Reviewer 3:

General comments:

This manuscript mapped global 100-m resolution SOC through compiling 84,880 topsoil, 44,304 subsoil SOC samples and covariates to multi-sources remote sensing and other data products as extra layers for training biome-specific random forest models. This high spatial resolution product is an important resource for future studies on soil carbon management, thus the major outcome from this work is useful and timely needed for soil biogeochemistry and carbon cycle modeling communities. The comprehensive data inputs authors used, random forest based geospatial predictive mapping are popular and robust methods, thus the quality of the produced SOC map is partially justified. However, the writing of the discussion section in this manuscript, justification of bias correction and missing attribution of different uncertainty sources, have not been addressed properly. In addition, I checked the data product and found lots of missing points in the uploaded geotiff file. Overall, this is a solid and interesting work, but I would only recommend it for publication at ESSD after major revisions to address my following concerns.

Author Response:

Thank you for taking the time to review our manuscript and provide detailed feedback. We appreciate your comments. Regarding the missing points in the uploaded GeoTIFF, we checked that the dataset is complete. The file is in COG format with overviews, and some overviews may take slightly longer to load than others.

Specific comments:

1. Histogram-based bias correction can be tricky. Authors claim the better match is achieved after bias-correction, but the correlation of SOC stock to environmental covariates and categorical datasets for training are forced to be changed. Since Soilgrid's product did not show biased results after training, I wonder if any of these covariates and categorical maps added to layers in your training model are the reason? Could you explain the reason causing this bias and justify that the overall uncertainty of the data product is not largely affected by your bias-correction?

Author Response:

In the previous version of the soil carbon map, a global histogram matching bias correction was applied to the combined outputs of the three broad models (global, mangrove, peatland) to reduce systematic over- or underestimation relative to observed values. In the current version, we modeled soil carbon separately for 14 biomes and ecosystems, with each model trained on a more homogeneous subset of data. This biome-specific approach inherently reduces systematic bias, as each model is better able to capture local ecological patterns. Therefore, no additional bias correction was applied, and model outputs are reported directly from the biome-specific models.

This clarification has been added to **Methods Section 2.6**, as follows:

> "The biome-specific approach inherently reduces systematic bias, allowing each model to better capture local patterns. Consequently, no additional bias correction (i.e. histogram-based bias correction) was applied, and the reported values reflect the direct outputs of the biome-specific models."

2. The highlight of this work is the ultra fine 100-m resolution product, which is unique product. But in your manuscript, I cannot find any discussion on how you get a meaningful high resolution data product. By just using soil profiles to evaluate your model, I entrust your model to mapping the non-linear correlation of SOC stock to other covariates. But without high resolution input, your 100m resolution is less persuasive, such as Soilgrids 250m product using some 250m input covariates like MODIS products. In your study, I did not see any very high resolution products for model training. I would suggest explaining what high resolution products you're used to increase the credibility of your high resolution product.

Author Response:

We thank the reviewer for this important comment. We have revised **Methods Section 2.4** to detail our approach. As described in the manuscript:

> "All geospatial input raster layers were reprojected to a standardized spatial reference system and a uniform 100 m spatial resolution prior to model training and inference. The input datasets, described in previous sections, included: Copernicus Global Land Cover (100 m), Landsat 8 bands (30 m), MODIS Land Surface Temperature (1 km), MODIS Evapotranspiration (500 m), ALOS PALSAR (25 m), SoilGrids soil properties (250 m), GPM2.0 global peatland map (1 km), Global Mangrove Watch forest extent (25 m), NCSCDv2 permafrost map (1 km), WRB soil classification (250 m), Congo Basin peatland map (50 m), Peruvian peatland map (50 m), and MarSOC tidal marsh map (10 m). Differences in native spatial resolution among datasets were addressed through interpolation-based resampling. The harmonized 100 m layers were then used as predictor variables in the RF modeling framework to generate the final global 100 m resolution product."

3. Authors described SOC stock results over specific regions/categories, but also wrote lengthy discussion with lots of statements only weakly related to the SOC stock product itself. I feel like reading a review paper and lots of results are just descriptions of categorical or environmental covariates input. I recommend authors to simplify your writings in the results section and focus on discussing SOC stock. I have one example in additional points, but expect authors to double check the whole manuscript.

