, F., Jiang, L., Koven, C., Jackson, R. B., ... Zhou, T. (2016). Toward more
3315 realistic projections of soil carbon dynamics by Earth system models. *Global*
3316 *Biogeochemical Cycles*, 30(1), 40–56. <https://doi.org/10.1002/2015GB005239>
- 3317 McDowell, N., Allen, C. D., Anderson-Teixeira, K., Brando, P., Brienen, R., Chambers, J.,
3318 Christoffersen, B., Davies, S., Doughty, C., Duque, A., Espirito-Santo, F., Fisher, R.,
3319 Fontes, C. G., Galbraith, D., Goodsman, D., Grossiord, C., Hartmann, H., Holm, J.,
3320 Johnson, D. J., ... Xu, X. (2018). Drivers and mechanisms of tree mortality in moist
3321 tropical forests. *New Phytologist*, 219(3), 851–869. <https://doi.org/10.1111/nph.15027>
- 3322 Meinshausen, N. (2017). *quantregForest: Quantile Regression Forests* (1.3-7). [https://cran.r-](https://cran.r-project.org/package=quantregForest)
3323 [project.org/package=quantregForest](https://cran.r-project.org/package=quantregForest)
- 3324 Mekonnen, Z. A., Riley, W. J., Berner, L. T., Bouskill, N. J., Torn, M. S., Iwahana, G., Breen,
3325 A. L., Myers-Smith, I. H., Criado, M. G., Liu, Y., Euskirchen, E. S., Goetz, S. J., Mack,
3326 M. C., & Grant, R. F. (2021). Arctic tundra shrubification: a review of mechanisms and
3327 impacts on ecosystem carbon balance. *Environmental Research Letters*, 16(5), 53001.
3328 <https://doi.org/10.1088/1748-9326/abf28b>
- 3329 Melillo, J. M., Frey, S. D., DeAngelis, K. M., Werner, W. J., Bernard, M. J., Bowles, F. P., Pold,
3330 G., Knorr, M. A., & Grandy, A. S. (2017). Long-term pattern and magnitude of soil carbon

- 3331 feedback to the climate system in a warming world. *Science*, 358(6359), 101–105.
3332 <https://doi.org/10.1126/science.aan2874>
- 3333 Meyer, H., & Pebesma, E. (2022). Machine learning-based global maps of ecological variables
3334 and the challenge of assessing them. *Nature Communications*, 13(1), 1–4.
3335 <https://doi.org/10.1038/s41467-022-29838-9>
- 3336 Microsoft Corporation, & Weston, S. (2022). *doParallel: Foreach Parallel Adaptor for the*
3337 *“parallel” Package* (1.0.17). <https://cran.r-project.org/package=doParallel>
- 3338 Moquedace, C. M., Baldi, C. G. O., Siqueira, R. G., Cardoso, I. M., Souza, E. F. M. de, Fontes,
3339 R. L. F., Francelino, M. R., Gomes, L. C., & Fernandes-Filho, E. I. (2024). High-resolution
3340 mapping of soil carbon stocks in the western Amazon. *Geoderma Regional*, 36, e00773.
3341 <https://doi.org/10.1016/j.geodrs.2024.e00773>
- 3342 Moreno-Martínez, Á., Camps-Valls, G., Kattge, J., Robinson, N., Reichstein, M., van Bodegom,
3343 P., Kramer, K., Cornelissen, J. H. C., Reich, P., Bahn, M., Niinemets, Ü., Peñuelas, J.,
3344 Craine, J. M., Cerabolini, B. E. L., Minden, V., Laughlin, D. C., Sack, L., Allred, B.,
3345 Baraloto, C., ... Running, S. W. (2018). A methodology to derive global maps of leaf traits
3346 using remote sensing and climate data. *Remote Sensing of Environment*, 218, 69–88.
3347 <https://doi.org/10.1016/j.rse.2018.09.006>
- 3348 Naidu, D. G. T., & Bagchi, S. (2021). Greening of the earth does not compensate for rising soil
3349 heterotrophic respiration under climate change. *Global Change Biology*, 27(10), 2029–
3350 2038. <https://doi.org/10.1111/gcb.15531>
- 3351 Nissan, A., Alcolombri, U., Peleg, N., Galili, N., Jimenez-Martinez, J., Molnar, P., & Holzner,
3352 M. (2023). Global warming accelerates soil heterotrophic respiration. *Nature*
3353 *Communications*, 14(1), 3452. <https://doi.org/10.1038/s41467-023-38981-w>

- 3354 Nottingham, A. T., Meir, P., Velasquez, E., & Turner, B. L. (2020). Soil carbon loss by
3355 experimental warming in a tropical forest. *Nature*, *584*(7820), 234–237.
3356 <https://doi.org/10.1038/s41586-020-2566-4>
- 3357 Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N.,
3358 Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J.,
3359 Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., &
3360 Kassem, K. R. (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth:
3361 A new global map of terrestrial ecoregions provides an innovative tool for conserving
3362 biodiversity. *BioScience*, *51*(11), 933–938. [https://doi.org/https://doi.org/10.1641/0006-](https://doi.org/https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2)
3363 [3568\(2001\)051\[0933:TEOTWA\]2.0.CO;2](https://doi.org/https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2)
- 3364 O'Neill, B. C., Kriegler, E., Riahi, K., Ebi, K. L., Hallegatte, S., Carter, T. R., Mathur, R., &
3365 van Vuuren, D. P. (2014). A new scenario framework for climate change research: The
3366 concept of shared socioeconomic pathways. *Climatic Change*, *122*(3), 387–400.
3367 <https://doi.org/10.1007/s10584-013-0905-2>
- 3368 Patel, K. F., Fansler, S. J., Campbell, T. P., Bond-Lamberty, B., Smith, A. P., Roychowdhury,
3369 T., McCue, L. A., Varga, T., & Bailey, V. L. (2021). Soil texture and environmental
3370 conditions influence the biogeochemical responses of soils to drought and flooding.
3371 *Communications Earth and Environment*, *2*(1), 127. [https://doi.org/10.1038/s43247-021-](https://doi.org/10.1038/s43247-021-00198-4)
3372 [00198-4](https://doi.org/10.1038/s43247-021-00198-4)
- 3373 Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The*
3374 *R Journal*, *10*(1), 439. <https://doi.org/10.32614/RJ-2018-009>
- 3375 Phillips, C. A., & Wurzburger, N. (2019). Elevated rates of heterotrophic respiration in shrub-
3376 conditioned arctic tundra soils. *Pedobiologia*, *72*, 8–15.
3377 <https://doi.org/10.1016/j.pedobi.2018.11.002>

- 3378 Plaza, C., García-Palacios, P., Berhe, A. A., Barquero, J., Bastida, F., Png, G. K., Rey, A.,
3379 Bardgett, R. D., & Delgado-Baquerizo, M. (2022). Ecosystem productivity has a stronger
3380 influence than soil age on surface soil carbon storage across global biomes.
3381 *Communications Earth and Environment*, 3(1), 1–8. [https://doi.org/10.1038/s43247-022-](https://doi.org/10.1038/s43247-022-00567-7)
3382 00567-7
- 3383 Poggio, L., De Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., &
3384 Rossiter, D. (2021). SoilGrids 2.0: Producing soil information for the globe with quantified
3385 spatial uncertainty. *Soil*, 7(1), 217–240. <https://doi.org/10.5194/soil-7-217-2021>
- 3386 Quan, Q., Tian, D., Luo, Y., Zhang, F., Crowther, T. W., Zhu, K., Chen, H. Y. H., Zhou, Q., &
3387 Niu, S. (2019). Water scaling of ecosystem carbon cycle feedback to climate warming.
3388 *Science Advances*, 5(8). <https://doi.org/10.1126/sciadv.aav1131>
- 3389 R Core Team. (2025). R: A Language and Environment for Statistical Computing. Em *R*
3390 *Foundation for Statistical Computing* (4.4.3). <http://www.r-project.org>
- 3391 Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S.,
3392 Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., &
3393 Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial,
3394 hierarchical, or phylogenetic structure. *Ecography*, 40(8), 913–929.
3395 <https://doi.org/10.1111/ecog.02881>
- 3396 Ruehr, S., Keenan, T. F., Williams, C., Zhou, Y., Lu, X., Bastos, A., Canadell, J. G., Prentice, I.
3397 C., Sitch, S., & Terrer, C. (2023). Evidence and attribution of the enhanced land carbon
3398 sink. *Nature Reviews Earth and Environment*, 4(8), 518–534.
3399 <https://doi.org/10.1038/s43017-023-00456-3>
- 3400 Sáez-Sandino, T., García-Palacios, P., Maestre, F. T., Plaza, C., Guirado, E., Singh, B. K., Wang,
3401 J., Cano-Díaz, C., Eisenhauer, N., Gallardo, A., & Delgado-Baquerizo, M. (2023). The soil
3402 microbiome governs the response of microbial respiration to warming across the globe.

- 3403 *Nature Climate Change*, 13(12), 1382–1387. <https://doi.org/10.1038/s41558-023-01868->
3404 1
- 3405 Shangguan, W., Dai, Y., Duan, Q., Liu, B., & Yuan, H. (2014). A global soil data set for earth
3406 system modeling. *Journal of Advances in Modeling Earth Systems*, 6(1), 249–263.
3407 <https://doi.org/10.1002/2013MS000293>
- 3408 Shukla, P. R., Skea, J., Calvo Buendia, E., Masson-Delmotte, V., & [P.R. Shukla, J. Skea, E.
3409 Calvo Buendia, V. Masson-Delmotte, H.- O. Pörtner, D. C. Roberts, P. Zhai, R. Slade, S.
3410 Connors, R. van Diemen, M. Ferrat, E. Haughey, S. Luz, S. Neogi, M. Pathak, J. Petzold,
3411 J. Portugal Pereira, P. Vyas, E. Huntley, K. Kissick, J. M. (2007). IPCC, 2019: Climate
3412 Change and Land: an IPCC special report on climate change, desertification, land
3413 degradation, sustainable land management, food security, and greenhouse gas fluxes in
3414 terrestrial ecosystems. Em *Climate change 2007: Impacts, adaptation and vulnerability.*
3415 *Contribution of Working Group II to the Fourth Assessment Report of the*
3416 *Intergovernmental Panel of Climate Change (IPCC)* (Números vii, 973). www.ipcc.ch
- 3417 Signorell, A. (2023). *DescTools: Tools for Descriptive Statistics* (0.99.49).
- 3418 Singh, S., Mayes, M. A., Shekoofa, A., Kivlin, S. N., Bansal, S., & Jagadamma, S. (2021). Soil
3419 organic carbon cycling in response to simulated soil moisture variation under field
3420 conditions. *Scientific Reports*, 11(1), 10841. <https://doi.org/10.1038/s41598-021-90359-4>
- 3421 Siqueira, R. G., Moquedace, C. M., Fernandes-Filho, E. I., Schaefer, C. E. G. R., Francelino,
3422 M. R., Sacramento, I. F., & Michel, R. F. M. (2024). Modelling and prediction of major
3423 soil chemical properties with Random Forest: Machine learning as tool to understand soil-
3424 environment relationships in Antarctica. *Catena*, 235, 107677.
3425 <https://doi.org/10.1016/j.catena.2023.107677>
- 3426 Siqueira, R. G., Moquedace, C. M., Francelino, M. R., Schaefer, C. E. G. R., & Fernandes-
3427 Filho, E. I. (2023). Machine learning applied for Antarctic soil mapping: Spatial prediction

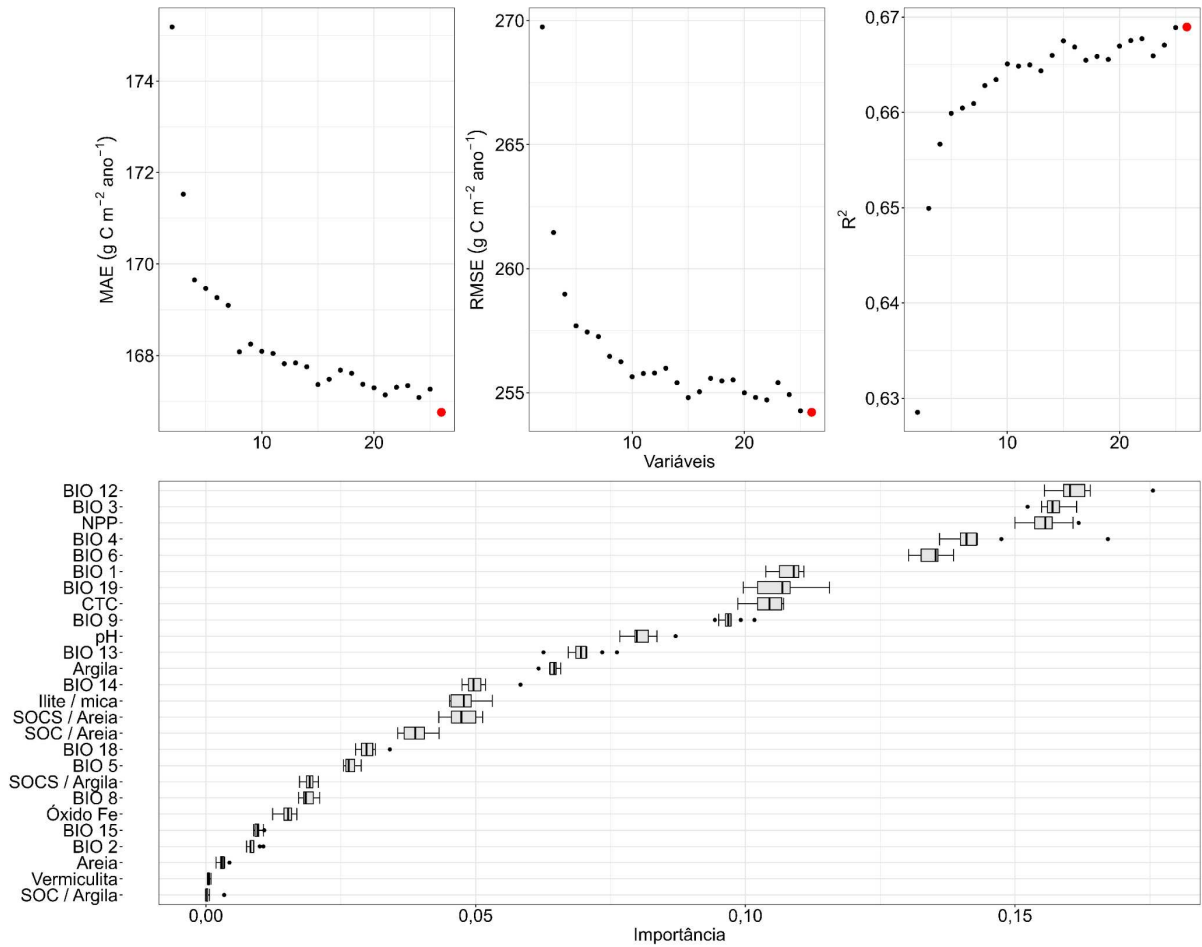
- 3428 of soil texture for Maritime Antarctica and Northern Antarctic Peninsula. *Geoderma*, 432,
3429 116405. <https://doi.org/10.1016/j.geoderma.2023.116405>
- 3430 Soong, J. L., Castanha, C., Hicks Pries, C. E., Ofiti, N., Porras, R. C., Riley, W. J., Schmidt, M.
3431 W. I., & Torn, M. S. (2021). Five years of whole-soil warming led to loss of subsoil carbon
3432 stocks and increased CO₂ efflux. *Science Advances*, 7(21).
3433 <https://doi.org/10.1126/sciadv.abd1343>
- 3434 Stell, E., Warner, D., Jian, J., Bond-Lamberty, B., & Vargas, R. (2021). Spatial biases of
3435 information influence global estimates of soil respiration: How can we improve global
3436 predictions? *Global Change Biology*, 27(16), 3923–3938.
3437 <https://doi.org/10.1111/gcb.15666>
- 3438 Sulman, B. N., Phillips, R. P., Oishi, A. C., Shevliakova, E., & Pacala, S. W. (2014). Microbe-
3439 driven turnover offsets mineral-mediated storage of soil carbon under elevated CO₂.
3440 *Nature Climate Change* 2014 4:12, 4(12), 1099–1102.
3441 <https://doi.org/10.1038/nclimate2436>
- 3442 Tamocai, C., Canadell, J. G., Schuur, E. A. G., Kuhry, P., Mazhitova, G., & Zimov, S. (2009).
3443 Soil organic carbon pools in the northern circumpolar permafrost region. *Global*
3444 *Biogeochemical Cycles*, 23(2). <https://doi.org/10.1029/2008GB003327>
- 3445 Tang, X., Pei, X., Lei, N., Luo, X., Liu, L., Shi, L., Chen, G., & Liang, J. (2020). Global patterns
3446 of soil autotrophic respiration and its relation to climate, soil and vegetation
3447 characteristics. *Geoderma*, 369, 114339. <https://doi.org/10.1016/j.geoderma.2020.114339>
- 3448 Tennekes, M. (2018). tmap: Thematic Maps in R. *Journal of Statistical Software*, 84(6).
3449 <https://doi.org/10.18637/jss.v084.i06>
- 3450 van den Hoogen, J., Geisen, S., Routh, D., Ferris, H., Traunspurger, W., Wardle, D. A., de
3451 Goede, R. G. M., Adams, B. J., Ahmad, W., Andriuzzi, W. S., Bardgett, R. D., Bonkowski,
3452 M., Campos-Herrera, R., Cares, J. E., Caruso, T., de Brito Caixeta, L., Chen, X., Costa, S.

- 3453 R., Creamer, R., ... Crowther, T. W. (2019). Soil nematode abundance and functional
3454 group composition at a global scale. *Nature*, 572(7768), 194–198.
3455 <https://doi.org/10.1038/s41586-019-1418-6>
- 3456 Varney, R. M., Chadburn, S. E., Friedlingstein, P., Burke, E. J., Koven, C. D., Hugelius, G., &
3457 Cox, P. M. (2020). A spatial emergent constraint on the sensitivity of soil carbon turnover
3458 to global warming. *Nature Communications*, 11(1), 5544. [https://doi.org/10.1038/s41467-](https://doi.org/10.1038/s41467-020-19208-8)
3459 [020-19208-8](https://doi.org/10.1038/s41467-020-19208-8)
- 3460 Wadoux, A. M. J. C., Samuel-Rosa, A., Poggio, L., & Mulder, V. L. (2020). A note on
3461 knowledge discovery and machine learning in digital soil mapping. *European Journal of*
3462 *Soil Science*, 71(2), 133–136. <https://doi.org/10.1111/ejss.12909>
- 3463 Warner, D. L., Bond-Lamberty, B., Jian, J., Stell, E., & Vargas, R. (2019). Spatial Predictions
3464 and Associated Uncertainty of Annual Soil Respiration at the Global Scale. *Global*
3465 *Biogeochemical Cycles*, 33(12), 1733–1745. <https://doi.org/10.1029/2019GB006264>
- 3466 Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., & Schewe, J. (2014). The
3467 inter-sectoral impact model intercomparison project (ISI-MIP): Project framework.
3468 *Proceedings of the National Academy of Sciences of the United States of America*, 111(9),
3469 3228–3232. <https://doi.org/10.1073/pnas.1312330110>
- 3470 Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. *Em Journal of Statistical*
3471 *Software* (Vol. 35, Número Book Review 1). Springer-Verlag New York.
- 3472 Wieder, W. R., Sulman, B. N., Hartman, M. D., Koven, C. D., & Bradford, M. A. (2019). Arctic
3473 Soil Governs Whether Climate Change Drives Global Losses or Gains in Soil Carbon.
3474 *Geophysical Research Letters*, 46(24), 14486–14495.
3475 <https://doi.org/10.1029/2019GL085543>
- 3476 Xu, X., Shi, Z., Li, D., Rey, A., Ruan, H., Craine, J. M., Liang, J., Zhou, J., & Luo, Y. (2016).
3477 Soil properties control decomposition of soil organic carbon: Results from data-

- 3478 assimilation analysis. *Geoderma*, 262, 235–242.
3479 <https://doi.org/10.1016/j.geoderma.2015.08.038>
- 3480 Yao, X., Goodale, U. M., Yang, M., Liao, L., Yu, S., Wang, S., & Ye, S. (2021). Bidirectional
3481 nitrogen transfer and plant growth in a mixed plantation of n₂-fixing species and
3482 eucalyptus urophylla × e. Grandis under different n applications. *Forests*, 12(9), 1171.
3483 <https://doi.org/10.3390/f12091171>
- 3484 Yu, K., Ciais, P., Seneviratne, S. I., Liu, Z., Chen, H. Y. H., Barichivich, J., Allen, C. D., Yang,
3485 H., Huang, Y., & Ballantyne, A. P. (2022). Field-based tree mortality constraint reduces
3486 estimates of model-projected forest carbon sinks. *Nature Communications*, 13(1), 2094.
3487 <https://doi.org/10.1038/s41467-022-29619-4>
- 3488 Zhang, Z., Zhang, R., Cescatti, A., Wohlfahrt, G., Buchmann, N., Zhu, J., Chen, G., Moyano,
3489 F., Pumpanen, J., Hirano, T., Takagi, K., & Merbold, L. (2017). Effect of climate warming
3490 on the annual terrestrial net ecosystem CO₂ exchange globally in the boreal and temperate
3491 regions. *Scientific Reports*, 7(1), 3108. <https://doi.org/10.1038/s41598-017-03386-5>
- 3492 Zhao, Z., Peng, C., Yang, Q., Meng, F. R., Song, X., Chen, S., Epule, T. E., Li, P., & Zhu, Q.
3493 (2017). Model prediction of biome-specific global soil respiration from 1960 to 2012.
3494 *Earth's Future*, 5(7), 715–729. <https://doi.org/10.1002/2016EF000480>
- 3495 Zheng, P., Wang, D., Yu, X., Jia, G., Liu, Z., Wang, Y., & Zhang, Y. (2021). Effects of drought
3496 and rainfall events on soil autotrophic respiration and heterotrophic respiration.
3497 *Agriculture, Ecosystems and Environment*, 308, 107267.
3498 <https://doi.org/10.1016/j.agee.2020.107267>

3499

Material Suplementar



3500
3501
3502

Figura S1. Variáveis selecionadas pelo RFE variáveis usadas para modelar a respiração total do solo usando todas as covariáveis do solo.

