

Global hotspots of particulate organic carbon losses under climate change

Corresponding Author: Professor Ji Chen

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

This article explores the crucial role of soil organic carbon (SOC) in the global carbon cycle and its uncertain response to climate change, with a focus on particulate organic carbon (POC) and mineral-associated organic carbon (MAOC). Based on extensive data collection and machine learning methods, it aims to reveal the global distribution, key drivers, and future trends of POC and MAOC under different climate scenarios. While the study is conceptually innovative and methodologically sound, offering new insights into SOC dynamics, it has significant issues affecting the accuracy and reliability of the results. Below are the major and minor questions.

Major questions

1. Only 2% of observations are from polar regions, which may lead to an underestimate of SOC losses there. Polar regions are crucial in the global carbon cycle, and insufficient data makes it hard to ensure the global applicability of the conclusions.
2. Random forest models, though good at handling large datasets and nonlinear relationships, cannot directly explain the complex biogeochemical mechanisms of SOC dynamics. A deeper understanding of these mechanisms is needed for more accurate predictions.
3. POC and MAOC interact in complex ways, such as during microbial decomposition. The article does not adequately explore these interactions and their impact on overall SOC dynamics, which is essential for a comprehensive understanding of SOC cycling and climate responses.

Detailed questions

1. The data comes from published studies and unpublished sources, which may vary in sampling and measurement methods. How is the consistency and comparability of the data ensured? More details are needed on data standardization and quality control.
2. Land cover types are critical for the results. How were they determined for each observation point? The accuracy and reliability of this classification method, and whether it accounts for land cover changes affecting SOC, need to be addressed.
3. Environmental variables like MAT and MAP come from different databases. Do their temporal and spatial scales match the SOC observation data? Mismatches may introduce errors, which should be discussed.
4. With large datasets and many variables, random forest models may overfit. Overfitting could affect the reliability of future predictions.
5. The article uses three SSPs to predict SOC changes but lacks in-depth sensitivity analysis. Different SSPs imply varying social, economic, and policy factors that could differently impact SOC dynamics. The uncertainty of these scenarios on results needs more exploration.
6. The definitions of POC and MAOC are based on particle size, density, or both, which may be subjective. How do different separation methods affect the results? This should be discussed.
7. The study standardizes soil depth data to 0-30 cm. Is this method suitable for all soil types and environmental conditions?
8. Soil microorganisms play a key role in SOC decomposition and formation, especially for POC. Were changes in microbial communities and functions considered? Their omission could be a significant oversight.

Reviewer #2

(Remarks to the Author)

Sun, Chen, and colleagues compiled a valuable dataset on soil organic carbon (SOC) fractions in various ecosystems. They drew attention to the vulnerability of particulate organic carbon (POC) in high-latitude regions, particularly in tundra and boreal forests. While this focus is timely and relevant to advancing the understanding of the terrestrial carbon cycle, the manuscript currently lacks sufficient empirical evidence to robustly support these claims. Significant additional analyses are necessary to substantiate the conclusions. Currently, the manuscript reads more like a perspective or hypothesis-driven commentary than a research article grounded in rigorous data analysis.