Author Response:

We appreciate this helpful comment. We consolidated the Results and Discussion to focus on the SOC stock dataset. The manuscript structure was reorganized (Results 3.1–3.6; Discussion 4.1–4.3) to emphasize model performance, spatial patterns, feature effects, and uncertainty, in alignment with the data product and its interpretation. The revised Discussion highlights global patterns of SOC in comparison with existing global maps and briefly addresses land use and disturbance, providing context for map applications in carbon management, monitoring, and research.

Revised Manuscript Outline:

Results: 3.1 Global distribution of soil organic carbon, 3.2 Ground truth data across biomes and depths, 3.3 Model generalization across biomes, 3.4 Final model performance across biomes, 3.5 Feature effects on SOC predictions, 3.6 Global and regional uncertainty.

Discussion: 4.1 Global patterns of SOC, 4.1.1 Comparison with existing global map, 4.1.2 Critical ecosystems and patterns of SOC across biomes, 4.2 Land-use and disturbance, 4.2.1 Global fires and agriculture, 4.2.2 Regional SOC stocks relative to land-use and fire, 4.3 Implications for carbon management and conservation, 4.3.1 Baseline map for research and monitoring, 4.3.2 Policy and management applications.

4. This concern is related to my point #3. I found the discussion on uncertainty is missing in the results section. I appreciate authors separately discussing their results under different categories like fire-prone region, agriculture land and peatland, but a word or two to summarize the uncertainty from your data products over these regions can be valuable evaluation and probably can help you find the source of bias when training your model with raw data?

Author Response:

We appreciate this constructive comment. We have added a separate section (Results 3.6 Global and regional uncertainty) presenting global maps of uncertainty. We also computed regional uncertainties, including model variance and residual variance, using established zonal inference methods (Xu et al., 2017; McRoberts et al., 2019; 2022). These regional uncertainties are reported throughout the revised manuscript, providing a clearer evaluation of potential sources of bias.

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.

5. It is not reasonable to use bbox for calculating averaged or total SOC stock for specific geopolitical regions. Please use maps for masking geopolitical regions. Also, since some of the categorical maps are overlapping each other, for example, agriculture and fire-prone areas, I have an extra suggestion that you can prepare a global map to depict different categories you discussed with different colors, so readers will have a better idea.

Author Response:

Thank you for the suggestions. We have replaced all bounding boxes with GeoJSON masks and recalculated the statistics for geopolitical regions.

To address overlapping categories, we prepared a reprojected and combined figure showing fire-prone and agricultural areas for clarity and improved readability:

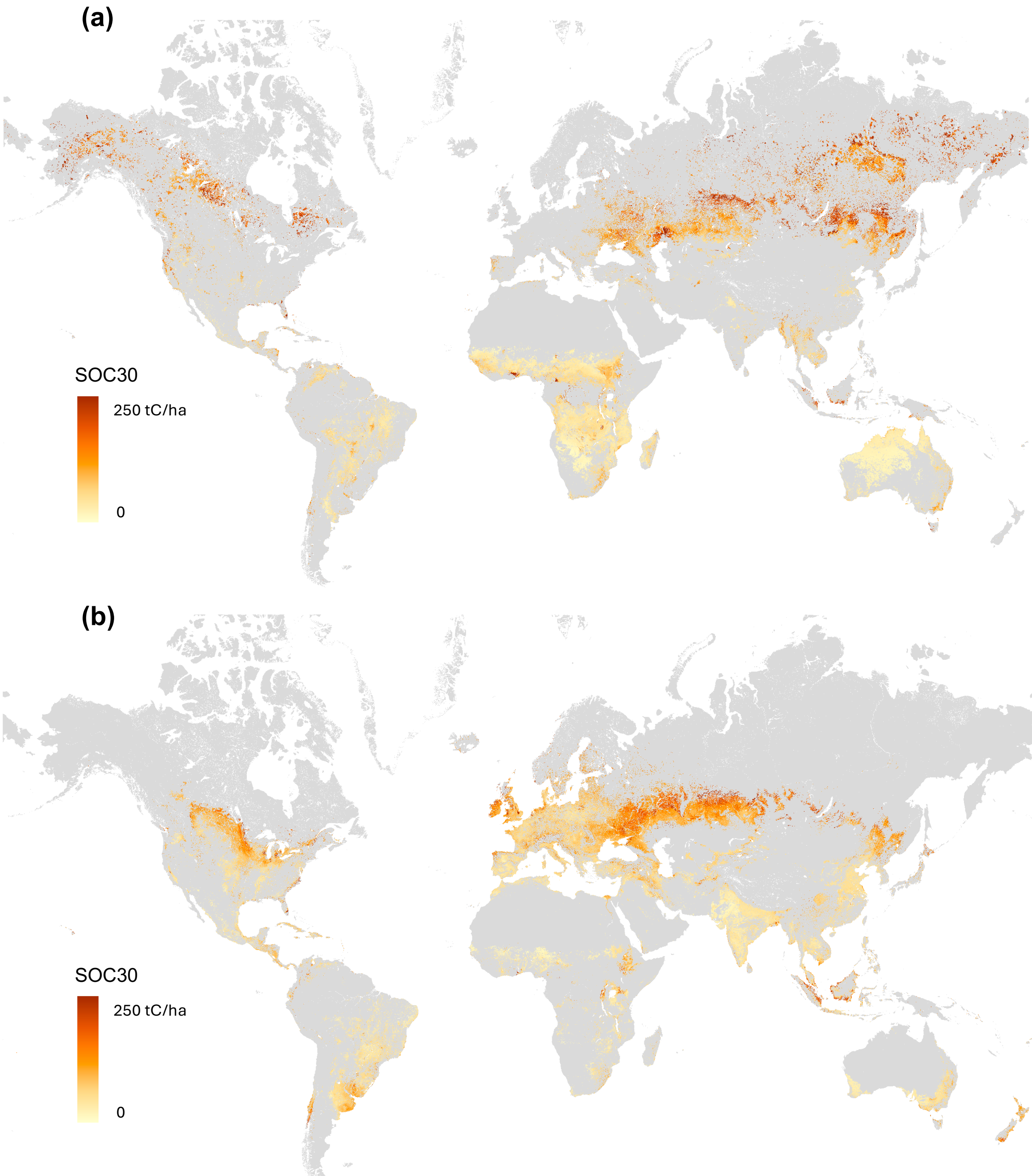

Figure 6. Soil organic carbon stocks (t C/ha) at 0‑30 cm (SOC30) for fire-prone and agricultural regions, showing fire-prone areas (≥1 burned day per year; MODIS/061/MDC64A1, 2000‑2023) (a) and agricultural land extent (LGRIP v001; Teluguntla et al., 2023) (b).

6. The data product contains lots of missing points for no reason. Shall double check your uploaded geotiff files or explain the reason why you have these missing values.

Author Response:

Thank you. We reviewed the dataset and can confirm that all points are present. The apparent gaps may be due to the COG format; some areas may appear empty at certain zoom levels, but all data points are visible when zoomed in.

Additional points:

Line 158: Explain how you bin the samples, by SOC stock?

Author Response:

Thank you for your comment. We have added further clarification in the revised manuscript, which now reads:

> "We applied a randomized subsampling approach to national and regional maps based on SOC values. SOC pixels were stratified into 10 quantile-based bins, from lowest to highest, with each bin containing approximately the same number of data points. A fixed number of samples per bin (10 bins; 50 samples per bin) were then randomly drawn, to get an equal representation across the full SOC distribution."

Line 173: Shall explain: D is the soil layer depth (cm), and equals to 30cm for topsoil and 70cm for subsoil in this study?

Author Response:

Thank you. We have updated section 2.1.4 of the manuscript to clarify this. It now reads:

> "We produced maps for two standard depth intervals commonly used in global SOC mapping: 0-30 cm and 0-100 cm."

Line 331: R square is a metric to show how large a fraction of the variance of a dependent variable explained by the independent variable. Since it is a fraction, the relative magnitude of SOC from dependent and independent variables cancels out each other, and does not affect the R square. Please revise or remove your conclusion.

Author Response:

Thank you. As we have refined the modeling approach by stratifying across distinct biomes and training separate machine learning models for each, this statement has been removed from the manuscript.

Line 355: What are these underrepresented regions?

Author Response:

Thank you. We have updated the manuscript, which now reads:

> "Critical peatland complexes have been largely overlooked in prior global maps, most notably those of the Congo Basin, now recognized as the world's largest tropical peatland (Dargie et al., 2017; UNEP, 2022). The extensive Brazilian and Peruvian Amazon peatlands are also underrepresented in previous datasets despite growing evidence of their significant extent and carbon densities (Draper et al., 2014; Hastie et al., 2022, 2024)."

Line 386: Delete one "and"

Author Response: Done.

Line 423: Which data you used for global fire?