3503
3504**Tabela S1.** Estatísticas descritivas para a respiração total anual do solo (Rs) e a respiração heterotrófica do solo (Rh) ponderada por área para o período passado-presente para cada bioma global usando o subconjunto selecionado de variáveis. Os valores são apresentados em $\text{gC m}^{-2} \text{ano}^{-1}$.

Biomassas	Min Rh	Min Rs	Média Rh	Média Rs	Rh/Rs	Mediana Rh	Mediana Rs	Max Rh	Max Rs	CV Rh	CV Rs	Soma Rh (PgC)	Soma Rs (PgC)
Florestas Boreais/Taiga	51	183	290	463	0,63	290	444	734	928	0,42	0,23	4,10	6,55
Desertos e Arbustais Xerófilos	184	88	343	392	0,87	332	318	1052	1989	0,25	0,57	9,52	10,91
Savanas e Campos Inundados	184	144	541	875	0,62	594	926	975	1862	0,34	0,40	0,54	0,88
Lagos	100	260	394	655	0,60	382	636	1001	1821	0,45	0,47	0,02	0,04
Manguezais	329	283	677	1318	0,51	682	1361	1034	2080	0,15	0,20	0,15	0,30
Florestas, Bosques e Arbustais Mediterrâneos	229	122	372	621	0,60	356	571	1050	1849	0,21	0,36	1,16	1,94
Campos e Arbustais Montanos	177	102	431	683	0,63	405	659	987	2008	0,29	0,35	2,18	3,47
Rochas e geleiras	87	221	362	555	0,65	406	569	643	1109	0,38	0,26	0,04	0,06
Florestas Temperadas de Folhas Largas e Mistas	127	40	429	732	0,59	420	714	1480	2686	0,25	0,22	5,17	8,82
Florestas de Coníferas Temperadas	126	98	424	718	0,59	432	684	910	3059	0,21	0,27	1,66	2,80
Campos Temperados, Savanas e Arbustais	182	74	345	574	0,60	308	525	1004	2078	0,29	0,34	3,39	5,65
Florestas de Coníferas Tropicais e Subtropicais	224	264	578	937	0,62	593	911	948	1911	0,25	0,19	0,39	0,64
Florestas Tropicais e Subtropicais Secas de Folhas Largas	297	260	667	1084	0,62	674	1048	1020	2603	0,18	0,22	1,95	3,17
Savanas e Arbustais Tropicais e Subtropicais	230	176	610	972	0,63	612	952	1155	2354	0,27	0,30	12,32	19,64
Florestas Tropicais e Subtropicais Úmidas de Folhas Largas	197	154	713	1243	0,57	717	1279	1117	2781	0,18	0,23	13,77	24,00
Tundra	61	133	218	396	0,55	179	385	579	985	0,50	0,20	1,50	2,73

3505

3506 **Artigo 4: Interações solo–biota–clima e explicabilidade em modelos globais de SOC**

3507 1. Introdução

3508 Em um momento crítico da história da Terra, no qual a estabilidade climática, a
3509 segurança alimentar e a conservação da biodiversidade convergem como prioridades globais, o
3510 solo emerge como um dos bens naturais mais estratégicos e vulneráveis (A. McBratney et al.,
3511 2014). Sua degradação não é uma metáfora alarmista, mas um fato científico: a estabilidade
3512 ecológica e climática planetária está diretamente ligada à integridade dos solos (Evangelista et
3513 al., 2023). Com a capacidade de armazenar entre 1.460 e 1.550 Pg de carbono orgânico, o solo
3514 contém mais carbono do que a atmosfera e a biomassa viva combinadas (Batjes, 1996; Lal,
3515 2010), ficando atrás somente do oceano.

3516 Este reservatório subterrâneo é fundamental para regular o ciclo global do carbono,
3517 sustentar serviços ecossistêmicos essenciais, manter a multifuncionalidade do solo e amortecer
3518 os impactos das mudanças climáticas (Evangelista et al., 2024; FAO & ITPS, 2015). Nesse
3519 contexto, o solo também fundamenta conceitos contemporâneos como *soil health* (a capacidade
3520 do solo de funcionar como um ecossistema vivo) e *soil security*, que reconhece o solo como um
3521 recurso finito, estratégico e indispensável à estabilidade socioeconômica global (A. McBratney
3522 et al., 2014).

3523 No entanto, a integridade desse bem natural vem sendo comprometida em ritmo
3524 preocupante. Práticas inadequadas de uso da terra, intensificação agrícola, urbanização
3525 descontrolada e mudanças climáticas vêm degradando os solos em escala global, o que resulta
3526 na emissão anual de 0,8 a 1,9 Pg de carbono para a atmosfera (Padarian et al., 2022). Cada
3527 tonelada perdida representa não apenas a deterioração silenciosa de um recurso vital, mas o
3528 fortalecimento de um ciclo de retroalimentação climática que ameaça os sistemas de suporte à
3529 vida.

3530 Por outro lado, a gravidade da ameaça ressalta a urgência de medidas eficazes: restaurar
3531 os estoques de carbono do solo e adotar práticas que recuperem sua funcionalidade e

3532 estabilidade são estratégias fundamentais. Embora não resolvam isoladamente a crise climática,
3533 essas ações podem contribuir significativamente para a mitigação das emissões globais de
3534 carbono, com estimativas apontando um potencial de compensação de até 35% ao ano (Lal,
3535 2014; Minasny et al., 2017).

3536 Nesse cenário, a construção de inventários nacionais e globais confiáveis de carbono do
3537 solo torna-se um pré-requisito inegociável para orientar políticas públicas, compromissos
3538 climáticos internacionais e práticas agrícolas sustentáveis (Batjes et al., 2019; Friedlingstein et
3539 al., 2025). Entretanto, a acurácia dessas estimativas ainda é prejudicada por limitações
3540 metodológicas e por falhas estruturais nos bancos de dados disponíveis (Batjes, 1996; Wadoux
3541 et al., 2020).

3542 A distribuição geográfica das amostras de solo permanece fortemente enviesada para
3543 áreas agrícolas de regiões temperadas, enquanto zonas ecologicamente estratégicas, como
3544 regiões boreais, áridas, semiáridas e florestas tropicais, continuam sub-representadas (Batjes et
3545 al., 2019; Minasny et al., 2017). Essa lacuna compromete a calibração dos modelos e resulta na
3546 subestimação dos estoques de SOC em ecossistemas de alta relevância climática, como as
3547 regiões dominadas por *permafrost*.

3548 Nesse contexto, a inteligência artificial (IA) tem se consolidado como um conjunto de
3549 abordagens promissoras para lidar com a complexidade e a alta dimensionalidade dos dados
3550 ambientais. Embora concebida nos anos 1950 com o objetivo de simular processos cognitivos
3551 humanos (Turing, 2009), a IA passou por um ressurgimento expressivo no século XXI,
3552 especialmente na ciência do solo, impulsionado pela ampliação do poder computacional, pela
3553 disponibilidade de grandes bases de dados e pelo avanço de algoritmos como as redes neurais
3554 profundas (Minasny & McBratney, 2025; Wadoux, 2025). O reconhecimento dessas inovações
3555 foi consolidado pelo Prêmio Nobel de Física de 2024, concedido a John J. Hopfield e Geoffrey

3556 E. Hinton por descobertas fundamentais que embasam os modelos modernos de aprendizado
3557 profundo, hoje amplamente aplicados na ciência do solo (Minasny & McBratney, 2024).

3558 Entre as aplicações da IA, o aprendizado de máquina (ML) destacou-se por sua
3559 capacidade de identificar padrões complexos e relações ocultas em grandes conjuntos de dados
3560 (Hengl et al., 2017; Mello et al., 2025; Moquedace et al., 2024; Poggio et al., 2018; Siqueira et
3561 al., 2024). Essa abordagem tem sido largamente adotada na pedometria, especialmente para
3562 modelar atributos como o carbono orgânico do solo em escalas regionais (Moquedace et al.,
3563 2024), nacionais (Gomes et al., 2019) e globais (Hengl et al., 2017; Poggio et al., 2021),
3564 superando limitações impostas por métodos estatísticos convencionais. No entanto, sua
3565 aplicação na ciência do solo tem sido marcada por uma ênfase quase exclusiva na previsão,
3566 com integração limitada ao conhecimento pedológico e pouca capacidade explicativa (Castro
3567 et al., 2024; Wadoux, 2025).

3568 Paralelamente, os modelos baseados em ML tendem a priorizar a maximização da
3569 acurácia preditiva global, muitas vezes em detrimento da captura de variações locais e
3570 contextuais (Padarian et al., 2022). A ênfase em métricas agregadas, como MAE, RMSE e R^2 ,
3571 pode ocultar erros sistemáticos em regiões ecologicamente sensíveis e comprometer a coerência
3572 pedoecológica das projeções, entendida aqui como a capacidade dos modelos de refletir
3573 adequadamente as relações bidirecionais entre o solo e os ecossistemas que ele suporta,
3574 especialmente as interações entre solo-organismos-ambiente. Além disso, a maioria dos
3575 algoritmos são operados como “caixas-pretas”, reproduzindo padrões estatísticos sem
3576 necessariamente explicar os processos pedoecológicos subjacentes à dinâmica do SOC (Heung
3577 et al., 2016; Padarian et al., 2020).

3578 Essas limitações são especialmente preocupantes quando se considera que o controle do
3579 SOC resulta de interações complexas, não lineares e contexto-dependentes entre clima, relevo,
3580 propriedades do solo, organismos e manejo da terra (Lal, 2004; A. B. McBratney et al., 2003).

3581 A modelagem precisa desses estoques exige ferramentas capazes de integrar múltiplas variáveis
3582 e traduzir suas interações de forma compreensível e operacionalizável, respeitando a natureza
3583 tridimensional, heterogênea e dinâmica dos solos (Poggio & Gimona, 2014).

3584 Como resposta a esses desafios, emergem as abordagens de inteligência artificial
3585 explicável, que conciliam poder preditivo com transparência e coerência pedoecológica
3586 (Beucher et al., 2022; Chandra et al., 2023; Wadoux et al., 2023; Wadoux & Molnar, 2022).
3587 Técnicas como importância por permutação, gráficos de dependência parcial e valores de
3588 Shapley permitem interpretar a estrutura interna dos modelos, revelar hierarquias funcionais de
3589 preditores e identificar padrões espaciais de controle do SOC alinhados ao conhecimento
3590 pedoecológico (Padarian et al., 2020; Wadoux, 2025).

3591 Aliar interpretabilidade, fundamentação pedoecológica e desempenho quantitativo
3592 representa uma inflexão metodológica essencial para o avanço da modelagem de SOC. Essa
3593 mudança não apenas aprimora a capacidade preditiva dos modelos em diferentes escalas, mas
3594 também fortalece o papel da ciência do solo na formulação de estratégias ambientais baseadas
3595 em evidências, consolidando a segurança do solo e saúde do solo como pilar da sustentabilidade
3596 ecológica e da estabilidade socioeconômica global (A. McBratney et al., 2014).

3597 Diante desse panorama, este estudo propõe uma abordagem baseada em inteligência
3598 artificial explicável para integrar desempenho preditivo e coerência pedoecológica na
3599 modelagem global dos estoques de SOC. A proposta alinha-se a uma ideia de modelos que
3600 combinam aprendizado de máquina com conhecimento pedológico, promovendo o que se tem
3601 chamado de *Soil-Informed Machine Learning* (Castro et al., 2024; Minasny et al., 2024). Mais
3602 do que apenas predizer, esses modelos visam compreender os processos subjacentes. Além
3603 disso, permitem a formulação de hipóteses pedoecológicas com base em evidências
3604 quantitativas, espacialmente explícitas e fundamentadas no conhecimento do solo.

3605 Com base nessa abordagem, objetivou-se aprimorar a modelagem global dos estoques
3606 de carbono orgânico do solo por meio da integração entre aprendizado de máquina e técnicas
3607 de interpretabilidade. Especificamente, buscou-se: (i) avaliar a importância relativa e espacial
3608 dos preditores edafoclimáticos e biológicos na distribuição global do SOC; (ii) diagnosticar
3609 padrões de variação, inconsistência preditiva e lacunas de representatividade amostral; e (iii)
3610 propor uma estrutura metodológica replicável, baseada em inteligência artificial explicável, que
3611 amplie a robustez, a coerência pedoecológica e a transparência dos modelos pedométricos.

3612

3613 **2. Material e métodos**

3614 **2.1. Base de dados**

3615 Foram utilizadas duas bases de dados previamente compiladas. A primeira foi o WoSIS
3616 2023 snapshot, proveniente do World Soil Information Service, que reúne informações de 228
3617 mil perfis de solo georreferenciados, distribuídos por mais de 174 países e em diferentes
3618 profundidades (Calisto et al., 2023). A segunda base foi o GSOCS-LULCC (Global Soil
3619 Organic Carbon Stock dataset after Land Use and Land Cover Change), recentemente
3620 compilada por Chen et al. (2024). Essa base integra 5.982 registros provenientes de 1.206 locais
3621 amostrais, obtidos em diferentes profundidades e extraídos de 639 artigos científicos indexados
3622 na Web of Science até dezembro de 2023.

3623 As duas bases foram integradas e pré-processadas. Para o conjunto WoSIS, os estoques
3624 de carbono orgânico do solo (SOC) foram calculados para cada ponto que apresentava
3625 informações de densidade do solo, teor de fração grosseira e teor de SOC. No caso da base
3626 GSOCS-LULCC, os dados já estavam expressos diretamente como estoque de SOC.

3627 Adotou-se a camada de 0–30 cm como camada padrão para a análise. Considerando a
3628 heterogeneidade dos dados, oriundos de diferentes tipos de levantamentos e com profundidades
3629 não padronizadas, aplicou-se o método *mass-preserving spline* para a harmonização dos dados

3630 em profundidade. Essa padronização foi realizada com o auxílio do pacote “mp2spline”
3631 (O’Brien, 2022), em R (R Core Team, 2025). O cálculo do estoque de SOC foi realizado com
3632 base na seguinte equação (USDA, 2023):

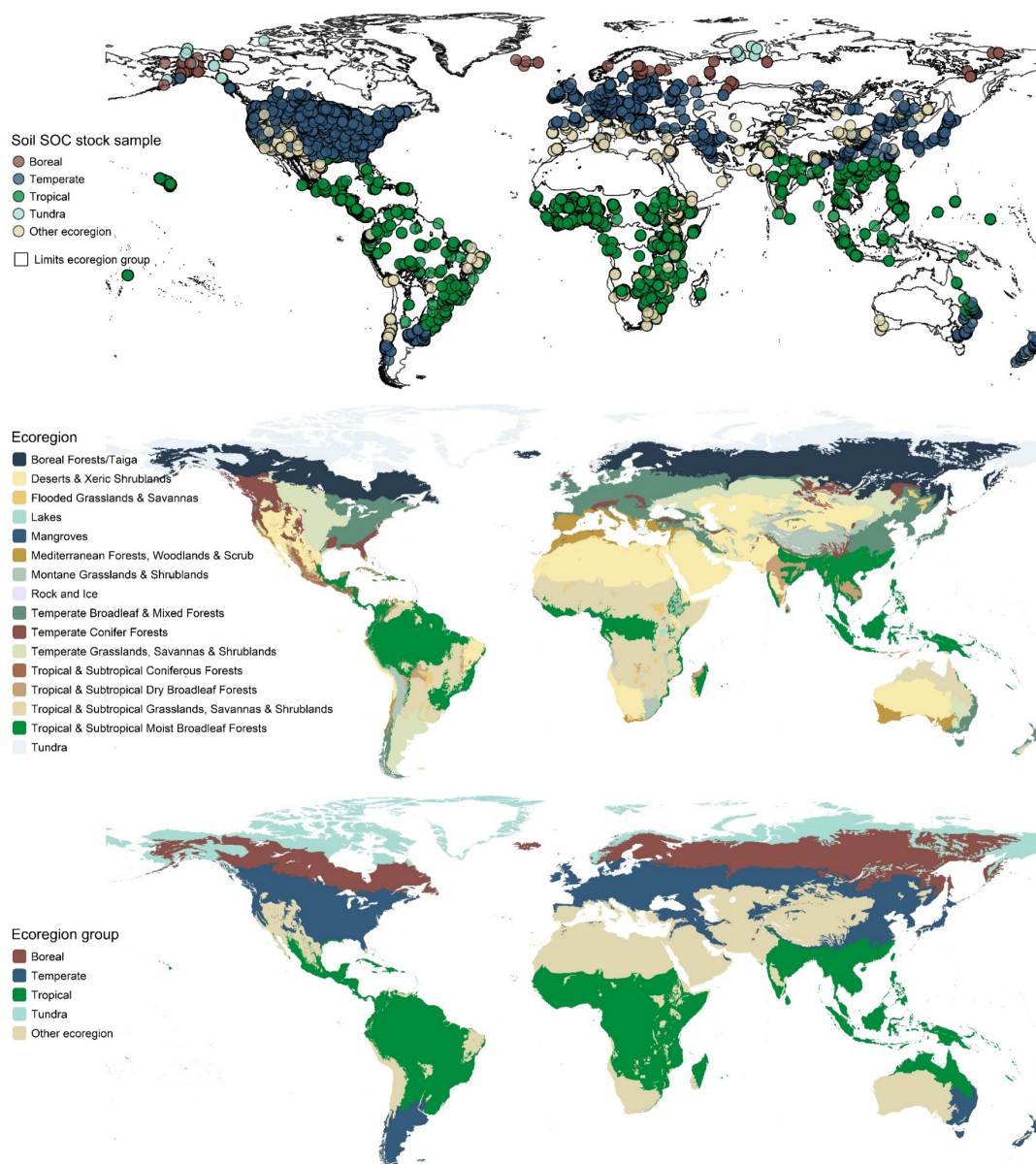
3633

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

3634 Estoque de SOC refere-se ao estoque de carbono orgânico em determinada camada, expresso em Mg ha⁻¹; SOC é
3635 o teor de carbono orgânico do solo, expresso em g kg⁻¹; BD é a densidade do solo em g cm⁻³; “e” é a espessura da
3636 camada em cm; CF representa a fração grosseira do solo (partículas > 2 mm), expressa em valor decimal; o fator
3637 10 é utilizado para conversão das unidades e obtenção do estoque final em Mg ha⁻¹.
3638

3639 Para a aplicação dos *splines*, os dados foram tratados por meio da remoção de
3640 informações espúrias, como registros com profundidades inferiores maiores que as superiores,
3641 outliers e coordenadas geográficas inconsistentes, resultando em um total de 29.153 amostras
3642 distribuídas globalmente (Figura 1).

3643



3644
 3645 **Figura 1.** Distribuição espacial das amostras (Calisto et al., 2023; S. Chen et al., 2024) de
 3646 estoque de carbono orgânico do solo (SOC) em relação às ecorregiões globais (Olson et al.,
 3647 2001) e seus grupos bioclimáticos.
 3648

3649 2.2. Preditores explicativos do estoque de SOC

3650 Para compor os preditores explicativos do estoque de SOC, foram utilizadas variáveis
 3651 bióticas e abióticas, classificadas conforme suas funções ecossistêmicas e organizadas em
 3652 domínios explicativos: clima, fungos do solo, respiração do solo, traços de raízes, atributos do
 3653 solo e topografia (Tabela 1). Todas as variáveis foram harmonizadas para 1km de resolução
 3654 espacial com o uso do pacote “terra” (Hijmans, 2023).

3655

3656 **Tabela 1.** Preditores ambientais utilizados na modelagem dos estoques de carbono orgânico do
 3657 solo (SOC), com respectivas domínios explicativos e referências.

