

Reviewer 1:

This study leveraged a rich collection of global soil organic carbon (SOC) data to create an updated, high-resolution map of the spatial distribution of SOC storage at the top 30 cm and 100 cm. The motivation for developing a spatially resolved, high-resolution soil C map for better land management, C accounting, and global C monitoring is strong. The resulting map can be valuable to the urgent need of accurate monitoring and estimation of soil carbon removal impacts through climate solutions around agricultural and natural ecosystems. Overall, this work is a good contribution to the soil community, but I have a few concerns about the delivery of results and the writing. Therefore, I recommend this manuscript to be published on ESSD if a major revision can fully address the main concerns.

First, the good things:

I appreciate that the authors harmonized and cleaned up a massive collection of layered soil data rigorously. For example, they used only samples with bulk density measurements, adjusted measurement values according to sampling protocols, and filtered out redundant points. The authors dealt with SOC stock calculations and layered data at different depths responsibly.

The authors also fitted separate models for peatland and mangroves, which seems reasonable. This approach implies an assumption that relationships between SOC storage and predictor layers are different for these two ecosystems compared to the rest of global ecosystems.

Author Response:

Thank you for reviewing our manuscript and sharing your feedback. We hope our revisions address your questions and concerns.

Major concern 1-1:

My first major concern is that the entire article showed lots of summarized total SOC storage at biome, regional, and global level, but none of these estimates come with an uncertainty range or confidence interval in the main text or the supplemental material. Although pixel level uncertainties were shown as main figures, the authors should still report uncertainties on the estimated global and regional SOC stocks in the abstract and throughout the result sections. This should be easy given the authors already fitted 20 models from bootstrapped training data. The CIs (and bootstrap distributions) can be derived once the regional and global

estimates were calculated from each of the 20 models. The lack of regional uncertainties makes some of the conclusions unconvincing.

Author Response:

Thank you for your comment regarding the need to report uncertainties for global and regional SOC stock estimates. We calculated the mean and total SOC stock for each region from our 20 bootstrapped model outputs, and computed corresponding uncertainty measures, including model variance and residual variance, following established zonal inference methods (Xu et al., 2017; McRoberts et al., 2019; 2022). These regional uncertainties are reported in the revised manuscript.

We also added the following section to the methodology:

2.10 Uncertainty assessment

“We assessed uncertainty in our SOC estimates by quantifying errors at both pixel and regional scales, incorporating uncertainties from model development and prediction. At the pixel level, uncertainty reflects both model residuals defined as the differences between observed and predicted values and parameter-estimation variability due to the finite size of the training dataset. To capture these uncertainties and their spatial variability, we implemented a bootstrap resampling framework: in each iteration, the SOC dataset was randomly split into 70% training and 30% testing sets, and 20 independent global Random Forest models (300 trees) were trained to generate separate SOC maps. Pixel-level uncertainty was quantified as the standard deviation across the 20 model outputs, and 95% confidence intervals were computed. This approach captures the inherent variability of the global SOC dataset, enabling evaluation of heteroscedasticity in model performance. For both 30 cm and 100 cm depths, we produced two uncertainty maps: pixel-based standard deviation and 95% confidence-interval maps (Fig. 4 & Fig. S4). At the regional scale, total uncertainty was derived from the 20 bootstrapped SOC prediction maps for each depth by summing model variance and residual variance following zonal inference approaches developed in previous studies (Xu et al., 2017; McRoberts et al., 2019, 2022).”

References:

Xu, L., Saatchi, S. S., Shapiro, A., Meyer, V., Ferraz, A., Yang, Y., Bastin, J.-F., Banks, N., Boeckx, P., Verbeeck, H., Lewis, S. L., Muanza, E. T., Bongwele, E., Kayembe, F., Mbenza, D., Kalau, L., Mukendi, F., Ilunga, F., and Ebuta, D.: Spatial Distribution of Carbon Stored in Forests of the Democratic Republic of Congo, *Scientific Reports*, 7, 15030, <https://doi.org/10.1038/s41598-017-15050-z>, 2017.