Author Response:

Thank you for the comment. In the Discussion section, we clarified the use of MODIS products for fire frequency. Additionally, we expanded **Methods Section 2.2**. The fire raster used as an input layer was derived from the MODIS/061/MCD64A1 dataset for 2000–2023 and represents the average number of burned days per pixel per year. This raster was then converted into a binary fire mask, with pixels having an average annual burned days ≥1 assigned a value of 1 (fire-prone) and all others assigned 0 (not fire-prone). Zonal statistics were subsequently computed using this binary mask.

Line 442: "Our value at 100 cm is similar to the carbon density of 361 t C/ha found by Sanderman et al. (2018)." But authors mention that total SOC stock to be 45% less from Sanderman's study compared to this work, with similar mangrove extent from both studies. Need to clarify.

Author Response:

Thank you for this helpful comment. We have clarified this section of the manuscript. In our study, we report the following at 100 cm depth:

- SOC stock = 5.07 ± 0.27 Pg C
- SOC density = 346 ± 19 t C/ha
- Mangrove extent = 15 Mha

Sanderman et al. (2018) reported:

- SOC stock = 6.4 Pg C
- SOC density = 361 t C/ ha
- Mangrove extent = 13.8 Mha

We have revised the manuscript to read:

> "Our SOC estimates at 100 cm (5.07 Pg C) are slightly lower than Sanderman et al. (2018) (6.4 Pg C), despite using a larger mapped extent. This difference arises from lower carbon densities in our dataset (346 t C/ha vs. 361 t C/ha) and may also result from differences in spatial resolution, aggregation methods and methodological factors, such as ground-truth datasets and the application of depth functions."

Line 464: I'm more interested in the peatland SOC stock between northern peat and tropical peat, not SOC stock from Northern-Hemisphere, since you discuss these two separated regions later.

Author Response:

Thank you for the comment. The section has been updated accordingly. The revised text now reads in **Section 4.1.2.**:

> "Northern peatlands store the bulk of peatland carbon, with 140 Pg C (85 % of peatland SOC) at 30 cm and 327 Pg C (80 %) at 100 cm across 759 Mha, highlighting the importance of high-latitude ecosystems in carbon storage. Tropical peatlands, despite their smaller extent (134 Mha), contribute 23 Pg C (14 % of peatland SOC) at 30 cm and 81 Pg C (20 %) at 100 cm, reflecting their high carbon density."

Line 470: How did you combine these datasets? By calculating the maximal extent that any dataset shows peatland coverage? Also, what threshold did you use (e.g., 1% of the grid cell)?

Author Response:

Thank you for the comment. We combined the peatland datasets by prioritizing the higher-resolution regional datasets (30 m) for the Congo Basin and Amazon/Peru regions over the coarser-resolution global peatland dataset (GPM2.0). We updated **Section 2.5** of the manuscript, which now reads:

> "To improve regional accuracy, we integrated high-resolution peatland datasets for the Congo Basin and Peruvian Amazon. These higher-resolution regional datasets were prioritized in overlapping areas, such that where both global and regional datasets

were available, the regional data replaced the global data, while areas outside these regions retained the GPM2.0 coverage."

Line 481: Did you calculate the peatland extent or you used a dataset? I got the feeling that you obtained the peatland extent through a dataset but you claimed this as "finding", which means the peatland extent is the output of your trained model? Please correct.

Author Response: The peatland extent is not an output of our work. We have updated all sections to clarify the source of the dataset: Global Peatland Assessment (GPM2.0) (UNEP, 2022).

Line 512: "Mapping peatlands across the Amazon Basin has largely depended on modelling approaches". I'm still curious about how you model the peatland extent. Is it just constrained by several datasets?

Author Response:

Thank you. To clarify, peatland extent is not an output of our model. We use the Global Peatland Assessment (GPM2.0) (UNEP, 2022) to define peatland areas. The sentence in Line 512 refers to previous studies that modelled peatland extent; our work does not model peatland extents but relies on the GPM2.0 dataset. We have clarified this in the revised manuscript.

Figure 5. It may be better to have another subplot to the right showing relative uncertainty (% of uncertainty to the SOC stock)

Author Response: Thank you. We have added regional uncertainties throughout the revised manuscript, as well as in **Figure 1** and **Table 3**.

Figure 6. What is the mask map you used for fire-prone land?

Author Response:

Thank you for the comment. We have added a reference to the MODIS product used to identify fire-prone land in the caption of **Figure 6**. The fire mask was derived from the MODIS/061/MCD64A1 dataset (2000–2023), where pixels with an average annual burned days ≥1 were classified as fire-prone and all others as not fire-prone.