Sigla	Preditor	Grupo	Ref.
BIO 1	Temperatura média anual	<i>Climatic conditions</i>	(Karger et al., 2017)
BIO 12	Precipitação Anual	<i>Climatic conditions</i>	(Karger et al., 2017)
SFA	Abundância de fungos no solo a 5cm	<i>Fungal attributes</i>	Modelado Art. 1
SFR	Riqueza de fungos no solo a 5cm	<i>Fungal attributes</i>	Modelado Art. 1
RCC	Concentração de carbono nas raízes	<i>Root traits</i>	Modelado Art. 2
RMC	Taxa de colonização micorrízica nas raízes	<i>Root traits</i>	Modelado Art. 2
RTR	Turnover das raízes	<i>Root traits</i>	Modelado Art. 2
SRR	Respiração específica das raízes	<i>Root traits</i>	Modelado Art. 2
CEC	Capacidade de trocar cátions	<i>Soil properties</i>	(Poggio et al., 2021)
Clay	Teor de argila	<i>Soil properties</i>	(Poggio et al., 2021)
N	Teor de nitrogênio	<i>Soil properties</i>	(Poggio et al., 2021)
pH	Potencial hidrogeniônico em água	<i>Soil properties</i>	(Poggio et al., 2021)
Sand	Teor de areia	<i>Soil properties</i>	(Poggio et al., 2021)
Rh	Respiração heterotrófica do solo	<i>Soil respiration</i>	Modelado Art. 3
Rs	Respiração total do solo	<i>Soil respiration</i>	Modelado Art. 3
DEM	Modelo digital de elevação	<i>Topographic attribute</i>	(Hawker et al., 2022)

3658

3659 2.3. Modelagem espacial

3660 Como o objetivo foi avaliar a contribuição das camadas ambientais para os estoques de
 3661 SOC, o processo de modelagem foi conduzido sem aplicação de seleção de variáveis por
 3662 importância prévia. Foram avaliadas apenas as correlações entre os preditores, com o intuito de
 3663 remover redundâncias por meio da função “findcorrelation” do pacote caret (Kuhn, 2022).

3664 Optou-se pela adoção do algoritmo quantile random forest – qrf (Meinshausen, 2017),
 3665 amplamente utilizado na literatura para o mapeamento de atributos do solo, em razão de sua
 3666 robustez e capacidade de modelar incertezas (Hengl et al., 2017; Mello et al., 2025; Moquedace
 3667 et al., 2024; Poggio et al., 2021; Siqueira et al., 2023). O processo de modelagem foi conduzido
 3668 em duas abordagens complementares: (i) a construção de um modelo global para a predição
 3669 espacial, utilizando todas as amostras disponíveis; e (ii) a avaliação de desempenho do modelo,
 3670 com divisão dos dados em conjuntos de treino e teste.

3671 Para ambas as abordagens, a parametrização seguiu os mesmos critérios: foi empregada
 3672 validação cruzada repetida, com 10 repetições de 10 *folds*, e ajuste do hiperparâmetro *mtry*,
 3673 testando-se 10 valores distintos dentro do intervalo admissível. A principal diferenciação
 3674 metodológica no modelo de avaliação de performance consistiu na divisão dos dados em treino
 3675 (80%) e teste (20%), repetindo-se esse processo 100 vezes. Essa repetição permitiu minimizar
 3676 a influência de divisões específicas dos dados e atenuar a aleatoriedade no ajuste do modelo.

3677 A performance do modelo foi avaliada por meio das seguintes métricas: erro médio
 3678 absoluto – MAE (1), raiz do erro quadrático médio – RMSE (2), coeficiente de correlação de
 3679 concordância - ρ_c (3) e coeficiente de determinação – R^2 (4). Como forma adicional de
 3680 validação, foi utilizado um modelo nulo, baseado na média global dos valores observados, cujas
 3681 métricas de erro foram calculadas e comparadas com os resultados do qrf, permitindo verificar
 3682 a efetividade do modelo em relação a uma referência (Moquedace et al., 2024; Siqueira et al.,
 3683 2023).

3684

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

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

$$\rho_c = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2 + (\mu_x - \mu_y)^2} \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (4)$$

3685 Em que: *n* representa o número de amostras; ρ_c é o coeficiente de correlação entre as duas variáveis, μ_x e μ_y são as
 3686 médias para as duas variáveis, σ_x^2 e σ_y^2 e são as variâncias correspondentes; P_i e O_i representam os valores previstos
 3687 e observados no local *i*, respectivamente e \bar{O} representa a média dos valores observados.
 3688

3689 **2.4. Contribuição dos preditores ambientais para o estoque de SOC**

3690 A importância dos preditores ambientais para os estoques de SOC foi avaliada com base
3691 em três abordagens complementares: importância relativa por permutação, curvas de
3692 dependência parcial e valores de Shapley. Essa triangulação metodológica baseou-se em captar
3693 tanto o impacto médio global das variáveis preditoras quanto os efeitos funcionais e contextuais
3694 mais complexos, o que ofereceu uma interpretação pedoecologicamente coerente dos fatores
3695 que determinam os estoques de SOC em diferentes ecorregiões.

3696 A importância por permutação foi empregada como primeira abordagem para estimar a
3697 sensibilidade do modelo em relação a cada covariável. Para isso, os valores de cada variável
3698 preditora foram aleatoriamente embaralhados, mantendo as demais constantes, e a consequente
3699 perda de desempenho preditivo do modelo foi registrada. Variáveis cuja permutação resultou
3700 em maior redução de acurácia foram consideradas mais relevantes. Os resultados foram
3701 expressos como a contribuição percentual de cada preditor para o desempenho global do
3702 modelo. Essa análise foi conduzida com a função “varImp” do pacote caret (Kuhn, 2022).

3703 As curvas de dependência parcial foram utilizadas para investigar a relação funcional
3704 média entre cada covariável e as previsões do modelo. A técnica consistiu em simular, para
3705 cada preditor, diferentes cenários em que todos os pontos da base de dados assumiram um
3706 mesmo valor fixo para a variável preditora em análise, enquanto os demais preditores foram
3707 mantidos com seus valores originais. Esse procedimento marginalizou os efeitos das outras
3708 variáveis preditoras, isolando a resposta média do modelo à variável focal (estoque de SOC).

3709 As curvas de dependência parcial são particularmente úteis para revelar efeitos não-
3710 lineares, zonas de saturação, limiares ecológicos e outras formas de relação que não são
3711 captadas por métodos de importância agregada. Essa etapa foi implementada com adaptações
3712 específicas nas funções do pacote “rfUtilities” (Evans & Murphy, 2015).

3713 Os valores de Shapley, baseados na teoria dos jogos cooperativos (Shapley, 1953), foram
3714 empregados como abordagem final para quantificar a contribuição marginal individual de cada
3715 variável preditora. Embora robustos, esses valores são computacionalmente custosos,
3716 especialmente em modelos com muitas variáveis, pois exigem a avaliação de todas as
3717 combinações possíveis de preditores. Para viabilizar o cálculo, utilizou-se uma aproximação
3718 eficiente (Monte Carlo), que estima as contribuições médias ponderadas em subconjuntos
3719 reduzidos, trade-off comum em aplicações práticas (Greenwell, 2021; Wadoux et al., 2023).

3720 Essa métrica forneceu uma medida interpretável da influência local de cada variável
3721 sobre as predições de estoque de SOC (em Mg C ha⁻¹), expressa em relação à diferença entre a
3722 predição individual e o valor médio global do modelo. Entretanto, é fundamental ressaltar que
3723 os valores de Shapley explicam o comportamento do modelo, não necessariamente relações
3724 causais no sistema pedoecológico. A aplicação foi realizada com o suporte dos pacotes fastshap
3725 (Greenwell, 2021) e shapviz (Mayer, 2025), que também viabilizaram a visualização espacial
3726 das contribuições preditivas.

3727 Diferentemente da importância por permutação e dos valores de Shapley, que indicam
3728 quanto uma variável influencia o modelo, as funções de dependência parcial revelam como essa
3729 influência ocorre, o que permitiu a interpretação ecológica da direção e forma da resposta
3730 modelada. O uso integrado dessas abordagens pode fornecer uma visão abrangente e explicável
3731 da relação entre fatores ambientais e os estoques de SOC.

3732 Todas as análises foram realizadas com o software R, com o suporte dos pacotes
3733 mencionados e, adicionalmente, mdsFuncs (Fernandes Filho et al., 2021), tmap (Tennekes,
3734 2018) e ggplot2 (Wickham, 2016) para organização dos dados, construção dos gráficos e
3735 elaboração dos mapas.

3736

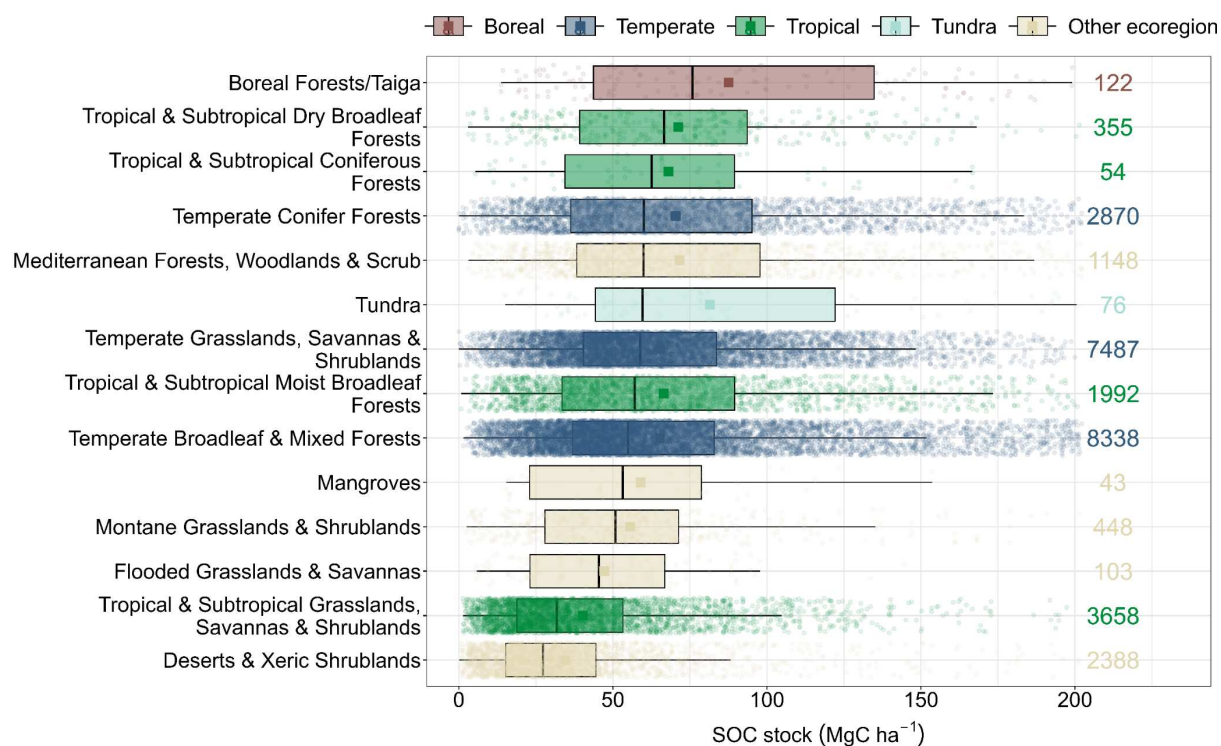
3737 3. Resultados

3738 3.1. Viés amostral e implicações para a modelagem global

3739 Na divisão por ecorregiões adotada, observa-se que a maioria das amostras de estoque
3740 de carbono das bases de dados utilizadas concentrou-se nas ecorregiões de clima temperado,
3741 em particular a Temperate Broadleaf & Mixed Forests e Temperate Grasslands, Savannas &
3742 Shublands, enquanto a menor representatividade ocorreu na ecorregião Boreal Forests/Taiga
3743 (Figura 2).

3744 Independentemente da densidade amostral, os valores de estoque de carbono orgânico
3745 do solo (SOC) se distribuíram em três grupos principais: (i) estoques elevados, entre
3746 aproximadamente 90 e 150 MgC ha⁻¹, predominam em ecorregiões de clima frio e/ou anóxico,
3747 como a Boreal Forests/Taiga e a Tundra; (ii) estoques intermediários, variando entre 50 e 70
3748 MgC ha⁻¹, são comuns em regiões de clima temperado e tropical úmido; e (iii) estoques baixos,
3749 inferiores a 20 MgC ha⁻¹, ocorrem principalmente em regiões áridas e semiáridas, incluindo
3750 zonas desérticas.

3751



3752
3753
3754
3755

Figura 2. Distribuição dos estoques de carbono orgânico do solo (SOC), com splines aplicados, por ecoregiões globais.

3756

3.2. Acurácia do modelo e limitações na predição de extremos

3757

3758

3759

3760

3761

3762

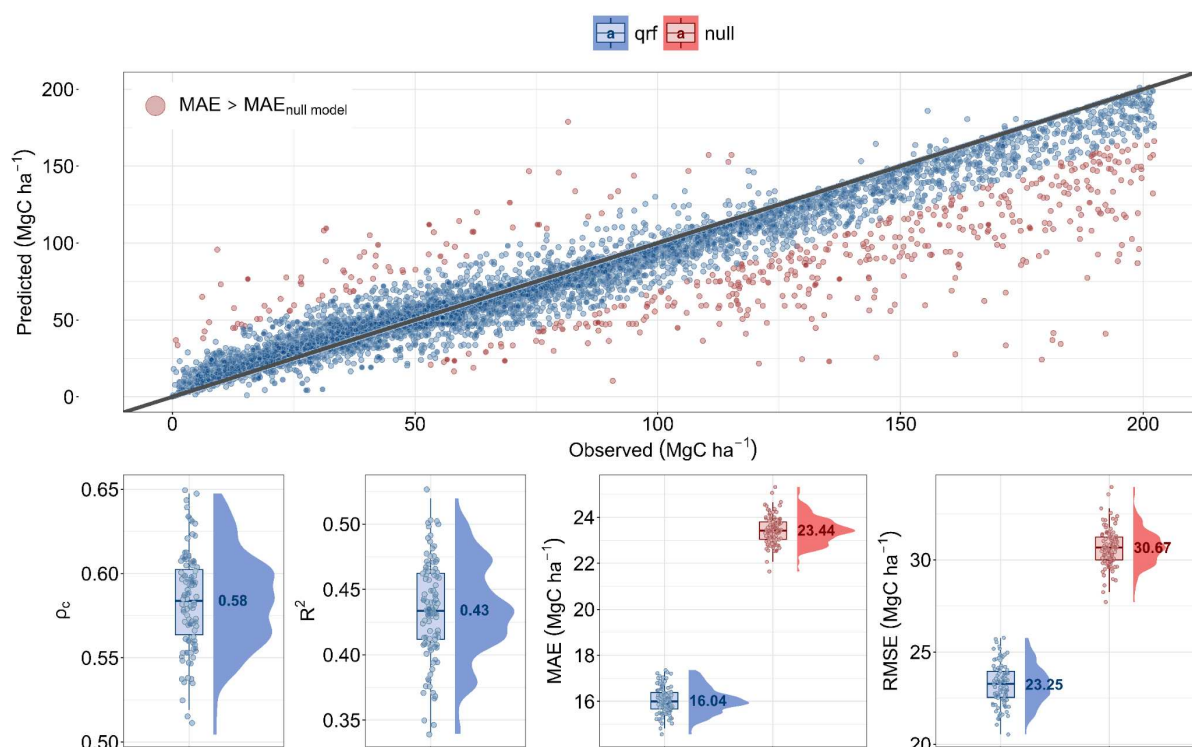
3763

3764

3765

3766

O modelo quantile random forest (qrf) apresentou desempenho satisfatório, com uma performance aproximadamente 30% superior ao modelo nulo nos indicadores de erro médio absoluto (MAE) e raiz do erro quadrático médio (RMSE) (Figura 3). Em relação ao coeficiente de concordância e ao coeficiente de determinação (R^2), os valores observados foram superiores a 0,4, o que reflete a elevada complexidade espacial dos fatores que controlam os estoques de SOC em escala global. Embora os resultados nos gráficos de valores preditos versus observados tenham evidenciado boa aderência geral do modelo, é possível notar maior dificuldade na predição de valores extremos, com desvios significativos da linha 1:1 especialmente para estoques de SOC superiores a 150 MgC ha⁻¹.



3767
3768
3769
3770

Figura 3. Dados preditos vs. observados e métricas de desempenho do modelo qrf na predição dos estoques de carbono orgânico do solo (SOC), em comparação ao modelo nulo.

3771

3.3. Padrões geográficos e fatores ambientais controladores do SOC

3772

3773

3774

3775

3776

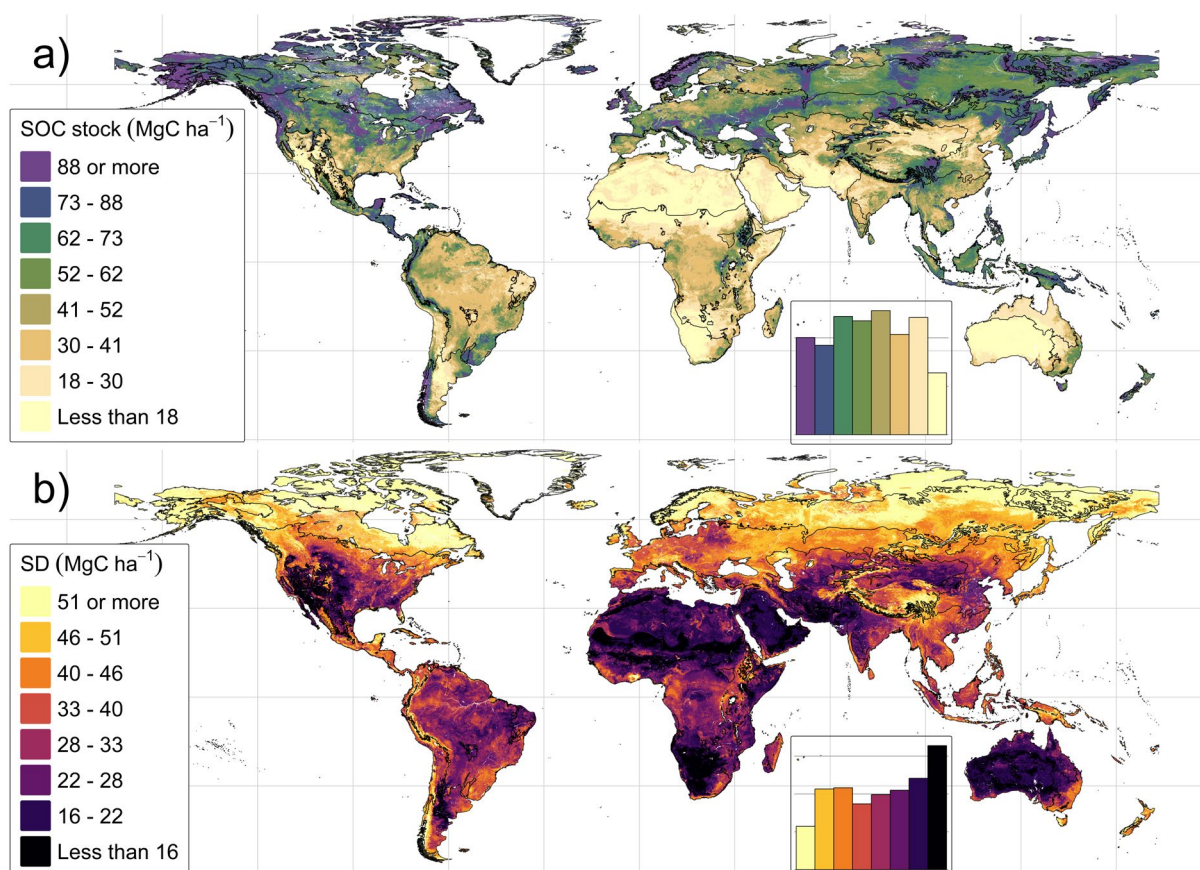
3777

3778

3779

3780

As ecorregiões agrupadas Boreais e Tundra — abrangendo o norte do Canadá, Alasca, norte da Europa e Rússia — apresentaram estoques de SOC superiores a 73 MgC ha⁻¹, chegando a ultrapassar 88 MgC ha⁻¹ nas regiões mais úmidas (Figura 4). Por sua vez, grande parte das zonas temperadas, como as porções centrais da América do Norte e a Europa Ocidental, exibiram valores entre 41 e 73 MgC ha⁻¹. Já as regiões semiáridas e desérticas — incluindo o Sahel, o norte da África, o Oriente Médio e o interior da Austrália — apresentaram os menores estoques, com valores < 30 MgC ha⁻¹, podendo ser ainda menores (abaixo de 18 MgC ha⁻¹) em áreas mais áridas.



3781
 3782 **Figura 4.** Mapas globais dos estoques de carbono orgânico do solo (SOC) e da incerteza
 3783 associada às predições do quantile random forest - qrf.
 3784

3785 Nas zonas tropicais úmidas, observou-se uma transição mais gradual nos estoques, com
 3786 florestas tropicais localizadas na Amazônia, na bacia do Congo e no sudeste asiático alcançando
 3787 valores intermediários, entre 30 e 60 MgC ha⁻¹. Quanto às incertezas associadas às estimativas,
 3788 as regiões boreais concentraram os maiores desvios, com valores superiores a 50 MgC ha⁻¹,
 3789 especialmente em latitudes elevadas do hemisfério norte.