McRoberts, R. E., Næsset, E., Saatchi, S., and Quegan, S.: Statistically rigorous, model-based inferences from maps, *Remote Sensing of Environment*, 279, 113028, <https://doi.org/10.1016/j.rse.2022.113028>, 2022.

McRoberts, R. E., Næsset, E., Liknes, G. C., Chen, Q., Walters, B. F., Saatchi, S., and Herold, M.: Using a Finer Resolution Biomass Map to Assess the Accuracy of a Regional, Map-Based Estimate of Forest Biomass, *Surveys in Geophysics*, 40, 1001–1015, <https://doi.org/10.1007/s10712-019-09507-1>, 2019.

Major concern 1-2:

For example, in section “3.2.1 Global fires”, the authors concluded that adding fire as a predictor layer increased global top 30cm SOC but decreased top 1m SOC. However, the difference for SOC at 1m is only 1% of the total estimated SOC (28 out of 2822, line 552-553). I don’t know whether the CIs of the estimated 2850 PgC and 2822 PgC overlap with each other. If so, this evidence can’t convince me that “fire can be a significant source of carbon emission”.

Author Response: We thank the reviewer for this comment. Following your suggestion, we have revised the discussion of fire-related effects in the manuscript. Specifically, we have made the following updates:

1. We have reorganized **Section 4.2 Land-use and disturbance** to highlight three main aspects: (i) how fire overlaps with SOC distribution, (ii) the influence of the fire feature on SOC predictions based on partial dependence analysis, and (iii) differences in SOC response across biomes.
2. We no longer suggest that fire is a major source of global carbon emissions. Instead, we emphasize how fire overlaps SOC stocks regionally, with peatlands showing the largest fire-related effects, and PDP results indicating changes of 37 t C/ha at 30 cm and –58 t C/ha at 100 cm depth.
3. Total uncertainty combining model variance and residual variance has been computed and added to the manuscript.

Major concern 1-3:

Another concern about this paragraph (line 551-555) is that, although adding fire as a predictor seems ecologically reasonable, I’d still expect to see the authors show that the model performance increased after adding fire as a predictor, in order to make that final conclusion. Without proving fire as a statistically justifiable predictor, this result might point to some artefacts rather than an improvement in the model due to fire data. Another way to look into fire’s effect is to predict the global SOC stock from the fitted model (with the fire layer), in a scenario where no fire happens on Earth, and then compare the no-fire prediction to global SOC stock estimate with real fire frequencies (Perhaps this is what you did. Am I misunderstanding?).

Author Response:

We thank the reviewer for this comment. To address this concern, we have revised the **Methods section 2.2.** to clarify the role of the fire layer. The updated manuscript now reads:

“For each biome model and soil depth, we compared the model performance (R^2) on the full dataset with and without the fire variable. Including fire as a predictor had a negligible to very small effect of ~ 0.01 on model performance across biomes and soil depths. The minimal change in R^2 suggests that adding fire does not introduce artifacts or model instability. Thus, we retain fire as a predictor because it captures an important ecological process influencing soil carbon dynamics.”

Regarding the second point, we no longer compare global stock predictions with and without fire as an input layer to the model. Instead, we focus on the results of the fire feature effects on SOC predictions per biome, and find large fire effects in tundra/boreal regions and in peatlands, consistent with known ecological patterns.

Major concern 2:

It seems like land cover is not an explicit predictor in the model. It gives me two questions.

(1) It is hard for me to understand that without this predictor, how does the model discern different SOC contents between a forest and a nearby cropland that shares essentially the same environmental condition? Perhaps you expect the Landsat bands to give such information implicitly?

Author Response:

We appreciate the reviewer's comment. The Landsat 8 bands we used (Red, NIR, SWIR1, and SWIR2) capture spectral information related to vegetation and land cover, providing an implicit signal that distinguishes forests, croplands, and other land cover types. Even without an explicit land cover layer, these bands allow the model to differentiate areas with similar environmental conditions (e.g., climate, elevation) but differing in vegetation type or management.