1. The study uses machine learning models that are trained and optimized with present-day spatial relationships between soil organic carbon (SOC), including the particulate (POC) and mineral-associated (MAOC) fractions, and environmental predictors. These models are then applied to predict changes in SOC under future scenarios. However, it is widely recognized that the spatial relationships between SOC and environmental drivers are likely to change under a changing climate or land use. The current modeling approach does not explicitly consider potential shifts in these relationships over time, raising concerns about the validity of future projections.
2. The manuscript places considerable emphasis on high-latitude ecosystems, particularly tundra and boreal forests. However, data from high-latitude regions account for only 2% of the total dataset. Specifically, observations from tundra and boreal forests comprise approximately 8% of the total dataset. In contrast, over 90% of the data used for model training and prediction are derived from low-latitude regions and ecosystems, such as forests, grasslands, and croplands. Given this imbalance, the conclusions regarding soil organic carbon (SOC) and its fractions in tundra and boreal systems may not be well supported by the underlying data. To strengthen the reliability of these region-specific interpretations, I recommend training and validating separate models for tundra and boreal forest ecosystems exclusively. Then, the performance and variable importance rankings from these targeted models should be compared to those obtained from models trained on the full dataset. This comparison would help determine whether general patterns derived from global data are applicable to high-latitude systems or if distinct mechanisms are at play in these vulnerable regions.
3. Soil organic carbon (SOC) temperature sensitivity is influenced by the intrinsic kinetics of carbon compounds and by factors such as microbial accessibility, physical protection within soil aggregates, and substrate limitation. For example, free particulate organic carbon (POC) tends to be more sensitive to temperature, whereas POC encapsulated within soil aggregates is substantially more protected and thus less responsive to warming. Given these considerations, the assumption that warming will universally accelerate POC decomposition in high-latitude ecosystems due to POC's high temperature sensitivity warrants further scrutiny (see Introduction, Lines 271–274). This hypothesis oversimplifies the complex interactions that regulate SOC stability under changing thermal regimes and may require additional empirical support or mechanistic justification.
4. Numerous studies have investigated the environmental and biological drivers of particulate organic carbon (POC) in soils across different ecosystems and spatial scales. To provide appropriate context, I recommend that the authors summarize and cite relevant literature examining key determinants of POC dynamics. These determinants include factors such as vegetation type, climate variables, soil texture, and land use history. Given this existing body of research, the manuscript should clearly justify the necessity and novelty of identifying the key drivers of POC in the present study (see lines 282–283). A stronger rationale for this analysis will clarify how the study advances the current understanding of this topic.
5. The manuscript reports the selection of 16 predictors for modeling and analysis but does not adequately explain the basis for this selection. Important variables, such as bulk density and elevation, which have been shown to significantly influence soil organic carbon (SOC) and its fractions, appear to have been omitted. The authors should clarify the criteria used to select the predictor variables and explain why certain potentially important factors were excluded. A transparent explanation would enhance the interpretability and reproducibility of the analysis.
6. The manuscript repeatedly emphasizes that POC is more temperature-sensitive than MAOC, especially in tundra and boreal forest soils. However, Supplementary Figure 8 is the only piece of evidence supporting this claim. Since the differential temperature sensitivity between POC and MAOC is central to the study's conclusions, it is essential to provide a more robust and comprehensive analysis (e.g., Lines 336–341, 428–430, 455–457, and 469–474). I strongly recommend incorporating multi-model comparisons to assess the robustness of the temperature sensitivity estimates, as well as clearly describing these procedures in the Methods section. Additionally, a separate analysis focusing on high-latitude regions is necessary to substantiate the claim that POC is more temperature-sensitive than MAOC in these ecosystems. Without these additional analyses, the conclusions may be overstated or insufficiently supported by the available data.
7. Although the authors introduce the Biogeochemistry-Informed Neural Network (BINN) model as supplementary analysis, it still does not resolve the fundamental inconsistency between the scope of model training and the focus of interpretation. Specifically, the BINN model is trained on a global dataset, yet the discussion focuses primarily on changes in soil organic carbon (SOC) fractions in specific regions (e.g., high-latitude ecosystems). This discrepancy raises concerns about the validity of region-specific conclusions derived from a globally trained model. To address this issue, regionally stratified modeling or the inclusion of region-specific interactions should be considered to better align the model's structure with the study's interpretive focus.
8. The regional variation in the temperature sensitivity of POC and its underlying drivers has not been adequately explored. This spatial heterogeneity is essential to supporting the study's key conclusions, especially those discussed in lines 462–478. I recommend that the authors conduct a more detailed analysis of the spatial patterns in POC temperature sensitivity

and identify the dominant environmental controls across different regions. This analysis would strengthen the manuscript's empirical basis and provide deeper insights into the mechanisms that regulate SOC responses to warming in diverse ecosystems.

9. Selecting and optimizing the random forest model is a core component of the study's methodology. However, the manuscript lacks sufficient detail about this process. It is important to clarify whether the training–testing data split was performed using random sampling or a different strategy, such as spatial blocking, and how the hyperparameters were tuned, such as through grid search, random search, or cross-validation. The evaluation metric used to select the optimal model configuration should also be explicitly stated (e.g., RMSE, R^2 , or MAE). The method used to compute variable importance (e.g., impurity-based or permutation importance) also requires clarification. Due to the significance of this modeling step to the study's findings, I strongly recommend that the authors provide the complete code or a reproducible workflow (in the response letter) to ensure transparency and reproducibility.

10. To account for inter-model variability and better capture structural uncertainty, it is important to incorporate multiple models instead of relying on a single modeling approach. I recommend that the authors apply a suite of complementary models (e.g., different machine learning algorithms or different parameter settings within the same algorithm) and report the multi-model mean as well as the range or standard deviation of the predictions across models. This would provide a more robust representation of prediction uncertainty and strengthen the credibility of the results, particularly when projecting future scenarios where model assumptions can substantially influence outcomes.

Reviewer #3

(Remarks to the Author)

I thoroughly enjoyed reading the manuscript leading by Sun et al. This manuscript presents a well-compiled global dataset on particulate organic carbon (POC) and mineral-associated organic carbon (MAOC), drawing from both published and previously unpublished sources. The inclusion of data from underrepresented regions such as South America, Africa and Russia helps improve the spatial coverage compared to existing datasets. Using this dataset, the authors identify land cover and mean annual temperature as key predictors of global POC and MAOC stocks. Their projections under future climate scenarios indicate notable POC losses in tundra and boreal forest soils by 2100, contributing substantially to total soil organic carbon (SOC) decline in these regions. This focus on POC complements earlier studies that have largely emphasized MAOC.