3790 Em contraste, as zonas tropicais e temperadas, notadamente Europa, América do Norte,
 3791 América do Sul e África tropical, apresentaram incertezas relativamente baixas, entre 20 e
 3792 35 MgC ha⁻¹. Já as regiões áridas e semiáridas, como o Sahel, o Oriente Médio e o interior da
 3793 Austrália, mostraram níveis inferiores de incerteza, com desvios padrão < 20 MgC ha⁻¹.
 3794

3795 **3.4. Interações solo–biota–clima e a importância funcional dos preditores**

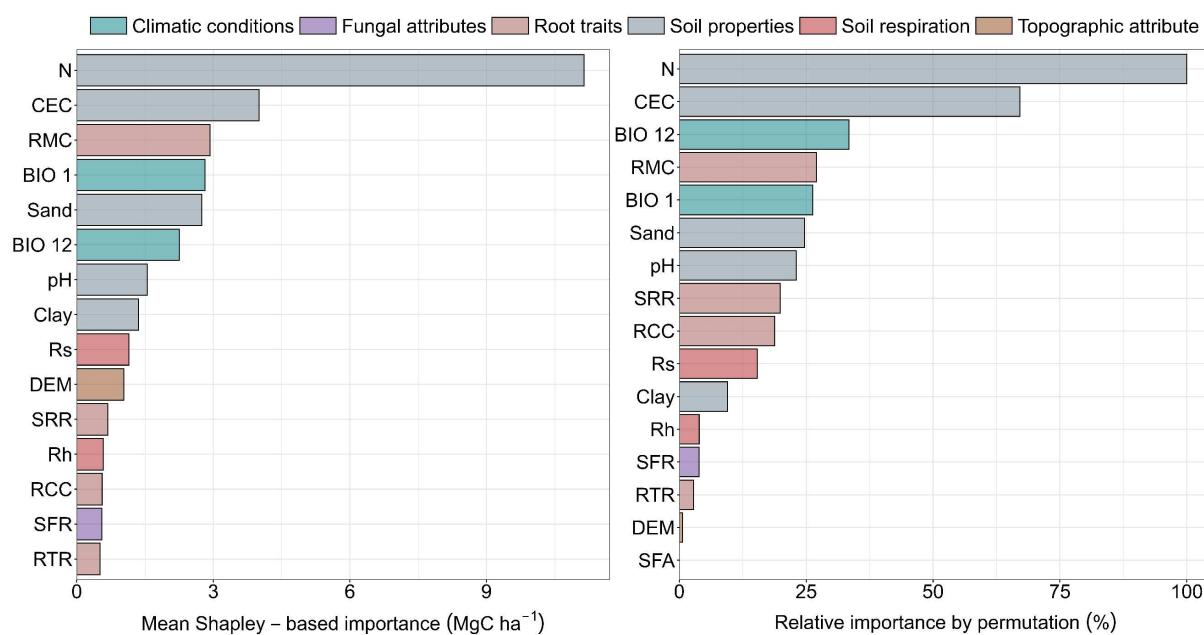
3796 **3.4.1. Importância global dos preditores ambientais**

3797 A contribuição das variáveis, conforme mensurada pelos valores de Shapley,
3798 representou o efeito marginal médio de cada preditor sobre a estimativa do estoque de SOC
3799 (Figura 5). Por sua vez, a importância por permutação quantificou a perda percentual de
3800 desempenho do modelo quando os valores de um preditor foram embaralhados, o que refletiu
3801 a sensibilidade do modelo a modificações nessa variável.

3802 O nitrogênio (N) ocupou a primeira posição em ambos os rankings, destacando-se como
3803 o principal fator de controle, associado à regulação da produtividade primária líquida. Em
3804 seguida, a capacidade de troca de cátions (CEC) apresentou o segundo maior impacto, o que
3805 reflete a relação entre cargas negativas e a matéria orgânica do solo.

3806 Na sequência, a taxa de colonização micorrízica (RMC) e a precipitação anual (BIO 12)
3807 compartilharam relevância. No ranking de Shapley, a RMC preponderou, o que evidenciou o
3808 papel das interações planta-fungo; já na permutação, BIO 12 se sobressaiu, o que indica sua
3809 influência provável sobre o aporte de resíduos e os regimes redox do solo. Ambas foram
3810 reconhecidas como variáveis-chave na determinação dos estoques de SOC.

3811



3812
3813 **Figura 5.** Importância dos preditores ambientais para os estoques de carbono orgânico do solo
3814 (SOC) com base nos valores de Shapley médio e na importância por permutação.
3815

3816 Variáveis relacionadas à respiração do solo, como a respiração total (Rs), a respiração
3817 heterotrófica (Rh), a respiração específica das raízes (SRR) e a concentração de carbono nas
3818 raízes (RCC), bem como atributos microbiológicos como a riqueza e a abundância de fungos
3819 no solo, ocuparam posições intermediárias a baixas, captando variações médias menos
3820 expressivas nos estoques de SOC.

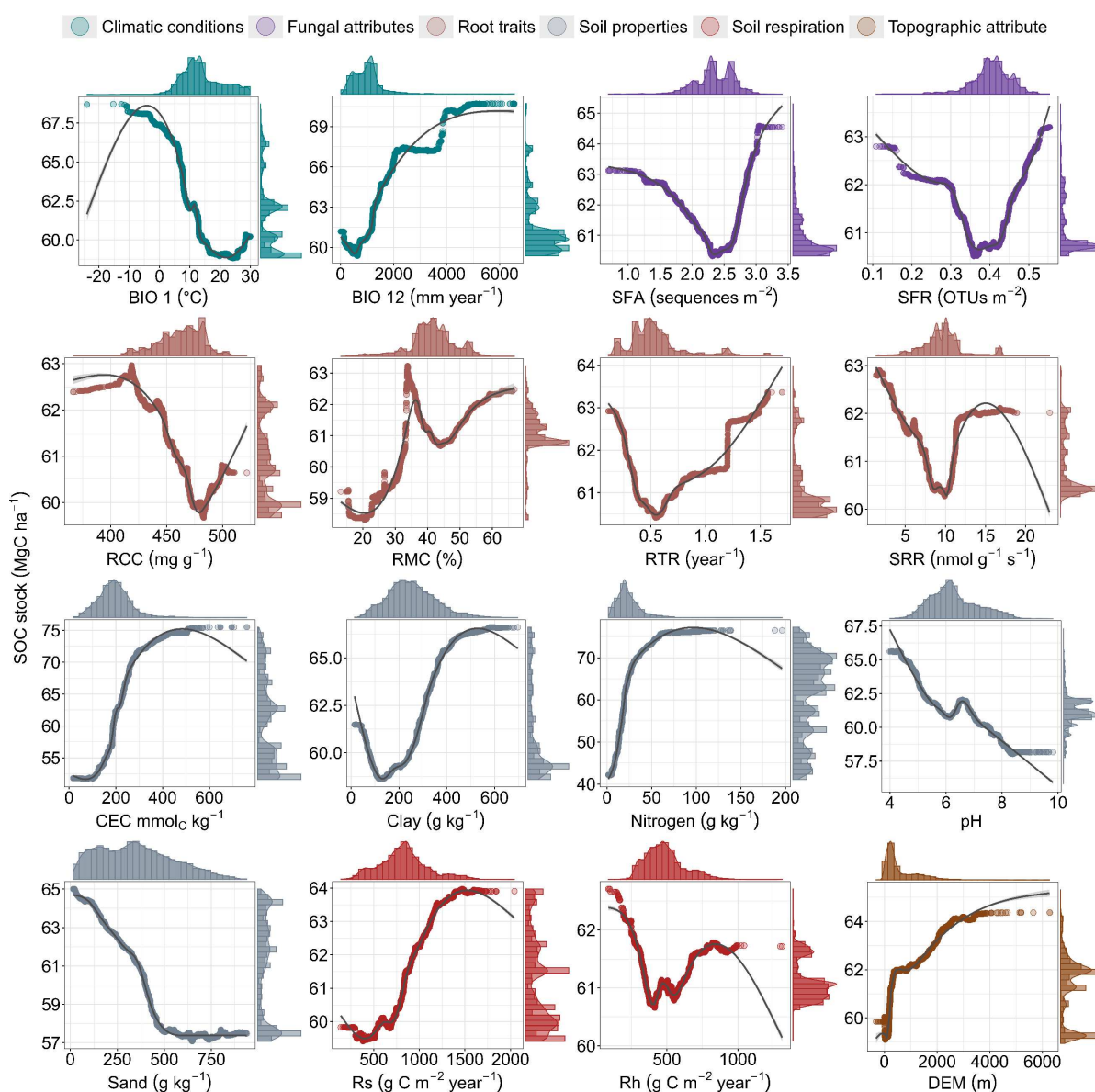
3821 O modelo digital de elevação (DEM) apareceu no ranking de Shapley com contribuição
3822 modesta ($\approx 1,5 \text{ MgC ha}^{-1}$) e impacto praticamente nulo na importância por permutação. A
3823 hierarquia de importância evidenciou que as propriedades intrínsecas do solo (como nutrientes
3824 e capacidade de troca de cátions), seguidas por fatores climáticos (precipitação e temperatura)
3825 e processos radiculares simbióticos, governaram a variação global nos estoques de SOC.

3826

3827 3.4.2. Respostas funcionais às variáveis ambientais

3828 A análise de dependência parcial revelou padrões distintos de resposta entre os
3829 diferentes grupos de variáveis (Figura 6). Dentre as variáveis climáticas, a temperatura média

3830 anual (BIO 1) apresentou uma relação predominantemente linear negativa: à medida que a
 3831 temperatura aumentou, os estoques de SOC diminuíram. Os valores máximos de SOC foram
 3832 observados entre 0 e 10 °C, com uma queda acentuada entre 10 e 15 °C.
 3833



3834
 3835 **Figura 6.** Dependência parcial dos estoques de carbono orgânico do solo (SOC) aos preditores
 3836 ambientais do modelo quantile random forest – qrf.
 3837

3838 A precipitação anual (BIO 12) apresentou uma curva com comportamento assintótico
 3839 positivo. Houve um aumento abrupto nos estoques de SOC até aproximadamente 4000 mm de
 3840 precipitação anual, seguido de uma estabilização. Regiões com menos de 1000 mm ano⁻¹

3841 apresentaram os menores estoques de SOC, enquanto valores superiores a 5000–6000 mm ano⁻¹
3842 não resultaram em ganhos adicionais, possivelmente devido à saturação hídrica ou à lixiviação
3843 excessiva.

3844 Entre as variáveis edáficas, a capacidade de troca de cátions (CEC) apresentou uma
3845 resposta linear positiva, com tendência de saturação em torno de 500 mmol_c kg⁻¹. Solos com
3846 maior capacidade de troca apresentaram maior retenção de compostos orgânicos, embora o
3847 efeito se estabilize em solos altamente reativos.

3848 A fração argila (Clay) exibiu uma relação não linear: os estoques de SOC tenderam a
3849 diminuir até cerca de 150 g kg⁻¹, cresceram entre 150 e 450 g kg⁻¹, e estabilizaram-se em teores
3850 mais elevados. A presença de argila, em geral, favorece a proteção do SOC por mecanismos de
3851 complexação e agregação. No entanto, teores muito elevados podem levar à saturação da
3852 capacidade de retenção mineral e à limitação da aeração do solo.

3853 A fração areia (Sand) apresentou um padrão de declínio quase logarítmico nos estoques
3854 de SOC até cerca de 500 g kg⁻¹, com estabilização acima desse valor. Solos mais arenosos
3855 tendem a perder a capacidade de reter e proteger o carbono, em razão da menor área superficial
3856 e maior capacidade de drenagem.

3857 Por fim, o pH do solo demonstrou um comportamento quase linear negativa até
3858 aproximadamente pH 8, com estabilização a partir desse ponto. Solos ácidos (pH < 5) podem
3859 inibir a atividade microbiana, o que, paradoxalmente, pode favorecer o acúmulo de SOC ao
3860 reduzir a taxa de decomposição da matéria orgânica.

3861 No que se refere às variáveis relacionadas às raízes, a taxa de colonização micorrízica
3862 (RMC) apresentou uma relação não linear positiva, com dois picos principais observados em
3863 torno de 35% e 60%. Esses resultados indicam que altos níveis de colonização micorrízica
3864 favorecem a formação de agregados estáveis e a incorporação de carbono no solo, ressaltando
3865 o papel das interações simbióticas na dinâmica do SOC.

3866 A concentração de carbono nas raízes (RCC) apresentou um padrão mais complexo: os
3867 estoques de SOC mantiveram-se relativamente estáveis até aproximadamente 430 mg g^{-1} ,
3868 seguidos por uma queda acentuada entre 430 e 475 mg g^{-1} . Após esse intervalo, observou-se um
3869 novo aumento. Esse comportamento sugere a existência de faixas de saturação e possíveis
3870 efeitos competitivos ou de turnover da biomassa radicular, que podem temporariamente limitar
3871 o acúmulo de SOC.

3872 A taxa de renovação radicular (RTR) também demonstrou uma relação não linear com
3873 o SOC. Valores de RTR inferiores a $0,5 \text{ ano}^{-1}$ e superiores a $1,4\text{--}1,5 \text{ ano}^{-1}$ estiveram associados
3874 a maiores estoques de SOC. Por outro lado, taxas intermediárias, próximas a uma renovação
3875 anual, corresponderam aos menores estoques. Isso indica que tanto baixas quanto altas taxas de
3876 renovação radicular podem favorecer a entrada de carbono no solo, enquanto taxas
3877 intermediárias parecem ser menos eficazes nesse processo, possivelmente por não promoverem
3878 nem o acúmulo expressivo nem a rápida mineralização da biomassa radicular.

3879 A respiração específica das raízes (SRR) apresentou um padrão de decréscimo até cerca
3880 de $10 \text{ nmol g}^{-1} \text{ s}^{-1}$, seguido por um aumento abrupto até aproximadamente $12 \text{ nmol g}^{-1} \text{ s}^{-1}$, com
3881 posterior estabilização. Os maiores estoques de SOC foram observados nos extremos da
3882 distribuição (valores < 6 ou > 12), sugerindo que raízes altamente ativas ou, inversamente,
3883 muito inativas estão associadas a estratégias contrastantes que promovem maior acúmulo de
3884 SOC.

3885 No grupo das variáveis associadas à respiração do solo, a respiração total (R_s)
3886 apresentou uma relação predominantemente linear positiva com os estoques de SOC, com
3887 crescimento até cerca de $1400 \text{ g C m}^{-2} \text{ ano}^{-1}$, seguido de uma tendência à estabilização. Esse
3888 padrão sugere que maiores fluxos respiratórios totais refletem maior entrada de SOC, seja via
3889 raízes ativas ou por decomposição de matéria orgânica recente.

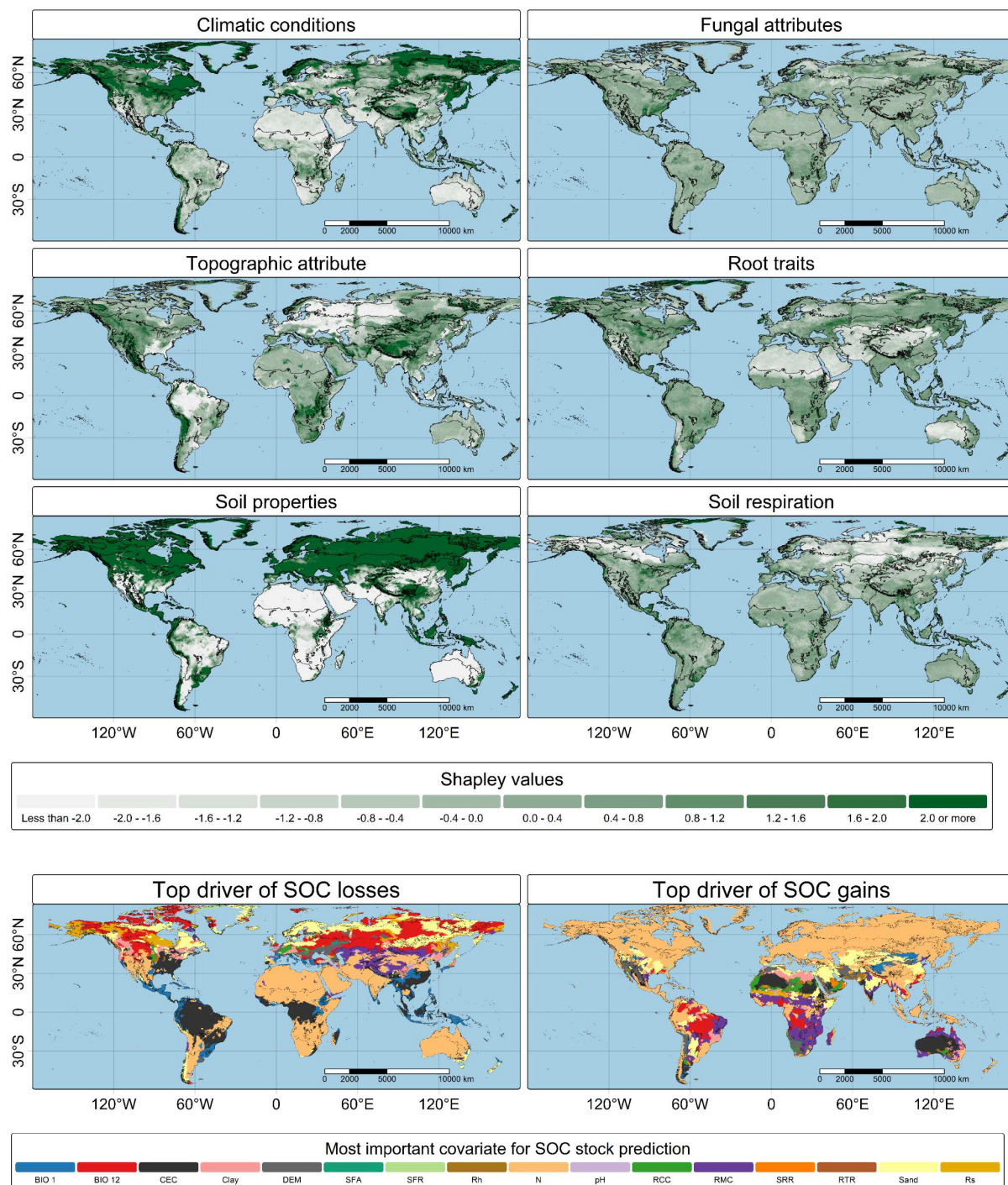
3890 Por outro lado, a respiração heterotrófica (Rh) demonstrou uma relação em forma de
3891 “U”: houve redução abrupta nos estoques de SOC até aproximadamente $500 \text{ g C m}^{-2} \text{ ano}^{-1}$,
3892 seguida por um aumento até cerca de $700 \text{ g C m}^{-2} \text{ ano}^{-1}$ e posterior estabilização. Assim como
3893 em outras variáveis radiculares, os extremos de Rh estiveram associados a maiores estoques de
3894 SOC. Já os valores intermediários indicaram condições de mineralização líquida mais intensa,
3895 que resultaram na redução do carbono armazenado no solo.

3896 Em relação à topografia, o modelo digital de elevação (DEM) apresentou um padrão de
3897 crescimento acentuado nos estoques de SOC até aproximadamente 3000 m de altitude, seguido
3898 por uma estabilização em forma de platô. Esse comportamento sugere que o aumento da
3899 altitude, ao reduzir as temperaturas médias, contribuiu para o acúmulo e a maior permanência
3900 do SOC. Em áreas montanhosas mais frias, essas condições pode promover menor taxa de
3901 decomposição e maior proteção física da matéria orgânica, favorecendo a formação e
3902 manutenção de estoques mais elevados de SOC.

3903

3904 **3.4.3. Padrões espaciais de dominância de importância com base nos valores de** 3905 **Shapley**

3906 Diferentemente dos valores médios de importância e dependência parcial, os mapas
3907 baseados nos valores de Shapley revelaram, de forma espacializada, a atuação de grupos e a
3908 dominância local da variável mais influente sobre os estoques de SOC (Figura 7). É importante
3909 ressaltar que as relações estabelecidas pelo modelo com base nos valores de Shapley indicaram
3910 como cada variável contribuiu para a predição do estoque de SOC em um determinado *pixel*,
3911 considerando o contexto local. Todas as variáveis influenciaram, em algum grau, as predições
3912 locais, mas com diferentes níveis de relevância. Neste ponto da análise, o foco está na
3913 dominância relativa de cada variável em relação às demais, ou seja, qual variável teve maior
3914 influência para explicar a predição em determinada localidade.