(2) The peatland and mangrove models were fitted by points within their remote sensing extents. Is it possible to find ground-truth information on whether a soil sampling belongs to peatland or mangroves? Because small-scale farming can make patchy drained lands within such systems (as you mentioned), it is likely that some points within the peatland/mangrove

extent weren't in fact peatland or mangrove, raising potential bias to the biome-specific models. It would be great if the authors could validate the peatland/mangrove extent with their ground truth land cover—if you have some site-level information from the data source, what proportion of peatland/mangrove sites were and weren't included in the peatland and mangrove masks? And how many sites that weren't peatland/mangrove end up within the masks?

Author Response: Thank you. We have added further clarification in **Section 2.5** of the Methods:

“To minimize the inclusion of non-mangrove or non-peatland sites, we relied on the most accurate and spatially explicit global datasets available: the Global Mangrove Watch v3.0 (2020) (Bunting et al., 2022) and the Global Peatland Assessment Database v2 (GPM2.0, 2022) from UNEP, both of which account for small-scale land use changes such as local farming and deforestation.”

“While not all SOC ground-truth points had land cover metadata, 58% of points within the mangrove extent at 30 cm (71% at 100 cm) were explicitly identified as mangrove, and 14% of points within the peatland extent at 30 cm (35% at 100 cm) were identified as peatland.”

Although some uncertainty remains in the land cover data, using high-quality remote sensing masks for mangroves and peatlands along with the best available ground-based data provides a reliable basis for biome-specific model training.

Major concern 3:

The soil data were sampled at different times spanning 1950s to 2020s, but the goal of this work is to represent a static map of global SOC stock distribution. It is common and maybe justifiable to use all historical data for mapping a static global SOC stock distribution, but I wish the authors could extend more discussion on how they expect readers to use this data product. The authors motivated this work with the recent need for soil carbon removal and land management, but how will this map of static SOC stock estimation help with these motivations, building upon the previous global SOC maps? I was excited to read an insightful reasoning of how this work is needed in the context of climate solution, policy making, and land management (which was pitched as the motivation), but I find it lacking in both the introduction and discussion. I hope the authors can elaborate more explicitly and clearly on the real impact and value of this data product, acknowledging that land use management and carbon removal programs would deal with SOC change over a short period of time (5-10 years). Some relevant literature on the perspective of natural climate solutions and carbon

removal include <https://www.nature.com/articles/s41558-024-02182-0> and <https://www.nature.com/articles/s41467-023-44425-2>.

Author Response:

Thank you. We appreciate the reviewer for raising this important point. To address it, we have added the following sections to the manuscript.

4.3.1 Baseline map for research and monitoring

“We present a high-resolution (100 m) soil organic carbon map at two depths, representing baseline SOC stocks circa 2022. Soil profile data collected from the 1950s to the 2020s were combined with contemporary remote sensing covariates, using Landsat 8 bands to capture landscape conditions around 2022. The use of long-term soil datasets aligns with previous global mapping efforts, such as SoilGrids v1-v2 and GSOCmap v1.5, which integrate multi-decadal data to capture large-scale spatial patterns. The primary goal of this work was to represent SOC variation across diverse landscape gradients with a resolution fine enough to capture local differences while maintaining global coverage. The resulting map can support applications in land management, nature-based climate solutions, and future assessments of SOC dynamics.

SOC data inevitably span decades, while Earth observation covariates reflect more recent conditions. In landscapes without major human disturbances, SOC is assumed to remain relatively stable over time, so historical soil measurements can reliably represent baseline stocks when combined with current remote sensing data. In areas affected by land-use change, degradation, or recovery, such as forests regrowing after wildfire, agricultural rotations, or restored wetlands, SOC may have changed since sampling. For instance, samples from western CONUS encompass forests at different successional stages recovering from events such as wildfires or storms. Remote sensing captures spectral signals related to land cover and vegetation recovery, helping to contextualize historical soil measurements. Consequently, the SOC map reflects not only static soil conditions but also the imprints of temporal changes across landscapes. To address potential temporal discrepancies at the pixel level, we generated an uncertainty map alongside the SOC estimates, allowing users to assess confidence in areas where SOC may have changed substantially since sampling.