Line 537: "This represents 13% and 12% of the global SOC total". This is interesting to show a large difference from Pellegrini's work on Nature Geosciences. My understanding is still that your fire-prone area mask is much smaller compared to the previous one. Please explain.

Author Response:

Thank you. In consolidating the discussion section of the revised manuscript, this section has been removed. We also acknowledge that the estimated SOC stocks largely depend on the definition of fire-prone areas, which varies between studies. The revised **Section 4.2.1** of the manuscript now reads:

> "Fire-prone areas were mapped from the MODIS/061/MDC64A1 dataset (2000-2023), with pixels averaging ≥1 burned day per year classified as fire-prone. [...] From 2000 to 2023, fires affected an average of ~2,107 Mha of land per year, encompassing an estimated 134 $\pm$ 2 Pg C at 30 cm and 340 $\pm$ 5 Pg C at 100 cm, equivalent to 13% and 12% of global SOC, respectively."

Line 562: I have a feeling that lots of discussion in this manuscript is not closely related to the data product itself. For example, "In tropical forests, fires can generate feedback loops as they alter forest understory fuels, making forests more vulnerable to further fire degradation (Dwomoh & Wimberly, 2017; Wimberly, 2024). The impact of fire on long-term carbon stocks remains under investigation." stated the importance of understanding how aboveground carbon stock and residence time respond to wildfire, which is not closely relevant to your fine resolution SOC stock dataset. I would suggest either shortening or removing unnecessary discussions. Also see my major concern #3.

Author Response: Thank you for this note. As outlined in our response to Comment #3, we have revised the Results and Discussion to maintain focus on the SOC stock dataset.

Line 615: "We find that 17% of fire-prone land (351/2,109 Mha) is located in agricultural zones.". Here I found evidence that you have overlaid analysis from different categories. This may cause potential confusions so I would suggest clearly defining the boundary of each category in a map. See my major concern #5.

Author Response:

Thank you. We have provided further clarification in the manuscript and added the combined **Figure 6**. The line now reads:

> "Based on MODIS fire data (2000-2023) and LGRIP v001 (circa 2015), fires on agricultural land account for 17% of global fire-prone areas (351/2,107 Mha), containing 26 ± 1 Pg C at 30 cm and 72 ± 2 Pg C at 100 cm."

Line 631: "bbox: -25.136719, -34.597042, 55.722656, 38.822591 degrees" Does this mean your calculated statistics for Africa based on this bbox extent? Or you accounted for the whole geographical Africa? Since Africa is a geopolitical region with a defined extent, you shall use a certain regional map as a mask.

Author Response:

Thank you. We have recalculated the statistics using a continent-level GeoJSON mask rather than the bounding box.

Line 667: "Our findings further confirm that agricultural activities place substantial pressure on the region's natural peatlands." I doubt a one time snapshot of SOC product itself can show how agricultural activities place substantial pressure on peatland. Here is another example that discussions are not closely related to your data product and not enough reasoning in your statement. Either simplify/remove or present your complete logical reasoning on how you "confirmed" your conclusions.

Author Response:
Thank you. We have corrected our statement. The manuscript now reads:

> "Agricultural land (derived from LGRIP 2015) occupies 16 Mha, and fire-prone land (derived from MODIS, 2000-2023) covers 8 Mha, with a 5 Mha overlap, suggesting that 41% of peatlands coincide spatially with agricultural or fire-prone zones."

Line 689: "Our map highlights high fire activity in the Matopiba region, likely linked to land clearings." Another example of weak relevance to your SOC stock data product. I doubt you can find evidence of land clearings from your SOC stock data product. Please revise.

Author Response:
Thank you for the comment. The revised section 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 703: "Our data, accounting for spatial variation, suggests that fire impacts on soil carbon in the Cerrado may be mostly noticeable in deeper soils." Authors should also add "fire and agriculture impacts" here.

Author Response: We have updated the sentence to include both fire and agriculture impacts. The revised text now reads:

> "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."

Fig S3. Missing colorbar.

Author Response:

Thank you. The figure has been updated and is now **Figure 2**, which includes a colorbar and shows the performance of the revised biome-specific models.