3915
 3916 **Figura 7.** Padrões espaciais, baseados nos valores de Shapley, da importância funcional e
 3917 direcional dos preditores ambientais sobre os estoques de carbono orgânico do solo (SOC).
 3918

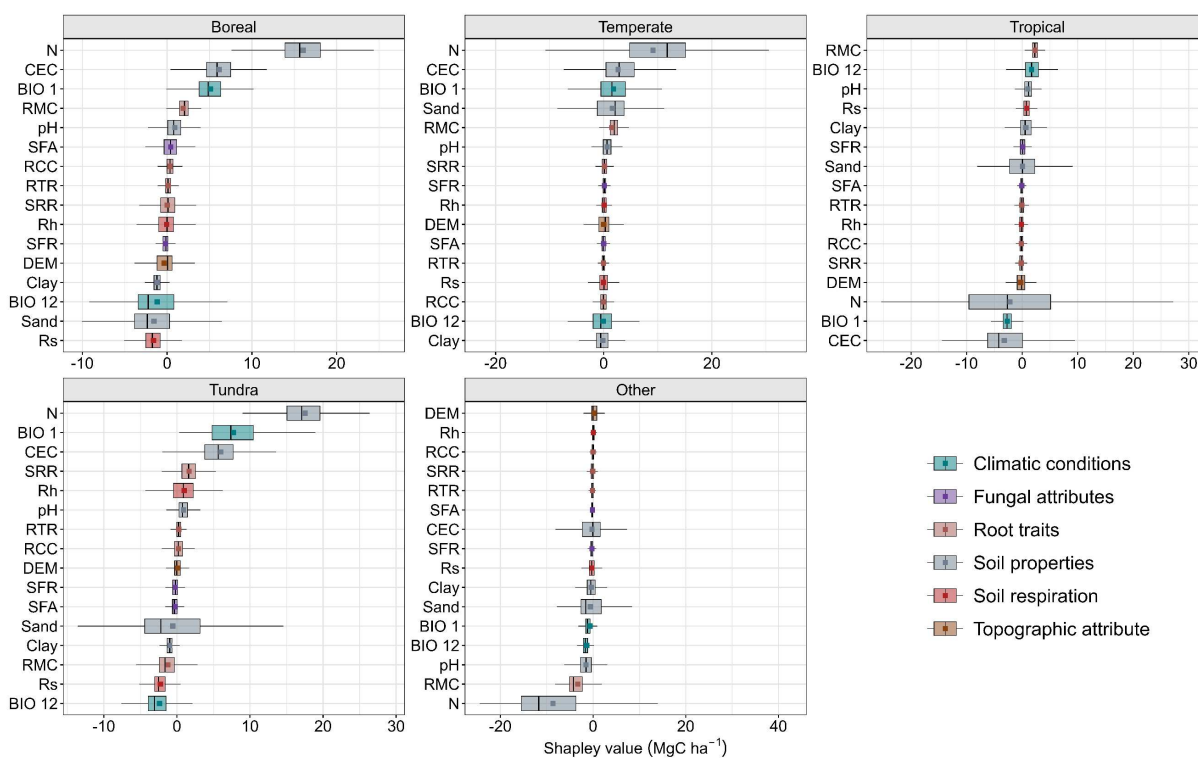
3919 Cabe destacar que essa análise não implica relação causal, tampouco permite inferir
 3920 direcionalidade do efeito. Por exemplo, observar que o nitrogênio (N) apresenta um valor de
 3921 Shapley negativo em uma região não significa, necessariamente, que maiores teores de N

3922 reduzem os estoques de SOC, nem que a adição de N leva à perda de carbono. O valor de
3923 Shapley apenas quantifica o quanto aquela variável contribuiu, positiva ou negativamente, na
3924 predição final do modelo em relação à média global, considerando o conjunto específico de
3925 interações naquele ponto.

3926 Assim, valores como +2 ou -2 Mg C ha⁻¹ devem ser interpretados como contribuições
3927 relativas ao *baseline* para o desvio local em relação à predição média, e não como indícios de
3928 efeito direto ou monotônico. Trata-se, portanto, de uma contribuição aditiva marginal dentro da
3929 estrutura explicativa do modelo

3930 As ecorregiões Boreal, Temperate e de Tundra apresentaram forte influência das
3931 variáveis climáticas sobre os estoques de SOC, especialmente da temperatura média anual
3932 (BIO 1), conforme evidenciado pelos padrões espaciais (Figura 7) e pelos valores de Shapley
3933 nas ecorregiões (Figura 8). A contribuição positiva da BIO 1 nessas regiões indicou que
3934 temperaturas mais baixas favoreceram o acúmulo de SOC.

3935



3936
3937
3938
3939

Figura 8. Importância dos preditores ambientais para os estoques de carbono orgânico do solo (SOC) por grupo de ecorregiões, com base nos valores de Shapley.

3940

O nitrogênio (N) foi identificado como um dos preditores mais influentes, associado ora

3941

ao aumento, ora à redução dos estoques, dependendo do contexto edafoclimático (Figura 7).

3942

Essa ambivalência também se refletiu nas análises por ecorregião (Figura 8), nas quais o N se

3943

destacou tanto em ambientes frios quanto em regiões tropicais secas.

3944

A capacidade de troca de cátions (CEC) apresentou efeitos contrastantes: contribuiu

3945

positivamente em solos frios, onde refletiu a presença de matéria orgânica mais perene, e

3946

negativamente em solos tropicais, possivelmente associados a alta intensidade de intemperismo

3947

e menores estoques de SOC. A fração arenosa (Sand) foi consistentemente associada a estoques

3948

reduzidos de SOC, com ampla distribuição espacial e destaque nas ecorregiões frias, onde a

3949

baixa capacidade de retenção de matéria orgânica em solos arenosos limitou o acúmulo de

3950

carbono.

3951

Quanto aos atributos fúngicos, a taxa de colonização micorrízica (RMC) apresentou

3952

contribuição positiva expressiva nos trópicos, conforme indicado pelos padrões espaciais e pela

3953 análise por ecorregiões. Os efeitos foram mais intensos em florestas tropicais úmidas, como a
3954 Amazônia, a bacia do Congo e o Sudeste Asiático, sugerindo que a elevada atividade fúngica
3955 em solos de climas quentes e úmidos podem favorecer o acúmulo de carbono. Em contrapartida,
3956 em ambientes extremos, como desertos quentes e regiões de tundra, a variável relacionada aos
3957 fungos limitou a formação de estoques de SOC.

3958 A precipitação anual (BIO 12) e o pH do solo também se destacaram como fatores
3959 positivos nos trópicos. A BIO 12 foi dominante no Cerrado brasileiro, no norte da Amazônia e
3960 no centro-sul da África, atuando como principal fator climático associado ao aumento dos
3961 estoques. O pH contribuiu positivamente ao modular a acidez em solos tropicais, o que
3962 favoreceu a atividade microbiana.

3963 A topografia também exerceu influência relevante. Em regiões montanhosas, como os
3964 Andes, Himalaia, Montanhas Rochosas, Rift africano e Nova Guiné, o modelo indicou
3965 contribuição positiva para o SOC. Em contrapartida, nas grandes planícies sedimentares, a
3966 influência topográfica foi predominantemente negativa. O modelo digital de elevação (DEM),
3967 em particular, apresentou contribuição positiva em regiões montanhosas, ilhas e regiões
3968 isoladas, refletindo sua associação com gradientes climáticos de temperatura e precipitação.

3969 Os atributos radiculares apresentaram efeitos positivos mais uniformes globalmente,
3970 com algum destaque para regiões tropicais e temperadas. Já os efeitos negativos predominaram
3971 em regiões desérticas. As variáveis associadas à respiração do solo, de modo geral,
3972 apresentaram contribuição positiva relativamente homogênea para os estoques de SOC, com
3973 exceção das latitudes elevadas, onde atividade respiratória se associou a efeitos negativos.

3974 Em síntese, o nitrogênio (N) destacou-se como a variável dominante em ambientes frios,
3975 Boreal, Temperate e de Tundra; a RMC, nos trópicos; e o DEM, em regiões montanhosas e
3976 isoladas. As variáveis mais relevantes para o aumento e a redução dos estoques de SOC foram,

3977 em sua maioria, as mesmas, como N, CEC e Sand, porém seus efeitos variaram espacialmente
3978 de acordo com as condições edafoclimáticas predominantes.

3979

3980 **4. Discussão**

3981 **4.1. Viés amostral e implicações para a modelagem global**

3982 Os dados utilizados são predominantemente provenientes de regiões temperadas,
3983 seguidos pelas regiões tropicais, sendo escassos nas florestas de taiga e em zonas áridas e
3984 semiáridas. Essa lacuna é preocupante, uma vez que as florestas de taiga e a ecorregião boreal
3985 concentram alguns dos maiores estoques globais de carbono orgânico do solo (SOC) (DeLuca
3986 & Boisvenue, 2012).

3987 A sub-representatividade dessas áreas pode gerar implicações negativas e desequilíbrios
3988 nos modelos preditivos, resultando em viés com potencial impacto sobre inventários de carbono
3989 e estratégias de mitigação climática (Barkov et al., 2024; Heuvelink, 2018). Os dados
3990 observados confirmam a ocorrência dos maiores estoques de SOC nessas regiões,
3991 frequentemente superiores a 90 MgC ha⁻¹.

3992 A predominância das amostras em regiões temperadas e tropicais úmidas está
3993 relacionada, em grande parte, à maior acessibilidade logística, infraestrutura de coleta de dados
3994 e à disponibilidade de repositórios públicos de solos, e não necessariamente à relevância
3995 climática ou ecológica dessas áreas em escala global (Calisto et al., 2023; S. Chen et al., 2024).

3996 Além disso, regiões áridas e semiáridas, que concentram os menores estoques de SOC,
3997 também tenderam a ser subamostradas. Essa sub-representação tanto em zonas de estoques
3998 muito elevados quanto muito baixos compromete a capacidade de compreender os limites
3999 superiores e inferiores dos estoques de SOC sob condições edafoclimáticas extremas.

4000 Dessa forma, a atual distribuição espacial das amostras revela um viés geográfico
4001 expressivo, no qual regiões de interesse climático e ecológico prioritário, como os ecossistemas

4002 boreais, permanecem subrepresentadas. Essa limitação pode comprometer a precisão das
4003 estimativas globais de SOC e impactar negativamente políticas públicas baseadas em dados
4004 ambientais.

4005 Tais evidências reforçam a necessidade de estratégias de amostragem mais equilibradas
4006 e direcionadas, com ênfase na expansão da cobertura em áreas de alta sensibilidade climática e
4007 baixa densidade, como a ecorregião boreal. Nesses ambientes, o degelo do *permafrost*
4008 representa um risco crítico (Kim et al., 2025; Schütte et al., 2019), pois pode desencadear
4009 emissões em larga escala de carbono anteriormente protegido no solo (Harris et al., 2023;
4010 Helbig et al., 2024; Jones et al., 2017).

4011

4012 **4.2. Acurácia do modelo e limitações na predição de extremos**

4013 Dada a complexidade e dinamicidade dos estoques de SOC em escala global, o modelo
4014 apresentou métricas de acurácia consideradas adequadas, especialmente em comparação com
4015 estudos prévios da literatura voltados à modelagem de atributos do solo em diferentes escalas
4016 (Gomes et al., 2019; Moquedace et al., 2024; Poggio et al., 2021; Siqueira et al., 2023, 2024).
4017 Uma forma mais objetiva de avaliar esse desempenho foi por meio da comparação com um
4018 modelo nulo, cuja performance foi expressivamente inferior, apresentando erros cerca de um
4019 terço maiores do que os do modelo qrf.

4020 É importante destacar que as métricas globais integradas, coeficiente de correlação de
4021 concordância (ρ_c), coeficiente de determinação (R^2), erro médio absoluto (MAE) e raiz do erro
4022 quadrático médio (RMSE), ofereceram uma visão geral útil do desempenho do modelo. No
4023 entanto, ao se observar a distribuição dos estoques de SOC por faixas de valor, nota-se que os
4024 maiores desvios ocorrem nos extremos da distribuição, tanto para estoques muito baixos quanto
4025 para estoques elevados. Nesses intervalos, os erros do modelo qrf superaram os do modelo nulo.

4026 Esse padrão foi particularmente evidente para valores elevados de SOC, acima de
4027 120 MgC ha⁻¹ nos quais a dispersão dos resíduos aumentou substancialmente. Essa observação
4028 é estatisticamente relevante por dois motivos principais: (i) indica que métricas globais de
4029 desempenho, embora informativas, podem mascarar erros localizados, sendo necessária uma
4030 análise estratificada por faixa de valor ou região; e (ii) evidencia uma limitação comum de
4031 modelos baseados em árvores de decisão, como o Random Forest, que tendem a apresentar
4032 previsões centradas na faixa média dos dados observados, reduzindo a acurácia na predição de
4033 valores extremos, mesmo quando esses representam variações naturais do sistema (Siqueira et
4034 al., 2024).

4035

4036 **4.3. Padrões geográficos e fatores ambientais que controlam o SOC**

4037 A elevada concentração de SOC observada nas altas latitudes era esperada, uma vez que
4038 o clima frio, a presença de *permafrost* e a ocorrência de solos turfosos favoreceram a retenção
4039 de grandes quantidades de matéria orgânica (Harden et al., 1992; Hobbie et al., 2000; Hugelius
4040 et al., 2014) ao longo de milhares a dezenas de milhares de anos durante o Quaternário Tardio
4041 (Gorham, 1991; Schirrmeister et al., 2011; Smith et al., 2004).

4042 Os estoques intermediários encontrados na faixa latitudinal de clima temperado refletem
4043 a interação entre clima ameno e forte sazonalidade, associada a intensos usos antrópicos do
4044 solo. A intensificação da prática agrícola e do manejo em pastagens nessas regiões reduzem
4045 significativamente os estoques de SOC, comumente resultando em menores teores de carbono
4046 nos solos cultivados (Poeplau et al., 2011; Singh et al., 2011; You et al., 2023).

4047 Nas regiões áridas e semiáridas, dois fatores principais explicam os baixos estoques de
4048 SOC: a limitação hídrica, que restringe a produtividade primária e, conseqüentemente, o aporte
4049 de matéria orgânica ao solo (Naorem et al., 2023); e as altas temperaturas, que intensificam os
4050 processos de mineralização e promovem a rápida decomposição do pouco carbono que é

4051 formado na superfície (Nicolás et al., 2012). Soma-se a isso a predominância de solos arenosos,
4052 nos quais a baixa quantidade de argila reduz a formação de agregados e a proteção físico-
4053 química da matéria orgânica, o que dificulta a sua permanência via interações organominerais
4054 (Van-Camp et al., 2004).

4055 Na faixa tropical, os estoques de SOC intermediários, são caracterizados por transições
4056 graduais ao longo da paisagem (Reichenbach et al., 2025; Richards et al., 2007). Apesar das
4057 elevadas taxas de decomposição, impulsionadas por alta disponibilidade hídrica e temperaturas
4058 elevadas, a alta produtividade primária garante um aporte contínuo de resíduos orgânicos,
4059 especialmente em ecossistemas nativos, o que contribui para compensar a mineralização
4060 acelerada (Hein et al., 2020). Além disso, os solos tropicais tendem a ser profundos, altamente
4061 intemperizados e ricos em argila, o que favorece a formação de agregados estáveis que
4062 protegem a matéria orgânica da decomposição, além de estabelecerem conexões
4063 organominerais que contribuem para a permanência de longo prazo do SOC (Elrys et al., 2025).

4064 As incertezas estiveram associadas, principalmente, às regiões com altas concentrações
4065 de estoque de SOC, notadamente na faixa andina e nas altas latitudes. Esses padrões se
4066 conectaram diretamente com os dados de valores preditos versus observados, evidenciando a
4067 dificuldade do modelo qrf em lidar com valores extremos (Moquedace et al., 2024). Nessas
4068 regiões, os altos estoques de SOC resultaram em maior dispersão dos resíduos e aumento da
4069 incerteza preditiva, um comportamento esperado em modelos baseados em árvores de decisão,
4070 cuja tendência é aumentar os erros em dados distantes da média.

4071

4072 **4.4. Interações solo–biota–clima e a importância funcional dos preditores**

4073 **4.4.1. Importância global dos preditores ambientais**

4074 Os resultados de importância dos preditores foram complementares ao revelar
4075 hierarquias funcionais entre os fatores que controlam os estoques de SOC em escala global. A

4076 dominância do nitrogênio (N) em ambos os métodos de interpretação — importância por
4077 permutação e valores médios de Shapley, reafirma seu papel central como regulador da
4078 produtividade primária líquida (O'Rourke et al., 2015), sendo um elemento-chave para a
4079 atividade biológica do solo e, conseqüentemente, para o acúmulo de SOC (Fornara & Tilman,
4080 2008; Pellegrini et al., 2018; Yildiz et al., 2020). Vale destacar, contudo, que grande parte do N
4081 comumente medido em estudos de solo (>90%) é derivado da própria matéria orgânica do solo
4082 (MOS), o que pode inflacionar a correlação entre N e SOC e dificultar a distinção entre efeitos
4083 causais e circulares, especialmente quando não se separam as frações mineral, livre e orgânica
4084 do N.

4085 A proeminência da capacidade de troca de cátions (CEC) ressalta a relevância da fração
4086 argilomineral na permanência da matéria orgânica, por meio da formação de complexos
4087 organominerais que protegem fisicamente e quimicamente o carbono no solo (Elrys et al.,
4088 2025). A emergência da taxa de colonização micorrízica (RMC) como uma variável de
4089 importância média a alta, frequentemente negligenciada em modelos globais, juntamente com
4090 a precipitação anual (BIO 12), evidencia a interação sinérgica entre processos biológicos e
4091 climáticos.

4092 A simbiose planta-fungo impacta diretamente o ciclo do carbono por meio do sistema
4093 radicular, sendo os solos rizosféricos reconhecidamente mais ativos biologicamente (Cui et al.,
4094 2024; Jiang et al., 2020; Tang et al., 2023; Yang et al., 2023). Por sua vez, o regime hídrico atua
4095 como regulador da entrada de resíduos orgânicos e dos processos redox, que modulam a
4096 decomposição e a mineralização da matéria orgânica.

4097 Em contrapartida, variáveis associadas à respiração do solo e à atividade
4098 microbiológica, embora reconhecidas por sua importância na multifuncionalidade ecológica
4099 dos solos, apresentaram influência menos expressiva na modelagem global. Esse resultado
4100 sugere que sua contribuição para o balanço de SOC pode ser mais localizada ou condicionada

4101 a contextos ambientais específicos, como tipos de solo, clima ou uso da terra. Além disso, a
4102 supressão da importância relativa desses fatores pode refletir um efeito de substituição de
4103 informação, no qual outras variáveis capturam indiretamente seus efeitos, diluindo seu peso
4104 específico nos modelos.

4105 Por fim, a baixa contribuição do relevo (DEM) reforça a hipótese de que, em escala
4106 global, os fatores edafoclimáticos e biológicos superam os efeitos topográficos diretos na
4107 predição dos estoques de SOC (Nyaupane et al., 2023; Peng et al., 2020; Xiang et al., 2023).
4108 Esses achados sustentam a priorização de atributos pedológicos e climáticos em modelos
4109 globais de SOC e indicam que interações simbióticas, como as micorrízicas, merecem maior
4110 atenção em estratégias baseadas na natureza para a fixação de carbono.

4111 Embora o aprendizado de máquina seja eficiente em detectar padrões numéricos
4112 complexos, sua aplicação em ciência do solo exige que esses padrões sejam interpretados à luz
4113 de processos pedológicos reais, com coerência biológica e aderência à realidade dos sistemas
4114 edáficos. Caso contrário, pode-se incorrer em produzir modelos estatisticamente robustos,
4115 porém cientificamente frágeis.

4116

4117 **4.4.2. Respostas funcionais às variáveis ambientais**

4118 Os padrões revelados pela análise de dependência parcial reforçaram a centralidade dos
4119 fatores climáticos e edáficos no controle do SOC, mas com nuances que ampliam a
4120 compreensão funcional desses preditores. A relação negativa da temperatura média anual
4121 (BIO 1) confirma o papel limitante do calor excessivo no acúmulo de SOC. A resposta
4122 assintótica positiva da precipitação anual (BIO 12) aponta que há um teto funcional na
4123 capacidade do sistema em converter água adicional em acúmulo de SOC, sugerindo que o
4124 excesso hídrico pode, inclusive, comprometer esse processo (Z. Chen et al., 2023; Deng et al.,
4125 2021; Xu et al., 2024; Zhou et al., 2016).

4126 As variáveis edáficas, como CEC e Clay, mantêm o padrão já observado: seus efeitos
4127 positivos reforçam a importância da fração mineral na permanência da matéria orgânica,
4128 embora apresentem limiares de saturação que devem ser considerados. Solos mais ácidos (pH
4129 < 5), a depender do local, podem inibir a decomposição, também limitar a atividade biológica,
4130 e solos com pH neutro a levemente alcalino surgiram como ideal para promover acúmulo de
4131 SOC, alinhando-se provavelmente com a atuação funcional dos microrganismos.

4132 Entre os fatores biológicos, destacaram-se os padrões não lineares de resposta das
4133 variáveis radiculares e respiratórias, sugerindo que o controle do SOC por esses fatores não se
4134 dá em gradientes simples, mas sim por meio de faixas funcionais. A taxa de colonização
4135 micorrízica (RMC) seguiu como variável-chave, com dois picos de acúmulo de SOC que
4136 refletiram diferentes estratégias ecológicas associadas à simbiose planta-fungo. Já os resultados
4137 para RTR, RCC e SRR mostraram que a dinâmica do carbono radicular é altamente sensível à
4138 intensidade e ao turnover da biomassa, com maiores estoques nos extremos funcionais
4139 (atividade muito baixa ou muito alta).