Our dataset integrates both natural and human-modified landscapes, including agricultural areas, managed forests, wetlands, and coastal ecosystems such as mangroves. By combining long-term soil measurements with remote sensing covariates and machine learning models, the map captures both spatial variation in SOC and the influence of historical land-use changes and recovery processes. Overall, the high-resolution SOC maps represent a valuable tool for carbon management and conservation. The maps enable fine-scale detection of spatial variability, supports assessments of land-use impacts, and provides a

foundation for monitoring SOC dynamics over time, even in regions with limited historical soil data.”

4.3.2 Policy and management applications

“The baseline SOC map serves as a benchmark with broad relevance across scientific, management, and policy applications. For carbon stock reporting, it enables estimation of SOC across regions, land cover types, and management units, with particularly high precision for areas larger than 1 hectare where pixel aggregation reduces uncertainty. At the policy level, the map provides national and sub-national jurisdictions with a means to assess soil carbon storage and sequestration potential, accounting for spatial variability across land uses and vegetation successional stages. Such information supports the design and implementation of land-use and restoration policies aimed at improving soil carbon conditions. The map also facilitates participation in carbon markets, where project developers require reliable SOC stock data to estimate emissions reductions or removal factors, including avoided losses and potential accumulation rates in disturbed versus natural areas. In addition, the high-resolution benchmark map at multiple depths can be integrated into biogeochemical models to quantify potential changes in soil carbon. Building on this work, we plan to combine repeated SOC measurements with process-based models to advance understanding of SOC dynamics. By leveraging this baseline, carbon management strategies can be refined and climate policies better informed.”

Major concern 4:

There are no plots showing the effect of each covariate on SOC storage in the fitted model. I understand if the work doesn't want to emphasize the fitted relationships in the model (which can be messy to interpret), but I think it should be reported in the supplemental material for transparency. A group of partial dependence plots for each biome-specific model would suffice, so that the readers don't wonder if model predictions are driven by one or two strong but non-physical effects.

Author Response:

Thank you for the suggestion. We conducted additional calculations using partial dependence analysis to evaluate feature influence on SOC predictions. These analyses were stratified by biome to highlight differences in covariate effects.

Methods are detailed in the new **Section 2.7** (“Partial dependence analysis”), and results are presented in **Section 3.5** (“Feature effects on SOC predictions”), where we discuss how the models' response patterns are ecologically interpretable.

Full results are provided in **Supplementary Figure S3**.

Minor comments:

Line 22-24, Abstract. “Our analysis indicates that annual wildfire dynamics and shifts in agricultural land can influence SOC by 132 Pg C and 140 Pg C at 30 cm, and by 345 Pg C and 368 Pg C at 100 cm, representing approximately 13% of the global stocks.”

These numbers seem to be calculated from the total estimated SOC stock in global areas that fall within the agricultural and fire-prone extent (line 619). Indeed, the theoretical maximum potential of SOC loss driven by any disturbances is always bounded by the total SOC stored in a given area, so the sentence is technically correct. However, this sentence reads like “we formally analyzed the effect of recent/near-future fire and agricultural dynamics on SOC change, which show that these activity poses a readily threat to this much SOC stock in the next several decades”. This does not accurately summarize the relevant results (line 619). In the fire-prone pixels defined by “average annual number of burned days >1” (line 296-297), such fire dynamics very likely do not pose an immediate threat to the entire soil C pool in every fire-prone pixel (the author also pointed out multiple times that fire has mixed effects to SOC change yet to be understood). The same point applies to agricultural activities. To avoid misrepresenting the main result of this work, please adjust the phrasing of this sentence. Maybe something like “estimate shows that XXX PgC of soil carbon sits in fire-prone area and/or area with ongoing agricultural activities...”