Overall, the manuscript is well-prepared and, with moderate revisions, would be suitable for publication. While the manuscript is generally well-structured and informative, several aspects would benefit from further clarification and refinement. I outline specific suggestions for improvement below.

1. While I appreciate the authors' efforts in harmonizing the dataset (e.g., standardizing to a common depth), the manuscript would benefit from a more detailed discussion of potential uncertainties. These include methodological differences in POC and MAOC quantification, variations in sampling depth, and potential disturbances from land use and agricultural management. A transparent acknowledgment of these uncertainties would enhance the interpretation of the results.

2. The identification of fPOC as a potential indicator of SOC vulnerability to climate change is intriguing. If validated, it could serve as a valuable proxy for assessing SOC stability at global scales. I encourage the authors to expand on this point and evaluate whether fPOC can be reliably applied across various ecosystems and climate regimes.

3. The authors estimate that projected POC losses could lead to a cumulative release of 224 Pg CO₂e by 2100. This is an important contribution, particularly given the historical underrepresentation of POC in global carbon models. However, the discussion of this figure remains somewhat superficial. The authors should clearly explain how excluding POC from carbon models could result in underestimating future CO₂ emissions.

4. The conclusion that boreal forests and tundra are hotspots for future POC and SOC losses is significant. It would be helpful to distinguish these two ecosystems more clearly. Are the mechanisms driving POC loss different between boreal forests and tundra? Are the underlying factors the same?

5. The use of random forest (RF) models to predict future changes in POC and MAOC is appropriate given the data distribution. The authors also employed the BINN model for validation, which is commendable. However, additional information on the BINN model—its assumptions, structure, and limitations—should be included, at least in the Supplementary Information. Were any inconsistencies observed between the RF and BINN projections?

6. A statement regarding data and code availability should be included.

Other Minor and Technical Comments

Line 232: Replace "SOC" with "soil organic carbon" for clarity.

Line 242: Clarify the meaning of "high latitudes"; the term appears abruptly.

Line 256: Use the full term "particulate organic carbon (POC)."

Line 268: Spell out “mineral-associated organic carbon (MAOC)” and ensure consistency throughout.

Line 282: Clarify what is meant by “POC drivers.”

Line 355: Justify the selection of climate scenarios. Are there significant differences across them?

Lines 372-374: The observed global increase in MAOC is interesting—please provide a mechanistic explanation.

Lines 377-388: Including model validation is good practice, but further methodological detail on the BINN model is needed.

Lines 397-398: The finding that POC loss may drive SOC decline in specific regions is noteworthy—consider emphasizing this more.

Line 417: Clarify why the focus is narrowed specifically to boreal forests at this point.

Line 428: Clarify the overlap and distinctions among boreal forest, tundra, and high-latitude regions. Are the dominant drivers of SOC loss the same across them?

Line 438: Provide complete information on the separate incubation study.

Lines 472-478: Add relevant citations to support the discussion.

Lines 492-507: The equal importance of POC and MAOC is effectively conveyed—consider providing concrete examples.

Lines 509-523: Expand the discussion on data and methodological uncertainties to aid interpretation.

Lines 525-533: Reiterate more clearly that POC loss is a key driver of SOC loss in some regions—this is a central conclusion.

Lines 546–550: Explain how the current dataset differs from previous global efforts.

Lines 553–557: Note any uncertainties or limitations associated with the earlier datasets.

Line 555: Ensure consistent decimal formatting across text and figures.

Line 562: Clarify whether all environmental variables used in modeling were at the same spatial resolution.

Lines 562-580: Justify the choice of specific open-access resources for environmental variables.

Line 582: How many data points were excluded during quality control? Please specify.

Line 589: Clarify how values can exceed 100%; explain the calculation method.

Line 594: Discuss whether the results would change significantly if original sampling depths were retained.

Line 636: Justify the choice of the 2081–2100 period for projections.

Lines 652-654: Describe the method used to convert POC loss into CO₂-equivalent emissions.

Line 674: Include a data and code availability statement.

Lines 680-701: Check for identical initials among authors and clarify as needed.

Lines 743-744: Spell out the full name of the journal. Review all references to ensure consistent formatting.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

While the authors have compiled an impressive global dataset and addressed the reviewers' comments with considerable effort and additional analyses, fundamental concerns regarding the robustness of the core scientific claims and the validity of the extrapolations remain unresolved. The study's central conclusion—that high-latitude soils are a global hotspot for future SOC loss driven predominantly by POC—is not adequately supported by the underlying data and modeling approach. The issues of data representativeness, model mechanistic limitations, and over-extrapolation are too significant to warrant publication in its current form.