4140 Esse padrão também se aplica à respiração do solo: enquanto a respiração total (R_s) se
4141 relacionou positivamente ao SOC, a respiração heterotrófica (R_h) exibiu uma relação em “U”,
4142 indicando que fluxos respiratórios intensos nem sempre significam perda de carbono — tudo
4143 depende da origem e destino desse fluxo. Por fim, a contribuição positiva do DEM, com
4144 altitudes até 3000 m reforça o argumento anterior de que, embora a topografia tenha baixa
4145 importância média global, sua ação sobre a temperatura, deposição de sedimentos e processos
4146 hidrológicos locais o torna um fator relevante em regiões de maior altitude.

4147

4148 **4.4.3. Padrões espaciais de dominância de importância com base nos valores de**
4149 **Shapley**

4150 Para fins interpretativos, valores de Shapley positivos ou negativos são aqui utilizados
4151 como indicativos de contribuições marginais que elevaram ou reduziram, respectivamente, a
4152 predição local do modelo em relação à *baseline*, sem implicar relação causal ou monotonicidade
4153 direta com os estoques de SOC. As análises espaciais dos valores de Shapley revelaram padrões
4154 de dominância funcional que complementam as métricas globais de importância e dependência
4155 parcial, indicando que as propriedades do solo exerceram a maior influência relativa sobre os
4156 estoques de carbono orgânico do solo (SOC), seguidas pelas variáveis climáticas.

4157 Em regiões frias do Hemisfério Norte, observou-se forte associação positiva com o SOC
4158 e a presença de temperaturas médias anuais reduzidas (BIO 1), o que está de acordo com a
4159 literatura clássica que associa climas frios à preservação da matéria orgânica (Gorham, 1991;
4160 Hobbie et al., 2000; Singh et al., 2011; You et al., 2023). A baixa atividade microbiana em
4161 ambientes gelados e frequentemente saturados por água desacelera os processos de
4162 decomposição, favorecendo a retenção de carbono ao longo do tempo.

4163 Padrões semelhantes foram registrados em áreas de alta montanha, como os Andes e o
4164 Himalaia, nas quais os gradientes altitudinais também modulam regimes térmicos e hídricos
4165 propícios ao acúmulo de SOC (L. Chen et al., 2017; Francis Justine et al., 2024; Xue et al.,
4166 2023). Por outro lado, em ambientes áridos e semiáridos, do cinturão do Saara à Austrália
4167 Central, as contribuições climáticas foram majoritariamente negativas. Nessas regiões, a
4168 combinação entre alta temperatura e déficit hídrico favorece a rápida mineralização da matéria
4169 orgânica e limita a produtividade primária, resultando em baixos estoques de SOC (Naorem et
4170 al., 2023; Nicolás et al., 2012).

4171 O nitrogênio (N) destacou-se como uma das variáveis mais influentes, exibindo
4172 comportamento ambivalente conforme o contexto ambiental. Em regiões áridas e semiáridas,

4173 marcadas por elevada vulnerabilidade à degradação, evapotranspiração intensa e baixa
4174 biomassa vegetal, os valores médios de Shapley foram predominantemente negativos,
4175 sugerindo que o N funcionou como um marcador de limitação funcional do solo, e não como
4176 causa direta da perda de carbono (Naorem et al., 2023; Nicolás et al., 2012). Já em ambientes
4177 frios e úmidos, o N associou-se positivamente aos estoques de SOC, refletindo a correlação
4178 espacial entre teores elevados de nitrogênio e a presença de matéria orgânica mais persistente
4179 (Batjes, 1996). Nesses contextos, o N atuou como um indicador químico edáfico típico de áreas
4180 com alta acumulação de carbono, em parte devido à lentidão dos processos de decomposição.

4181 É importante considerar, no entanto, que grande parte do N medido em solos é composta
4182 pelo próprio N orgânico contido na MOS, o que reforça sua co-ocorrência com altos teores de
4183 SOC e limita a interpretação de sua importância como evidência de causalidade direta
4184 (Cleveland & Liptzin, 2007; Van Groenigen et al., 2015). Assim, a relevância atribuída ao N
4185 pelo modelo pode refletir antes sua função como marcador funcional da MOS do que como
4186 preditor independente, exigindo cautela na inferência pedoecológica e espacial de sua
4187 influência.

4188 A variável BIO 1 apresentou comportamento não linear: enquanto um leve aquecimento
4189 em zonas frias pode estimular a produtividade vegetal sem promover perdas imediatas de SOC
4190 (Mello et al., 2025), em ambientes tropicais o mesmo aumento térmico tende a acelerar a
4191 decomposição e a reduzir o carbono armazenado (Piao et al., 2009). Alguns estudos apontam
4192 que, mesmo nos trópicos, o aquecimento pode elevar a produtividade primária líquida a ponto
4193 de compensar parcialmente as perdas por mineralização, dependendo do regime hídrico e do
4194 estágio sucessional do ecossistema (Gottschalk et al., 2012). Essas respostas não lineares
4195 reforçam a importância de considerar limiares ecológicos no entendimento do ciclo do carbono
4196 em escala global.

4197 Nos trópicos, a precipitação anual (BIO 12) e a taxa de colonização micorrízica (RMC)
4198 destacaram-se como variáveis-chave. A BIO 12 atuou como marcador climático de *hotspots* de
4199 carbono, especialmente em regiões como o Cerrado brasileiro, o norte da Amazônia e zonas de
4200 transição florestal, onde reforçou o sinal de sistemas biologicamente ativos com elevada
4201 produtividade subterrânea.

4202 Em áreas com alta RMC, a precipitação acentuou os efeitos biológicos, moldando
4203 ambientes favoráveis ao acúmulo de SOC por meio da interação entre água e atividade
4204 simbiótica. No entanto, a BIO 12 não deve ser interpretada como motor direto do acúmulo de
4205 carbono, mas sim como um indicador de condições ambientais propícias à sua permanência.

4206 A RMC emergiu como outra variável importante para o acúmulo de SOC, com forte
4207 expressão em regiões tropicais e de transição entre savana e floresta. Sua dominância funcional
4208 reflete a centralidade das interações planta-fungo na regulação do carbono, especialmente em
4209 solos altamente intemperizados, nos quais a simbiose com raízes finas promove a produção de
4210 exsudatos e a formação de agregados estáveis (Hawkins et al., 2023; Phillips et al., 2013). A
4211 presença da RMC como preditor de destaque representa um avanço conceitual em relação a
4212 modelos baseados em índices de vegetação derivados exclusivamente de sensoriamento remoto
4213 (Hengl et al., 2017; Poggio et al., 2021), que não captam adequadamente a complexidade
4214 funcional da biota edáfica.

4215 A fração arenosa (Sand) apresentou contribuições negativas amplamente distribuídas,
4216 especialmente em regiões frias e áridas, como a Sibéria, o Saara e o interior da Austrália. Nesses
4217 contextos, a baixa capacidade de retenção de água e a reduzida agregação física dificultam a
4218 permanência da matéria orgânica. No entanto, em algumas situações, a variável areia foi
4219 preponderante em zonas de acúmulo de SOC, como solos arenosos sob vegetação menos
4220 esparça, zonas ripárias ou ecossistemas transicionais com contrastes texturais marcantes,
4221 atuando como um marcador espacial da heterogeneidade do sistema. Esses casos podem refletir

4222 influência aluvial ou deposicional local, e não uma propriedade intrínseca da textura arenosa
4223 (Van-Camp et al., 2004).

4224 A capacidade de troca de cátions (CEC) também exibiu comportamento dependente do
4225 ambiente. Em regiões tropicais úmidas, associou-se com a redução do SOC, sugerindo baixa
4226 proteção mineral da matéria orgânica em solos intensamente intemperizados, frequentemente
4227 afetados por práticas agrícolas que removem frações orgânicas mais facilmente decomponível
4228 (L. Chen et al., 2021; Hu et al., 2024; Zhao et al., 2024). Já em ambientes frios, a CEC
4229 preponderou a contribuição no aumento dos estoques, atuando como indicadora de solos com
4230 maior reatividade superficial, nos quais as frações minerais contribuem para permanência
4231 organomineral efetiva.

4232 A topografia, expressa pelo modelo digital de elevação (DEM), apresentou importância
4233 reduzida em escala global, mas com efeitos marcantes em regiões montanhosas e ilhas isoladas.
4234 Nas cordilheiras dos Andes, Himalaia, Montanhas Rochosas e no Rift Africano, observou-se
4235 contribuição positiva da altitude para o acúmulo de SOC. Esses efeitos foram atribuídos tanto
4236 à influência climática altitudinal, como menor temperatura e maior umidade, quanto à
4237 deposição de matéria orgânica em zonas côncavas, resultado de processos erosivos e
4238 redistribuição de sedimentos (Tsui et al., 2013; Wiesmeier et al., 2019). Em contraste, a ausência
4239 de relevo em grandes planícies diluiu sua importância, limitando sua capacidade de atuar como
4240 fator redistributivo.

4241 As variáveis associadas à respiração do solo exibiram padrões espaciais divergentes. Em
4242 regiões tropicais com elevada produtividade primária, os fluxos respiratórios foram
4243 compensados pelo alto aporte de resíduos orgânicos, resultando em contribuições positivas para
4244 o SOC (Reichstein et al., 2003). Por outro lado, em altas latitudes, a respiração microbiana teve
4245 efeitos negativos, coerentes com a redução da matéria orgânica. Nessas regiões, alterações

4246 climáticas podem intensificar a decomposição, ampliando as perdas de carbono do solo (Liu et
4247 al., 2024).

4248 Os atributos radiculares demonstraram contribuições positivas generalizadas,
4249 especialmente em regiões tropicais e temperadas, confirmando o papel da biomassa radicular
4250 na entrada e permanência do carbono no solo (Hawkins et al., 2023; Li et al., 2023). Esses
4251 efeitos foram mais intensos em ambientes com vegetação densa e regime hídrico favorável. Em
4252 zonas desérticas, os efeitos negativos foram atribuídos à limitação do aporte de carbono
4253 decorrente da baixa produtividade primária (Kohler et al., 2017).

4254 Esses resultados sustentam que, embora o clima e o solo sejam os principais
4255 controladores globais dos estoques de SOC, raízes e fungos desempenham papéis fundamentais
4256 como controladores locais em ecossistemas com alta produtividade primária ou
4257 hidrologicamente singulares. A combinação entre espacialização dos valores de Shapley, a
4258 importância global dos preditores e a dependência parcial revelou hierarquias ecológicas e
4259 pedológicas robustas. Variáveis como o nitrogênio (N), a RMC, a BIO 12, a CEC e a fração
4260 arenosa emergiram como preditores dominantes em diferentes regiões, com efeitos fortemente
4261 condicionados pelas interações entre os fatores ambientais locais. Esse padrão demonstra o
4262 potencial dos modelos explicáveis para elucidar os mecanismos que regulam o ciclo do carbono
4263 no solo e orientar estratégias eficazes de conservação e manejo baseadas na natureza.

4264

4265 **4.5. Inteligência artificial explicável aplicada à ciência do solo**

4266 Os resultados deste estudo contribuem para o avanço da inteligência artificial explicável
4267 na ciência do solo, respondendo a recomendações centrais. Trabalhos recente, destacaram os
4268 riscos de uma aplicação puramente preditiva de algoritmos de machine learning, que ignoram
4269 o conhecimento pedoecológico acumulado e reduz os modelos à função de “caixa-preta”
4270 (Minasny & McBratney, 2025; Wadoux, 2025). Em contraste, este estudo exemplificou uma

4271 aplicação integrada, na qual os preditores foram analisados com base não apenas em sua
4272 capacidade preditiva, mas também em sua coerência funcional e espacial.

4273 A análise combinada da contribuição marginal por valores de Shapley, dos mapas de
4274 dominância local e das curvas de dependência parcial permitiu enxergar padrões complexos e,
4275 por vezes, contraintuitivos, materializando a “explicabilidade” em escala geográfica. O caso do
4276 nitrogênio (N) é ilustrativo: associado ao acúmulo de SOC em regiões frias, mas à sua perda
4277 em ambientes áridos e tropicais. O N demonstrou um comportamento bifásico que só pode ser
4278 compreendido dentro de contextos pedoecológicos específicos. Esse tipo de resultado vai além
4279 da descrição estatística e permite inferência funcional, dando suporte à geração de hipóteses
4280 sobre os mecanismos de entrada, proteção e perda de SOC.

4281 Além disso, incorporou-se preditores com base biológica direta, como traços funcionais
4282 de raízes, atributos fúngicos do solo e métricas respiratórias, o que supera a limitação apontada
4283 por Wadoux (2025) quanto à dependência de *proxies* biológicos derivados exclusivamente de
4284 sensoriamento remoto. Esse detalhamento permitiu identificar, por exemplo, a taxa de
4285 colonização micorrízica (RMC) como principal determinante dos estoques de SOC em regiões
4286 tropicais, algo raramente mensurado em estudos globais. Ao fazê-lo, o modelo reforçou o
4287 potencial do *machine learning* não apenas para prever, mas para descobrir mecanismos e
4288 relações pedoecológicas antes não quantificados. Com isso, este estudo representa uma
4289 contribuição para o desenvolvimento de abordagens de aprendizado de máquina mais sensíveis
4290 aos processos do solo, fundamentadas na ecologia do sistema edáfico e voltadas à explicação.

4291 Em síntese, os achados deste estudo demonstraram que é possível aliar desempenho
4292 preditivo à interpretação pedoecológica dos fatores que regulam os estoques globais de SOC.
4293 Ao empregar ferramentas de inteligência artificial explicável foi possível revelar relações não
4294 lineares e espacialmente condicionadas entre variáveis-chave, como o comportamento ambíguo

4295 do nitrogênio, a centralidade das micorrizas em solos tropicais e os efeitos divergentes da CEC
4296 e da textura conforme o contexto ambiental.

4297 Esses resultados dialogam com a literatura recente, que propõe o uso da inteligência
4298 artificial como aliada do raciocínio científico, e não apenas como ferramenta preditiva. Ao se
4299 basear em processos e interações pedoecológicas, este estudo contribui para o aprimoramento
4300 da modelagem global de SOC e para o fortalecimento de abordagens explicativas, fundamentais
4301 para orientar estratégias baseadas na natureza diante dos desafios climáticos.

4302

4303 **4.6. Limitações e perspectivas para o avanço da modelagem global de carbono no solo**

4304 Uma das principais limitações identificadas refere-se à sub-representação de
4305 ecorregiões-chave, como as florestas boreais e a taiga, que concentram alguns dos maiores
4306 estoques de SOC. Nesses ambientes, a escassez de dados não se restringe ao teor de SOC,
4307 frequentemente disponível, mas recai sobre variáveis indispensáveis ao cálculo do estoque,
4308 como densidade do solo e fração grosseira.

4309 A obtenção destes dados enfrenta obstáculos logísticos e técnicos significativos em
4310 campo. A coleta de anéis volumétricos para determinação da densidade do solo, por exemplo,
4311 é dificultada em solos com estrutura frágil, presença intensa de raízes, cascalho, blocos rochosos
4312 ou em solos congelados, onde o uso de anéis é tecnicamente limitado ou inviável. Já a
4313 determinação da fração grosseira exige coleta, triagem e pesagem do material em campo, um
4314 processo moroso e de alta complexidade operacional, especialmente em regiões remotas ou
4315 com infraestrutura precária. Dessa forma, muitas lacunas nos bancos de dados globais não
4316 decorrem de falta de interesse científico, mas da complexidade prática envolvida na coleta de
4317 dados essenciais.

4318 Outra limitação está na baixa disponibilidade de dados bióticos diretos, como traços
4319 funcionais de raízes, atributos de fungos do solo e informações sobre comunidades microbianas.

4320 Esses dados são fundamentais para compreender os mecanismos de entrada, transformação e
4321 permanência do SOC, mas seguem escassos ou inconsistentes. Além disso, sua coleta é
4322 tecnicamente exigente, onerosa e carece de padronização metodológica em escala global. Esse
4323 panorama evidencia a fragilidade de uma modelagem que ainda depende fortemente de *proxies*
4324 indiretos e de bases incompletas.

4325 Esse quadro é agravado pela ausência de estrutura e incentivo institucional à ciência
4326 aberta. Muitos dados gerados com financiamento público continuam inacessíveis, dispersos ou
4327 desprovidos de metadados adequados, dificultando sua reutilização. A falta de políticas robustas
4328 de compartilhamento, alinhadas aos princípios FAIR (*Findable, Accessible, Interoperable,*
4329 *Reusable*), compromete tanto a replicabilidade científica quanto a evolução coletiva da
4330 modelagem pedológica em larga escala. Diante desses desafios, é fundamental que haja:

4331 i. Apoio a campanhas de campo específicas, voltadas à coleta padronizada de densidade
4332 do solo, fração grosseira e atributos biológicos, com foco em regiões sub-representadas ou com
4333 condições edafoclimáticas desafiadoras;

4334 ii. Criação e consolidação de repositórios especializados, interoperáveis e vinculados a
4335 redes institucionais consolidadas, como Sistema Global de Informação sobre Solos, GLOSI
4336 (<https://www.fao.org/global-soil-partnership/areas-of-work/soil-information-and-data/en/>), a
4337 Rede Global de Laboratórios de Solos – GLOSOLAN (<https://www.fao.org/global-soil-partnership/glosolan/en/>), a Rede de Observação da Biodiversidade do Solo – SoilBON
4338 (<https://geobon.org/bons/thematic-bon/soil-bon/>) e a Rede de Biodiversidade Funcional do Solo
4339 – FunFun (<https://www.wur.nl/en/research-results/chair-groups/environmental-sciences/soil-biology-group/research/research-topics/functional-soil-biodiversity.htm>), com foco específico
4340 em dados edafobiológicos;
4341
4342

4343 iii. Desenvolvimento de modelos híbridos, que combinem o poder preditivo de
4344 algoritmos de *machine learning* com regras baseadas em processos pedoecológicos reais, o que
4345 promoverá maior explicabilidade funcional;

4346 iv. Fomento à cultura da ciência aberta, cooperativa e distribuída, em que os dados de
4347 solo sejam tratados como bens científicos comuns, com protocolos claros de publicação,
4348 validação e atualização contínua;

4349 v. Consolidação no uso da inteligência artificial explicável na ciência do solo, para isso
4350 é necessário integrá-la à interpretação funcional dos modelos, utilizando ferramentas como
4351 valores de Shapley, curvas de dependência parcial e mapas de dominância local. Essa
4352 abordagem permite revelar relações ecológicas, hierarquias funcionais e padrões de controle
4353 edáfico-biológico, tornando a IA um instrumento de descoberta científica, e não apenas de
4354 predição. Esse avanço depende da capacitação técnica da comunidade, da oferta de ferramentas
4355 auditáveis e da validação empírica dos padrões via estudos de campo. Quando transparente e
4356 contextualizada, a IA explicável pode fortalecer o raciocínio pedológico e tornar a ciência do
4357 solo mais robusta frente às mudanças climáticas.

4358 Em síntese, as limitações enfrentadas nesta pesquisa não indicam apenas lacunas
4359 técnicas de modelagem, mas revelam a necessidade de amadurecimento institucional e
4360 científico. A construção de estimativas globais de SOC que sejam simultaneamente precisas,
4361 úteis e ecologicamente coerentes exige mais do que algoritmos sofisticados: exige dados
4362 confiáveis, coleta coordenada, infraestrutura de compartilhamento e uma leitura crítica e
4363 sistêmica dos processos que regem as interações solo–biota–clima.

4364 Nesse contexto, a inteligência artificial explicável representa um ponto de inflexão
4365 metodológico, ao permitir que algoritmos complexos revelem hierarquias funcionais, limiares
4366 ecológicos e padrões espaciais de controle biológico e edáfico. Para que essa abordagem
4367 realmente amplie o conhecimento pedológico, é fundamental investir na capacitação técnica da

4368 comunidade, na disponibilização de ferramentas auditáveis e acessíveis, e na validação
4369 empírica dos padrões identificados, por meio de estudos de campo bem conduzidos. Sem isso,
4370 corremos o risco de continuar descrevendo médias globais enquanto os extremos permanecem
4371 invisíveis aos nossos modelos.

4372

4373 **5. Conclusão**

4374 O modelo ajustado apresentou desempenho expressivamente superior ao modelo nulo,
4375 demonstrando alta acurácia geral, apesar de maiores erros na predição de valores extremos. A
4376 incorporação direta de variáveis edafobiológicas proporcionou uma abordagem funcional e
4377 biologicamente fundamentada, o que permitiu quantificar processos até então negligenciados
4378 em estudos baseados exclusivamente em *proxies* indiretas obtidas por sensoriamento remoto.
4379 Nesse sentido, a identificação da taxa de colonização micorrízica (RMC) como principal
4380 preditor em ambientes tropicais representa um ponto de inflexão conceitual marcante: torna-se
4381 claro que a modelagem dos estoques de carbono orgânico do solo (SOC) não pode continuar
4382 desconsiderando o papel crítico desempenhado pela biota do solo na formação, transformação
4383 e permanência desse carbono.