Author Response:

Thank you for this comment. We have revised the sentence to clarify that the reported values represent SOC stocks located in fire-prone and agricultural areas. The sentence in the manuscript now reads:

“Our estimates indicate 134 ± 2 Pg C and 340 ± 5 Pg C sit in fire prone areas at 30 cm and 100 cm depth, and that 140 ± 2 Pg C and 384 ± 8 Pg C are in areas of ongoing agricultural activity at 30 cm and 100 cm depth, representing about 13% of global SOC stocks.”

Line 505. “Our study suggests considerably greater carbon storage in the Amazon Basin.”

Is it because your study better predicts the large carbon storage in the Amazon peatland (your next paragraph)? If so, explain explicitly here.

Author Response:

Thank you. We have revised this section of the manuscript, which now reads:

“Across the entire Amazon Basin (593 Mha), we estimate 37 ± 1 Pg C at 30 cm and 162 ± 7 Pg C at 100 cm, representing 4 % and 6 % of the global SOC total, respectively. Our results at 100 cm are higher than previous Amazon Basin assessments, including 47 Pg C at 100 cm (Moraes et al., 1995), 46.5 Pg C (Batjes and Dijkshoorn, 1999), and 36.1 Pg C

(Gomes et al., 2019), due in part to our use of the recently updated peatland extent dataset that better captures carbon-rich areas.”

Line 551-555.

See major concern 1.

Author Response: We have addressed this section in our response to Major Concern 1.

Line 685 “Our data indicates that 35% of the Cerrado is used for agriculture (72/204 Mha)”

And **Line 689** “Our map highlights high fire activity in the Matopiba region”

If I understand correctly, the agricultural land extent is cited from another product, and the fire activity extent is indicated by the MODIS fire frequency map. How is that “your map” and your data? (I would’ve assumed these terms to refer to your SOC data product in this manuscript.) It kind of confuses and distracts readers from focusing on what’s truly your valuable map and your data product —the SOC stock maps at two depths.

Author Response:

Thank you for this comment. The agricultural land extent and fire activity are derived from external datasets (LGRIP 2015 and MODIS 2000–2023, respectively) rather than generated directly from our SOC mapping effort.

To clarify, we have revised the text to explicitly attribute these datasets to their original sources. The manuscript now reads:

“In Brazil’s Cerrado, our model estimates that soils store 8.7 ± 0.2 Pg C at 30 cm and 28.1 ± 10.8 Pg C at 100 cm across 204 Mha. Although the Cerrado contains moderate SOC stocks compared to other biomes, 38% of the region is classified as fire-prone and 35% as agricultural land, with a 63% spatial overlap with either category, based on LGRIP 2015 and MODIS 2000-2023 datasets (Figure 8). Fire-prone areas contain an estimated 3.3 ± 0.1 Pg C (30 cm) and 10.0 ± 0.6 Pg C (100 cm) across 77 Mha, and agricultural areas contain 3.1 ± 0.1 Pg C (30 cm) and 9.4 ± 0.6 Pg C (100 cm) across 72 Mha. Fire-prone areas and agricultural land show larger SOC contrasts at 100 cm than at 30 cm compared to other areas, which may reflect differential SOC dynamics with depth. High fire activity in regions such as Matopiba could be associated with rapid agricultural expansion. These patterns suggest a combined role of fire and land use in shaping SOC dynamics (LGRIP 2015; MODIS, 2000–2023).”

Line 699-703

The authors have already discussed the mixed and unknown impact of fire on soil in the global fire section and the grassland section (lines 556-565, lines 424-426). Coming after these sections, these lines read redundant. Perhaps all of these can be consolidated better into the global fire section (the paragraph of lines 556-565)

Author Response: Thank you. We have reworked the manuscript outline and consolidated the discussion of fire impacts into a single coherent section. The revised outline now reads:

4.2 Land-use and disturbance

4.2.1 Global fires and agriculture (*now includes a consolidated global fire discussion*)

4.2.2 Regional SOC stocks relative to land-use and fire (*focuses on regional distribution of fire and agriculture*)