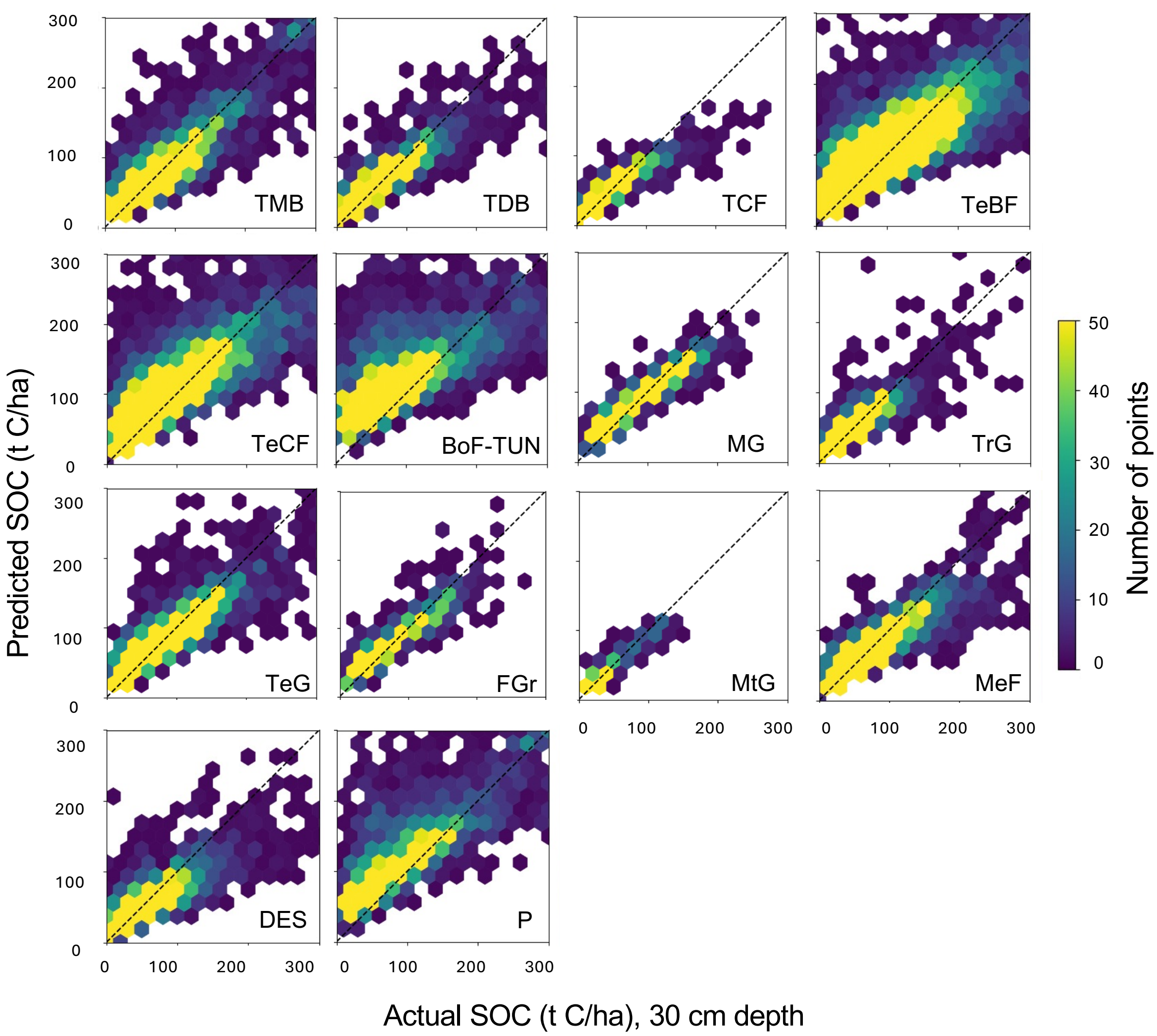

Figure 2. Predicted vs. observed SOC density (t C/ha; full dataset) of 14 biome- and ecosystem-models at 30 cm: (TMB) Tropical and subtropical moist broadleaf forest; (TDB) Tropical and subtropical dry broadleaf forests; (TCF) Tropical and subtropical coniferous forests; (TeBF) Temperate broadleaf and mixed forests; (TeCF) Temperate coniferous forest; (BoF-TUN) Boreal forests/taiga and tundra; (MG) Mangroves (Global Mangrove Watch, 2020); (TrG) Tropical and subtropical grasslands, savannas, and shrublands; (TeG) Temperate grasslands, savannas, and shrublands; (FGr) Flooded grasslands and savannas; (MtG) Montane grasslands and shrublands; (MeF) Mediterranean forests, woodlands, and scrub; (DES) Deserts and xeric shrublands; (P) Peatlands (UNEP, 2022). Extreme values outside axis limits are omitted for comparability.