4384 Além disso, a integração de métricas de importância de variáveis, análises espaciais e
4385 abordagens explicativas dos modelos (valores de Shapley, curvas de dependência parcial e
4386 análises de dominância local) permitiu revelar hierarquias funcionais e padrões condicionais
4387 complexos entre clima, solo e biota. Destacam-se especialmente as interações entre textura do
4388 solo, regime hídrico e atividade simbiótica, cujos efeitos frequentemente não lineares enfatizam
4389 que condições ambientais favoráveis ao acúmulo de SOC em determinadas regiões podem agir
4390 inversamente em outras.

4391 As limitações evidenciadas reforçam que, sem bancos de dados robustos, interoperáveis
4392 e globalmente coordenados, a modelagem seguirá incompleta e suscetível a vieses

4393 significativos. Finalmente, este estudo ressaltou que a integração da inteligência artificial
4394 explicável com o conhecimento pedológico oferece uma abordagem promissora para
4395 desenvolver modelos cientificamente úteis, pedoecologicamente coerentes e tecnicamente
4396 sólidos. A aplicação dessa metodologia transcende a simples predição numérica, priorizando a
4397 compreensão fundamentada em dados confiáveis e hipóteses claramente articuladas.

4398

4399 **Referências**

- 4400 Barkov, V., Schmidinger, J., Gebbers, R., & Atzmueller, M. (2024). An Efficient Model-
4401 Agnostic Approach for Uncertainty Estimation in Data-Restricted Pedometric
4402 Applications. *Proceedings - 2024 International Conference on Machine Learning and*
4403 *Applications, ICMLA 2024*, 198–205. <https://doi.org/10.1109/ICMLA61862.2024.00033>
- 4404 Batjes, N. H. (1996). Total carbon and nitrogen in the soils of the world. *European Journal of*
4405 *Soil Science*, 47(2), 151–163. <https://doi.org/10.1111/j.1365-2389.1996.tb01386.x>
- 4406 Batjes, N. H., Ribeiro, E., & van Oostrum, A. J. M. (2019). Standardised soil profile data to
4407 support global mapping and modelling (WoSIS snapshot 2019). *Earth System Science*
4408 *Data Discussions*, 1–46. <https://doi.org/10.5194/essd-2019-164>
- 4409 Beucher, A., Rasmussen, C. B., Moeslund, T. B., & Greve, M. H. (2022). Interpretation of
4410 Convolutional Neural Networks for Acid Sulfate Soil Classification. *Frontiers in*
4411 *Environmental Science*, 9, 809995. <https://doi.org/10.3389/fenvs.2021.809995>
- 4412 Calisto, L., Sousa, L. M., & Batjes, N. H. (2023). Standardised soil profile data for the world
4413 (WoSIS snapshot – December 2023). Em *Http://Dx. Doi. Org* (Vol. 10). ISRIC Data Hub.
4414 <https://doi.org/10.17027/isric-wdcsoils-20231130>
- 4415 Castro, J. P., Gasch, C. K., & Flores, P. (2024). Artificial intelligence in soil science: Where do
4416 we go now? *Agricultural and Environmental Letters*, 9(2), e20134.
4417 <https://doi.org/10.1002/ael2.20134>

- 4418 Chandra, H., Pawar, P. M., Elakkiya, R., Tamizharasan, P. S., Muthalagu, R., & Panthakkan, A.
4419 (2023). Explainable AI for Soil Fertility Prediction. *IEEE Access*, *11*, 97866–97878.
4420 <https://doi.org/10.1109/ACCESS.2023.3311827>
- 4421 Chen, L., Fang, K., Wei, B., Qin, S., Feng, X., Hu, T., Ji, C., & Yang, Y. (2021). Soil carbon
4422 persistence governed by plant input and mineral protection at regional and global scales.
4423 *Ecology Letters*, *24*(5), 1018–1028. <https://doi.org/10.1111/ele.13723>
- 4424 Chen, L., Jing, X., Flynn, D. F. B., Shi, Y., Kühn, P., Scholten, T., & He, J. S. (2017). Changes
4425 of carbon stocks in alpine grassland soils from 2002 to 2011 on the Tibetan Plateau and
4426 their climatic causes. *Geoderma*, *288*, 166–174.
4427 <https://doi.org/10.1016/j.geoderma.2016.11.016>
- 4428 Chen, S., Shuai, Q., Arrouays, D., Chen, Z., Dai, L., Hong, Y., Hu, B., Huang, Y., Ji, W., Li, S.,
4429 Liang, Z., Ma, Y., Richer-de-Forges, A. C., Schillaci, C., Su, Y., Teng, H., Wang, N., Wang,
4430 X., Wang, Y., ... Shi, Z. (2024). *GSOCS-LULCC: the Global Soil Organic Carbon Stock*
4431 *dataset after Land Use and Land Cover Change*. <https://doi.org/10.5194/essd-2024-373>
- 4432 Chen, Z., Wei, X., Ni, X., Wu, F., & Liao, S. (2023). Changing precipitation effect on forest soil
4433 carbon dynamics is driven by different attributes between dry and wet areas. *Geoderma*,
4434 *429*, 116279. <https://doi.org/10.1016/j.geoderma.2022.116279>
- 4435 Cleveland, C. C., & Liptzin, D. (2007). C:N:P stoichiometry in soil: Is there a “Redfield ratio”
4436 for the microbial biomass? *Biogeochemistry*, *85*(3), 235–252.
4437 <https://doi.org/10.1007/s10533-007-9132-0>
- 4438 Cui, Y., Xiao, Q., Wang, B., Li, Z., Zhai, Y., Xie, K., Cao, G., Li, X., & Chen, J. (2024).
4439 Rhizospheric soil bacterial and fungal co-occurrence networks of *Artemisia lavandulifolia*
4440 present opposite stability patterns in volcanic areas. *Rhizosphere*, *31*, 100927.
4441 <https://doi.org/10.1016/j.rhisph.2024.100927>

- 4442 DeLuca, T. H., & Boisvenue, C. (2012). Boreal forest soil carbon: Distribution, function and
4443 modelling. *Forestry*, 85(2), 161–184. <https://doi.org/10.1093/forestry/cps003>
- 4444 Deng, L., Peng, C., Kim, D. G., Li, J., Liu, Y., Hai, X., Liu, Q., Huang, C., Shanguan, Z., &
4445 Kuzyakov, Y. (2021). Drought effects on soil carbon and nitrogen dynamics in global
4446 natural ecosystems. *Earth-Science Reviews*, 214, 103501.
4447 <https://doi.org/10.1016/j.earscirev.2020.103501>
- 4448 Elrys, A. S., Zhang, J., Meng, L., Nardi, P., & Müller, C. (2025). Clay-to-carbon ratio: An
4449 overlooked but pivotal mediator of soil nitrogen mineralization and availability. *Soil and*
4450 *Tillage Research*, 251, 106533. <https://doi.org/10.1016/j.still.2025.106533>
- 4451 Evangelista, S. J., Field, D. J., McBratney, A. B., Minasny, B., Ng, W., Padarian, J., Román
4452 Dobarco, M., & Wadoux, A. M. J. C. (2023). A proposal for the assessment of soil security:
4453 Soil functions, soil services and threats to soil. *Soil Security*, 10, 100086.
4454 <https://doi.org/10.1016/j.soisec.2023.100086>
- 4455 Evangelista, S. J., Field, D. J., McBratney, A. B., Minasny, B., Ng, W., Padarian, J., Román
4456 Dobarco, M., & Wadoux, A. M. J. C. (2024). Soil security—Strategizing a sustainable
4457 future for soil. *Advances in Agronomy*, 183, 1–70.
4458 <https://doi.org/10.1016/bs.agron.2023.10.001>
- 4459 Evans, J. S., & Murphy, M. A. (2015). *rfUtilities: Random Forests Model Selection and*
4460 *Performance Evaluation* (2.1-5). [https://cran.r-](https://cran.r-project.org/web/packages/rfUtilities/index.html)
4461 [project.org/web/packages/rfUtilities/index.html](https://cran.r-project.org/web/packages/rfUtilities/index.html)
- 4462 FAO, & ITPS. (2015). Status of the World's Soil Resources (SWSR) - Main Report. *Em Status*
4463 *of the World's Soil Resources (SWSR) - Main Report. Food and Agriculture Organization*
4464 *of the United Nations and Intergovernmental Technical Panel on Soils, Rome, Italy.* (Vol.
4465 650). Food and agriculture organization of the United Nations and intergovernmental
4466 technical panel on soils. <https://hal.archives-ouvertes.fr/hal-01241064/>

- 4467 Fernandes Filho, E. I., Moquedace, C. M., & Veloso, G. V. (2021). *mdsFuncs: Auxiliar functions*
4468 *to pre-process raster data and modelling with machine learning models using caret*
4469 *package* (0.2.10).
- 4470 Fornara, D. A., & Tilman, D. (2008). Plant functional composition influences rates of soil
4471 carbon and nitrogen accumulation. *Journal of Ecology*, *96*(2), 314–322.
4472 <https://doi.org/10.1111/j.1365-2745.2007.01345.x>
- 4473 Francis Justine, M., Kaiwen, P., Tadesse, Z., Hongyan, Z., & Lin, Z. (2024). Cooling has
4474 stimulated soil carbon storage in forest ecosystems. *Environmental Research*, *245*, 118012.
4475 <https://doi.org/10.1016/j.envres.2023.118012>
- 4476 Friedlingstein, P., O’Sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Landschützer, P., Le
4477 Quéré, C., Li, H., Luijkx, I. T., Olsen, A., Peters, G. P., Peters, W., Pongratz, J.,
4478 Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., ... Zeng,
4479 J. (2025). Global Carbon Budget 2024. *Earth System Science Data*, *17*(3), 965–1039.
4480 <https://doi.org/10.5194/essd-17-965-2025>
- 4481 Gomes, L. C., Faria, R. M., de Souza, E., Veloso, G. V., Schaefer, C. E. G. R., & Filho, E. I. F.
4482 (2019). Modelling and mapping soil organic carbon stocks in Brazil. *Geoderma*,
4483 *340*(January), 337–350. <https://doi.org/10.1016/j.geoderma.2019.01.007>
- 4484 Gorham, E. (1991). Northern peatlands: role in the carbon cycle and probable responses to
4485 climatic warming. *Ecological Applications*, *1*(2), 182–195.
4486 <https://doi.org/10.2307/1941811>
- 4487 Gottschalk, P., Smith, J. U., Wattenbach, M., Bellarby, J., Stehfest, E., Arnell, N., Osborn, T. J.,
4488 Jones, C., & Smith, P. (2012). How will organic carbon stocks in mineral soils evolve
4489 under future climate? Global projections using RothC for a range of climate change
4490 scenarios. *Biogeosciences*, *9*(8), 3151–3171. <https://doi.org/10.5194/bg-9-3151-2012>

- 4491 Greenwell, B. (2021). *fastshap: Fast Approximate Shapley Values* (0.0.7). [https://cran.r-](https://cran.r-project.org/package=fastshap)
4492 [project.org/package=fastshap](https://cran.r-project.org/package=fastshap)
- 4493 Harden, J. W., Mark, R. K., Sundquist, E. T., & Stallard, R. F. (1992). Dynamics of Soil Carbon
4494 During Deglaciation of the Laurentide Ice Sheet. *Science*, 258(5090), 1921–1924.
4495 <https://doi.org/10.1126/science.258.5090.1921>
- 4496 Harris, L. I., Olefeldt, D., Pelletier, N., Blodau, C., Knorr, K. H., Talbot, J., Heffernan, L., &
4497 Turetsky, M. (2023). Permafrost thaw causes large carbon loss in boreal peatlands while
4498 changes to peat quality are limited. *Global Change Biology*, 29(19), 5720–5735.
4499 <https://doi.org/10.1111/gcb.16894>
- 4500 Hawker, L., Uhe, P., Paulo, L., Sosa, J., Savage, J., Sampson, C., & Neal, J. (2022). A 30 m
4501 global map of elevation with forests and buildings removed. *Environmental Research*
4502 *Letters*, 17(2), 024016. <https://doi.org/10.1088/1748-9326/ac4d4f>
- 4503 Hawkins, H. J., Cargill, R. I. M., Van Nuland, M. E., Hagen, S. C., Field, K. J., Sheldrake, M.,
4504 Soudzilovskaia, N. A., & Kiers, E. T. (2023). Mycorrhizal mycelium as a global carbon
4505 pool. *Current Biology*, 33(11), R560–R573. <https://doi.org/10.1016/j.cub.2023.02.027>
- 4506 Hein, C. J., Usman, M., Eglinton, T. I., Haghypour, N., & Galy, V. V. (2020). Millennial-scale
4507 hydroclimate control of tropical soil carbon storage. *Nature*, 581(7806), 63–66.
4508 <https://doi.org/10.1038/s41586-020-2233-9>
- 4509 Helbig, M., Daw, L., Iwata, H., Rudaitis, L., Ueyama, M., & Živković, T. (2024). Boreal Forest
4510 Fire Causes Daytime Surface Warming During Summer to Exceed Surface Cooling During
4511 Winter in North America. *AGU Advances*, 5(5), e2024AV001327.
4512 <https://doi.org/10.1029/2024AV001327>
- 4513 Hengl, T., De Jesus, J. M., Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A.,
4514 Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A.,
4515 Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I.,

- 4516 Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based
4517 on machine learning. *PLoS ONE*, *12*(2), e0169748.
4518 <https://doi.org/10.1371/journal.pone.0169748>
- 4519 Heung, B., Ho, H. C., Zhang, J., Knudby, A., Bulmer, C. E., & Schmidt, M. G. (2016). An
4520 overview and comparison of machine-learning techniques for classification purposes in
4521 digital soil mapping. *Geoderma*, *265*, 62–77.
4522 <https://doi.org/10.1016/j.geoderma.2015.11.014>
- 4523 Heuvelink, G. B. M. (2018). *Uncertainty and Uncertainty Propagation in Soil Mapping and*
4524 *Modelling*. 439–461. https://doi.org/10.1007/978-3-319-63439-5_14
- 4525 Hijmans, R. J. (2023). *terra: Spatial Data Analysis* (1.7-39). <https://rspatial.org/terra/>
- 4526 Hobbie, S. E., Schimel, J. P., Trumbore, S. E., & Randerson, J. R. (2000). Controls over carbon
4527 storage and turnover in high-latitude soils. *Global Change Biology*, *6*(SUPPLEMENT 1),
4528 196–210. <https://doi.org/10.1046/j.1365-2486.2000.06021.x>
- 4529 Hu, S., Zheng, L., Zhang, H., Yang, Y., Chen, G., Meng, H., Cheng, K., Guo, C., Wang, Y., Li,
4530 X., & Liu, T. (2024). Sequestration of Labile Organic Matter by Secondary Fe Minerals
4531 from Chemodenitrification: Insight into Mineral Protection Mechanisms. *Environmental*
4532 *Science and Technology*, *58*(25), 11003–11015. <https://doi.org/10.1021/acs.est.3c10134>
- 4533 Hugelius, G., Strauss, J., Zubrzycki, S., Harden, J. W., Schuur, E. A. G., Ping, C. L.,
4534 Schirmer, L., Grosse, G., Michaelson, G. J., Koven, C. D., O'Donnell, J. A., Elberling,
4535 B., Mishra, U., Camill, P., Yu, Z., Palmtag, J., & Kuhry, P. (2014). Estimated stocks of
4536 circumpolar permafrost carbon with quantified uncertainty ranges and identified data gaps.
4537 *Biogeosciences*, *11*(23), 6573–6593. <https://doi.org/10.5194/BG-11-6573-2014>
- 4538 Jiang, S., Hu, X., Kang, Y., Xie, C., An, X., Dong, C., Xu, Y., & Shen, Q. (2020). Arbuscular
4539 mycorrhizal fungal communities in the rhizospheric soil of litchi and mango orchards as

- 4540 affected by geographic distance, soil properties and manure input. *Applied Soil Ecology*,
4541 152, 103593. <https://doi.org/10.1016/j.apsoil.2020.103593>
- 4542 Jones, M. C., Harden, J., O'Donnell, J., Manies, K., Jorgenson, T., Treat, C., & Ewing, S.
4543 (2017). Rapid carbon loss and slow recovery following permafrost thaw in boreal
4544 peatlands. *Global Change Biology*, 23(3), 1109–1127. <https://doi.org/10.1111/gcb.13403>
- 4545 Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann,
4546 N. E., Linder, H. P., & Kessler, M. (2017). Climatologies at high resolution for the earth's
4547 land surface areas. *Scientific Data*, 4(1), 1–20. <https://doi.org/10.1038/sdata.2017.122>
- 4548 Kim, J., Kim, Y., Nam, S., Jung, J. Y., Kim, Y. J., Hwang, J. H., & Kim, M. (2025). Methane
4549 trapping in permafrost soils: a biogeochemical dataset across Alaskan boreal-Arctic
4550 gradient. *Scientific data*, 12(1), 110. <https://doi.org/10.1038/s41597-025-04463-5>
- 4551 Kohler, J., Roldán, A., Campoy, M., & Caravaca, F. (2017). Unraveling the role of hyphal
4552 networks from arbuscular mycorrhizal fungi in aggregate stabilization of semiarid soils
4553 with different textures and carbonate contents. *Plant and Soil*, 410(1–2), 273–281.
4554 <https://doi.org/10.1007/s11104-016-3001-3>
- 4555 Kuhn, M. (2022). *caret: Classification and Regression Training* (6.0-91).
4556 <https://doi.org/10.1887/0750303123/b365c43>
- 4557 Lal, R. (2004). Agricultural activities and the global carbon cycle. *Nutrient Cycling in*
4558 *Agroecosystems*, 70(2), 103–116. <https://doi.org/10.1023/B:FRES.0000048480.24274.0f>
- 4559 Lal, R. (2010). Managing soils and ecosystems for mitigating anthropogenic carbon emissions
4560 and advancing global food security. *BioScience*, 60(9), 708–721.
4561 <https://doi.org/10.1525/bio.2010.60.9.8>
- 4562 Lal, R. (2014). Soil conservation and ecosystem services. *International Soil and Water*
4563 *Conservation Research*, 2(3), 36–47. [https://doi.org/10.1016/S2095-6339\(15\)30021-6](https://doi.org/10.1016/S2095-6339(15)30021-6)

- 4564 Li, T., Yuan, Y., Mou, Z., Li, Y., Kuang, L., Zhang, J., Wu, W., Wang, F., Wang, J., Lambers, H.,
4565 Sardans, J., Peñuelas, J., Ren, H., & Liu, Z. (2023). Faster accumulation and greater
4566 contribution of glomalin to the soil organic carbon pool than amino sugars do under
4567 tropical coastal forest restoration. *Global Change Biology*, 29(2), 533–546.
4568 <https://doi.org/10.1111/gcb.16467>
- 4569 Liu, J., Hu, J., Liu, H., & Han, K. (2024). Global soil respiration estimation based on ecological
4570 big data and machine learning model. *Scientific Reports*, 14(1), 1–17.
4571 <https://doi.org/10.1038/s41598-024-64235-w>
- 4572 Mayer, M. (2025). *shapviz: SHAP Visualizations. Version 0.9.3 (0.9.7)*. [https://cran.r-](https://cran.r-project.org/package=shapviz)
4573 [project.org/package=shapviz](https://cran.r-project.org/package=shapviz)
- 4574 McBratney, A. B., Mendonça Santos, M. L., & Minasny, B. (2003). On digital soil mapping.
4575 *Geoderma*, 117(1–2), 3–52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)
- 4576 McBratney, A., Field, D. J., & Koch, A. (2014). The dimensions of soil security. *Geoderma*,
4577 213, 203–213. <https://doi.org/10.1016/j.geoderma.2013.08.013>
- 4578 Meinshausen, N. (2017). *quantregForest: Quantile Regression Forests (1.3-7)*. [https://cran.r-](https://cran.r-project.org/package=quantregForest)
4579 [project.org/package=quantregForest](https://cran.r-project.org/package=quantregForest)
- 4580 Mello, D. C., Francelino, M. R., Moquedace, C. M., Baldi, C. G. O., Silva, L. V., Siqueira, R.
4581 G., Veloso, G. V., Fernandes-Filho, E. I., Thomazini, A., Demattê, J. A. M., Ferreira, T. O.,
4582 Gomes, L. C., Senra, E. O., & Schaefer, C. E. G. R. (2025). Global warming may turn ice-
4583 free areas of Maritime and Peninsular Antarctica into potential soil organic carbon sinks.
4584 *Communications Earth and Environment*, 6(1), 1–14. [https://doi.org/10.1038/s43247-024-](https://doi.org/10.1038/s43247-024-01937-z)
4585 [01937-z](https://doi.org/10.1038/s43247-024-01937-z)
- 4586 Minasny, B., Bandai, T., Ghezzehei, T. A., Huang, Y. C., Ma, Y., McBratney, A. B., Ng, W.,
4587 Norouzi, S., Padarian, J., Rudiyanto, Sharififar, A., Styc, Q., & Widyastuti, M. (2024). Soil

- 4588 Science-Informed Machine Learning. *Geoderma*, 452, 117094.
4589 <https://doi.org/10.1016/j.geoderma.2024.117094>
- 4590 Minasny, B., Malone, B. P., McBratney, A. B., Angers, D. A., Arrouays, D., Chambers, A.,
4591 Chaplot, V., Chen, Z. S., Cheng, K., Das, B. S., Field, D. J., Gimona, A., Hedley, C. B.,
4592 Hong, S. Y., Mandal, B., Marchant, B. P., Martin, M., McConkey, B. G., Mulder, V. L., ...
4593 Winowiecki, L. (2017). Soil carbon 4 per mille. *Geoderma*, 292, 59–86.
4594 <https://doi.org/10.1016/j.geoderma.2017.01.002>
- 4595 Minasny, B., & McBratney, A. (2024). Nobel Prize in Physics, and Soil Science. *CSA News*,
4596 69(12), 30–33. <https://doi.org/10.1002/csan.21427>
- 4597 Minasny, B., & McBratney, A. B. (2025). Machine Learning and Artificial Intelligence
4598 Applications in Soil Science. *European Journal of Soil Science*, 76(2), e70093.
4599 <https://doi.org/10.1111/ejss.70093>
- 4600 Moquedace, C. M., Baldi, C. G. O., Siqueira, R. G., Cardoso, I. M., Souza, E. F. M. de, Fontes,
4601 R. L. F., Francelino, M. R., Gomes, L. C., & Fernandes-Filho, E. I. (2024). High-resolution
4602 mapping of soil carbon stocks in the western Amazon. *Geoderma Regional*, 36, e00773.
4603 <https://doi.org/10.1016/j.geodrs.2024.e00773>
- 4604 Naorem, A., Jayaraman, S., Dang, Y. P., Dalal, R. C., Sinha, N. K., Rao, C. S., & Patra, A. K.
4605 (2023). Soil Constraints in an Arid Environment—Challenges, Prospects, and
4606 Implications. *Agronomy*, 13(1), 220. <https://doi.org/10.3390/agronomy13010220>
- 4607 Nicolás, C., Hernández, T., & García, C. (2012). Organic amendments as strategy to increase
4608 organic matter in particle-size fractions of a semi-arid soil. *Applied Soil Ecology*, 57, 50–
4609 58. <https://doi.org/10.1016/j.apsoil.2012.02.018>
- 4610 Nyaupane, K., Mishra, U., Tao, F., Yeo, K., Riley, W. J., Hoffman, F. M., & Gautam, S. (2023).
4611 Observational benchmarks inform representation of soil organic carbon dynamics in land

- 4612 surface models. *Biogeosciences*, 21(22), 5173–5183. <https://doi.org/10.5194/BG-21->
4613 5173-2024
- 4614 O'Brien, L. (2022). *mpspline2: Mass-Preserving Spline Functions for Soil Data* (0.1.6).
4615 <https://cran.r-project.org/package=mpspline2>
- 4616 Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N.,
4617 Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J.,
4618 Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., &
4619 Kassem, K. R. (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth:
4620 A new global map of terrestrial ecoregions provides an innovative tool for conserving
4621 biodiversity. *BioScience*, 51(11), 933–938. <https://doi.org/https://doi.org/10.1641/0006->
4622 3568(2001)051[0933:TEOTWA]2.0.CO;2
- 4623 O'Rourke, S. M., Angers, D. A., Holden, N. M., & Mcbratney, A. B. (2015). Soil organic carbon
4624 across scales. *Global Change Biology*, 21(10), 3561–3574.
4625 <https://doi.org/10.1111/gcb.12959>
- 4626 Padarian, J., Minasny, B., & McBratney, A. B. (2020). Machine learning and soil sciences: A
4627 review aided by machine learning tools. *Soil*, 6(1), 35–52. <https://doi.org/10.5194/soil-6->
4628 35-2020
- 4629 Padarian, J., Minasny, B., McBratney, A., & Smith, P. (2022). Soil carbon sequestration
4630 potential in global croplands. *PeerJ*, 10, e13740. <https://doi.org/10.7717/peerj.13740>
- 4631 Pellegrini, A. F. A., Ahlström, A., Hobbie, S. E., Reich, P. B., Nieradzik, L. P., Staver, A. C.,
4632 Scharenbroch, B. C., Jumpponen, A., Anderegg, W. R. L., Randerson, J. T., & Jackson, R.
4633 B. (2018). Fire frequency drives decadal changes in soil carbon and nitrogen and
4634 ecosystem productivity. *Nature*, 553(7687), 194–198. <https://doi.org/10.1038/nature24668>
- 4635 Peng, Y., Schmidt, I. K., Zheng, H., Heděnc, P., Bachega, L. R., Yue, K., Wu, F., & Vesterdal,
4636 L. (2020). Tree species effects on topsoil carbon stock and concentration are mediated by

- 4637 tree species type, mycorrhizal association, and N-fixing ability at the global scale. *Forest*
4638 *Ecology and Management*, 478, 118510. <https://doi.org/10.1016/j.foreco.2020.118510>
- 4639 Phillips, R. P., Brzostek, E., & Midgley, M. G. (2013). The mycorrhizal-associated nutrient
4640 economy: A new framework for predicting carbon-nutrient couplings in temperate forests.
4641 *New Phytologist*, 199(1), 41–51. <https://doi.org/10.1111/nph.12221>
- 4642 Piao, S., Ciais, P., Friedlingstein, P., De Noblet-Ducoudré, N., Cadule, P., Viovy, N., & Wang,
4643 T. (2009). Spatiotemporal patterns of terrestrial carbon cycle during the 20th century.
4644 *Global Biogeochemical Cycles*, 23(4). <https://doi.org/10.1029/2008GB003339>
- 4645 Poeplau, C., Don, A., Vesterdal, L., Leifeld, J., Van Wesemael, B., Schumacher, J., & Gensior,
4646 A. (2011). Temporal dynamics of soil organic carbon after land-use change in the
4647 temperate zone - carbon response functions as a model approach. *Global Change Biology*,
4648 17(7), 2415–2427. <https://doi.org/10.1111/j.1365-2486.2011.02408.x>
- 4649 Poggio, L., De Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., &
4650 Rossiter, D. (2021). SoilGrids 2.0: Producing soil information for the globe with quantified
4651 spatial uncertainty. *Soil*, 7(1), 217–240. <https://doi.org/10.5194/soil-7-217-2021>
- 4652 Poggio, L., & Gimona, A. (2014). National scale 3D modelling of soil organic carbon stocks
4653 with uncertainty propagation - An example from Scotland. *Geoderma*, 232–234, 284–299.
4654 <https://doi.org/10.1016/j.geoderma.2014.05.004>
- 4655 Poggio, L., Simonetti, E., & Gimona, A. (2018). Enhancing the WorldClim data set for national
4656 and regional applications. *Science of The Total Environment*, 625, 1628–1643.
4657 <https://doi.org/10.1016/j.scitotenv.2017.12.258>
- 4658 R Core Team. (2025). R: A Language and Environment for Statistical Computing. Em *R*
4659 *Foundation for Statistical Computing* (4.4.3). <http://www.r-project.org>
- 4660 Reichenbach, M., Doetterl, S., Van Oost, K., Wilken, F., Muhindo, D., Bamba, F., & Fiener, P.
4661 (2025). Low but highly variable soil organic carbon stocks across deeply weathered

- 4662 eroding and depositional tropical landforms. *Catena*, 254, 108971.
4663 <https://doi.org/10.1016/j.catena.2025.108971>
- 4664 Richards, A. E., Dalal, R. C., & Schmidt, S. (2007). Soil carbon turnover and sequestration in
4665 native subtropical tree plantations. *Soil Biology and Biochemistry*, 39(8), 2078–2090.
4666 <https://doi.org/10.1016/j.soilbio.2007.03.012>
- 4667 Schirrmeister, L., Kunitsky, V., Grosse, G., Wetterich, S., Meyer, H., Schwamborn, G., Babiy,
4668 O., Derevyagin, A., & Siegert, C. (2011). Sedimentary characteristics and origin of the
4669 Late Pleistocene Ice Complex on north-east Siberian Arctic coastal lowlands and islands -
4670 A review. *Quaternary International*, 241(1–2), 3–25.
4671 <https://doi.org/10.1016/j.quaint.2010.04.004>
- 4672 Schütte, U. M. E., Henning, J. A., Ye, Y., Bowling, A., Ford, J., Genet, H., Waldrop, M. P.,
4673 Turetsky, M. R., White, J. R., & Bever, J. D. (2019). Effect of permafrost thaw on plant
4674 and soil fungal community in a boreal forest: Does fungal community change mediate
4675 plant productivity response? *Journal of Ecology*, 107(4), 1737–1752.
4676 <https://doi.org/10.1111/1365-2745.13139>
- 4677 Shapley, L. S. (1953). *A Value for n-Person Games*, in: *Contributions to the Theory of Games*
4678 (K. Harold William & T. Albert William, Orgs.; Vol. 28, A). Princeton University Press,
4679 Princeton.
- 4680 Singh, S. K., Pandey, C. B., Sidhu, G. S., Sarkar, D., & Sagar, R. (2011). Concentration and
4681 stock of carbon in the soils affected by land uses and climates in the western Himalaya,
4682 India. *Catena*, 87(1), 78–89. <https://doi.org/10.1016/j.catena.2011.05.008>
- 4683 Siqueira, R. G., Moquedace, C. M., Fernandes-Filho, E. I., Schaefer, C. E. G. R., Francelino,
4684 M. R., Sacramento, I. F., & Michel, R. F. M. (2024). Modelling and prediction of major
4685 soil chemical properties with Random Forest: Machine learning as tool to understand soil-

- 4686 environment relationships in Antarctica. *Catena*, 235, 107677.
4687 <https://doi.org/10.1016/j.catena.2023.107677>
- 4688 Siqueira, R. G., Moquedace, C. M., Francelino, M. R., Schaefer, C. E. G. R., & Fernandes-
4689 Filho, E. I. (2023). Machine learning applied for Antarctic soil mapping: Spatial prediction
4690 of soil texture for Maritime Antarctica and Northern Antarctic Peninsula. *Geoderma*, 432,
4691 116405. <https://doi.org/10.1016/j.geoderma.2023.116405>
- 4692 Smith, L. C., MacDonald, G. M., Velichko, A. A., Beilman, D. W., Borisova, O. K., Frey, K. E.,
4693 Kremenetski, K. V., & Sheng, Y. (2004). Siberian Peatlands a Net Carbon Sink and Global
4694 Methane Source since the Early Holocene. *Science*, 303(5656), 353–356.
4695 <https://doi.org/10.1126/science.1090553>
- 4696 Tang, Q., Li, Q., Tong, L., Wu, R., & Xu, J. (2023). Rhizospheric soil organic carbon
4697 accumulated but its molecular groups redistributed via rhizospheric soil microorganisms
4698 along multi-root *Cerasus humilis* plantation chronosequence at the karst rocky
4699 desertification control area. *Environmental Science and Pollution Research*, 30(28),
4700 72993–73007. <https://doi.org/10.1007/s11356-023-27588-9>
- 4701 Tennekes, M. (2018). tmap: Thematic Maps in R. *Journal of Statistical Software*, 84(6).
4702 <https://doi.org/10.18637/jss.v084.i06>
- 4703 Tsui, C. C., Tsai, C. C., & Chen, Z. S. (2013). Soil organic carbon stocks in relation to elevation
4704 gradients in volcanic ash soils of Taiwan. *Geoderma*, 209–210, 119–127.
4705 <https://doi.org/10.1016/j.geoderma.2013.06.013>
- 4706 Turing, A. M. (2009). Computing machinery and intelligence. *Parsing the Turing Test:
4707 Philosophical and Methodological Issues in the Quest for the Thinking Computer*, 23–65.
4708 https://doi.org/10.1007/978-1-4020-6710-5_3
- 4709 USDA. (2023). *Soil Organic Carbon Stock Monitoring - CEMA 221* (Número October, p. 1–
4710 8). Natural Resources Conservation Service. <https://www.nrcs.usda.gov>

- 4711 Van Groenigen, J. W., Huygens, D., Boeckx, P., Kuyper, T. W., Lubbers, I. M., Rütting, T., &
4712 Groffman, P. M. (2015). The soil n cycle: New insights and key challenges. *Soil*, *1*(1),
4713 235–256. <https://doi.org/10.5194/soil-1-235-2015>
- 4714 Van-Camp, L., Bujarrabal, B., Gentile, A. R., Jones, R. J. A., Montanerella, L., Olazabal, C., &
4715 Selvaradjou, S.-K. (2004). *Establishment under the thematic strategy for soil protection*.
4716 645.
- 4717 Wadoux, A. M. J. C. (2025). Artificial intelligence in soil science. *European Journal of Soil*
4718 *Science*, *76*(2), e70080. <https://doi.org/10.1111/ejss.70080>
- 4719 Wadoux, A. M. J. C., & Molnar, C. (2022). Beyond prediction: methods for interpreting
4720 complex models of soil variation. *Geoderma*, *422*, 115953.
4721 <https://doi.org/10.1016/j.geoderma.2022.115953>
- 4722 Wadoux, A. M. J. C., Saby, N. P. A., & Martin, M. P. (2023). Shapley values reveal the drivers
4723 of soil organic carbon stock prediction. *Soil*, *9*(1), 21–38. [https://doi.org/10.5194/soil-9-](https://doi.org/10.5194/soil-9-21-2023)
4724 [21-2023](https://doi.org/10.5194/soil-9-21-2023)
- 4725 Wadoux, A. M. J. C., Samuel-Rosa, A., Poggio, L., & Mulder, V. L. (2020). A note on
4726 knowledge discovery and machine learning in digital soil mapping. *European Journal of*
4727 *Soil Science*, *71*(2), 133–136. <https://doi.org/10.1111/ejss.12909>
- 4728 Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Em *Journal of Statistical*
4729 *Software* (Vol. 35, Número Book Review 1). Springer-Verlag New York.
- 4730 Wiesmeier, M., Urbanski, L., Hobley, E., Lang, B., von Lützow, M., Marin-Spiotta, E., van
4731 Wesemael, B., Rabot, E., Ließ, M., Garcia-Franco, N., Wollschläger, U., Vogel, H. J., &
4732 Kögel-Knabner, I. (2019). Soil organic carbon storage as a key function of soils - A review
4733 of drivers and indicators at various scales. *Geoderma*, *333*, 149–162.
4734 <https://doi.org/10.1016/J.GEODERMA.2018.07.026>

- 4735 Xiang, D., Wang, G., Tian, J., & Li, W. (2023). Global patterns and edaphic-climatic controls
4736 of soil carbon decomposition kinetics predicted from incubation experiments. *Nature*
4737 *Communications*, 14(1), 1–14. <https://doi.org/10.1038/s41467-023-37900-3>
- 4738 Xu, S., Wang, J., Sayer, E. J., Lam, S. K., & Lai, D. Y. F. (2024). Precipitation change affects
4739 forest soil carbon inputs and pools: A global meta-analysis. *Science of the Total*
4740 *Environment*, 908, 168171. <https://doi.org/10.1016/j.scitotenv.2023.168171>
- 4741 Xue, G., Cai, Y., Cheng, P., Lechleitner, F. A., Zhang, H., Zheng, Y., Wei, Y., Huang, S., Yang,
4742 L., Cheng, X., Lu, Y., Zhou, J., Ma, L., Cheng, H., & Edwards, R. L. (2023). The Climate
4743 Control of Soil Organic Carbon Dynamics Inferred From Speleothem Radiocarbon Ages.
4744 *Geophysical Research Letters*, 50(3), e2022GL101875.
4745 <https://doi.org/10.1029/2022GL101875>
- 4746 Yang, H., Liu, S., Wang, J., Schuler, J., Wang, Y., & Luan, J. (2023). Nitrogen-Fixing Plants
4747 Enhance and Stabilize Rhizospheric Soil Organic Carbon in Tropical Rainforests, Hainan
4748 Island, China. *Forest Science*, 69(6), 653–664. <https://doi.org/10.1093/forsci/foxad037>
- 4749 Yildiz, K., Karakaya, N., Kilic, S., & Evrendilek, F. (2020). Interaction effects of the main
4750 drivers of global climate change on spatiotemporal dynamics of high altitude ecosystem
4751 behaviors: process-based modeling. *Environmental Monitoring and Assessment*, 192(7),
4752 1–14. <https://doi.org/10.1007/s10661-020-08430-y>
- 4753 You, M., Liu, X., Bäumler, R., & Oliver, M. A. (2023). Grassland soils in the cool–arid–
4754 temperate ecozone (Steppe, grassland). *Encyclopedia of Soils in the Environment, Second*
4755 *Edition*, V4-299-V4-318. <https://doi.org/10.1016/B978-0-12-822974-3.00227-5>
- 4756 Zhao, J., Feng, X., Hu, J., He, M., Wang, S., Yang, Y., & Chen, L. (2024). Mineral and Microbial
4757 Properties Drive the Formation of Mineral-Associated Organic Matter and Its Response to
4758 Increased Temperature. *Global Change Biology*, 30(12), e70004.
4759 <https://doi.org/10.1111/gcb.70004>

4760 Zhou, X., Zhou, L., Nie, Y., Fu, Y., Du, Z., Shao, J., Zheng, Z., & Wang, X. (2016). Similar
4761 responses of soil carbon storage to drought and irrigation in terrestrial ecosystems but with
4762 contrasting mechanisms: A meta-analysis. *Agriculture, Ecosystems and Environment*, 228,
4763 70–81. <https://doi.org/10.1016/j.agee.2016.04.030>

4764 **Conclusão geral**

4765 Os resultados deste estudo avançam na compreensão da dinâmica subterrânea do
4766 carbono do solo, integrando fungos, raízes, respiração do solo e padrões espaciais do estoque
4767 de carbono sob uma perspectiva funcional e pedoecológica. A análise global da biodiversidade
4768 fúngica revelou que a distribuição desbalanceada de dados, aliada à alta variabilidade funcional
4769 dos microrganismos, compromete a acurácia dos modelos em ecossistemas tropicais e áridos,
4770 justamente onde a riqueza fúngica (nos trópicos) ou a especialização das comunidades (em
4771 regiões secas) são mais críticas. Essa lacuna reforça a necessidade de ampliar amostragens e
4772 incorporar explicitamente a diversidade microbiana em estratégias de conservação.

4773 A investigação dos traços funcionais radiculares destacou o papel central das raízes não
4774 apenas na aquisição de recursos, mas na estruturação da rizosfera e na modulação da ciclagem
4775 de nutrientes e carbono, mediadas por interações complexas entre fatores bióticos e abióticos.
4776 Paralelamente, a análise da respiração do solo (R_s e R_h) confirmou que a textura do solo,
4777 particularmente o teor de argila, é determinante na resposta ao aquecimento global: solos
4778 argilosos protegem o carbono orgânico (SOC) e reduzem a sensibilidade climática, enquanto
4779 solos arenosos amplificam os impactos, exigindo a integração de características pedológicas
4780 em projeções globais especialmente futuras para evitar generalizações frágeis.

4781 A modelagem dos estoques de SOC, ao incorporar inteligência artificial explicável e
4782 variáveis edafobiológicas, aponta novos controladores funcionais, como a taxa de colonização
4783 micorrízica (RMC), e hierarquias complexas entre clima, solo e biota. Esses avanços
4784 metodológicos evidenciaram interações não lineares e padrões condicionais, como a influência
4785 combinada de textura, regime hídrico e simbioses, destacando que condições favoráveis ao
4786 acúmulo de SOC em uma região podem ter efeitos opostos em outras.

4787 Contudo, as limitações persistem: a carência de bancos de dados robustos, harmonizados
4788 e globalmente representativos, especialmente em regiões críticas como trópicos e áreas áridas,

4789 restringe o potencial preditivo dos modelos. Cenários climáticos extremos (SSP5-8.5)
4790 amplificam esses desafios, intensificando riscos à dinâmica do carbono em altas latitudes e
4791 solos menos resilientes.

4792 Para enfrentar essas incertezas, são fundamentais iniciativas globais de padronização de
4793 dados, aliadas à priorização de práticas agrícolas sustentáveis e políticas de conservação que
4794 preservem a biodiversidade do solo. Avançar nessa agenda implicará: (i) ampliar redes de
4795 amostragem em ecossistemas negligenciados, como trópicos e zonas áridas; (ii) padronizar
4796 protocolos para a coleta de traços funcionais subterrâneos e medições de respiração do solo;
4797 (iii) integrar variáveis biológicas (e.g., diversidade fúngica, colonização micorrízica) em
4798 modelagem de carbono; e (iv) adotar modelos transparentes que esclareçam mecanismos
4799 ecológicos, não apenas gerem previsões numéricas. Assim a ciência do solo poderá fornecer
4800 estimativas confiáveis e orientações práticas à altura do desafio climático que se impõe.

4801 Em síntese, os capítulos convergem para uma visão integrada dos ecossistemas edáficos,
4802 onde a sinergia entre inteligência artificial explicável, conhecimento pedológico tradicional e
4803 dados empíricos robustos emerge como caminho indispensável. Essa abordagem
4804 multidisciplinar transcende a mera previsão numérica, priorizando a compreensão de processos
4805 ecológicos e a heterogeneidade edafobiológica. Ao unir transparência metodológica,
4806 fundamentação científica e aplicabilidade, este estudo oferece um arcabouço para orientar
4807 estratégias globais de manejo sustentável do solo, alinhando conservação da biodiversidade,
4808 segurança climática e resiliência dos ecossistemas terrestres.