1. The manuscript's pivotal claim hinges on high-latitude systems, yet only ~2% of the observational data originate from these regions. The authors' attempt to validate this using a model trained exclusively on the sparse high-latitude data is commendable but ultimately circular and statistically tenuous. Basing a sweeping conclusion about a global vulnerability hotspot on a dataset that is profoundly underrepresented in that very region constitutes a critical flaw. This concern is echoed by Reviewer #2, who noted the "fundamental inconsistency between the scope of model training and the focus of interpretation."
2. The reliance on Random Forest models, while useful for prediction, provides correlations rather than mechanistic understanding. The introduction of the BINN model, though a step towards process representation, does not fully resolve this issue. As Reviewer #2 points out, the BINN model is also trained globally and thus shares the same fundamental limitation in capturing region-specific mechanisms. The study identifies patterns but fails to provide a deeper, process-based explanation for why POC is so vulnerable in these systems, particularly without adequately accounting for critical factors like the distinction between free and occluded POC or explicit microbial community dynamics.
3. A key limitation of the spatial RF approach is the assumption of stationarity—that the relationships between predictors and SOC fractions will remain constant under future climate regimes. This is a strong and likely incorrect assumption for periods of rapid change. The authors acknowledge this but do not quantify how this limitation might impact their dramatic projections of POC loss and CO₂ release (e.g., 81 Pg CO₂e). The predictions of future carbon-climate feedbacks are therefore built on an unstable foundation.
4. The manuscript repeatedly presents specific quantitative findings (e.g., "POC losses account for 81 ± 10% of SOC losses") with a degree of confidence that is not justified given the methodological caveats and data limitations. The analysis lumps together distinct high-latitude ecosystems (e.g., tundra vs. boreal forest), and as noted by Reviewer #3, fails to distinguish the potentially different mechanisms driving POC loss in each. This leads to overly broad and generalized conclusions.
5. The authors correctly identify "Land Cover" as the most important predictor for POC and MAOC (Figure 2). However, their handling of future land cover is critically flawed. Their use of a global LUC simulation product ignores a crucial feedback loop: climate change itself will directly and drastically alter high-latitude land cover. For instance, Arctic warming is causing boreal forest northward expansion ("greening"), shrub encroachment, wetland formation and degradation ("browning"), and other rapid biome shifts. These changes are not primarily driven by socioeconomic scenarios (SSPs) but are direct biophysical processes forced by climate. The LUC product used likely reflects anthropogenic land-use change rather than capturing these climate-driven ecosystem transformations. Therefore, the "future tundra" upon which the model predictions are based may cease to exist or be replaced by other ecosystems by the end of the century. Using a static (or non-biophysically driven) future land cover map to predict soil carbon dynamics introduces potentially the largest unquantified error in the very regions identified as hotspots.
6. The authors' conversion of projected POC loss to CO₂ emissions by simply applying the 44/12 ratio is an overly simplistic assumption that likely leads to a severe overestimation of the actual climate feedback. This calculation implicitly assumes that 100% of the "lost" POC is mineralized to CO₂. In reality, the decomposition of POC involves several competing pathways: A portion can be transformed into MAOC (the "microbial carbon pump" process). A portion can be lost as dissolved organic carbon (DOC) through leaching. Decomposition products can be utilized for microbial anabolism, forming new microbial necromass carbon. The authors briefly mention the importance of promoting POC transformation into MAOC in the discussion, but this admission highlights the fundamental flaw in their CO₂ calculation. If a significant fraction of POC is not directly converted to CO₂, their proclaimed figure of "equivalent to 2-3 times current annual emissions" is highly misleading. A responsible estimate must discuss the partitioning among these different fates and conduct sensitivity analyses, rather than adopting the most extreme and direct conversion assumption.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

The authors very well addressed my questions. I have no further comments.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)

After thoroughly reviewing the revised manuscript, I find that the authors have provided adequate revisions to fully address the comments from me and other reviewers. This is an interesting work and I'm happy to see it can be published soon.

There are some minor comments with the aim to further improve the manuscript.

Line 238: should be "our results identify high-latitude soils as global hotspots of..."

Line 239: under what kinds of multiple climate scenarios? I think it should be more specific.

Line 241: should be "accounting for about 81 ± 10% of..."

Line 274-278: These two statements contain overlapping content and may be streamlined for clarity.

Line 330: What does it mean high-latitude model?

Line 417: controlling POC and MAOC storks or changes?

Line 497-498: "climatic sensitivity of POC losses" not clear

(Remarks on code availability)

It is able to run the code.

Reviewer #4

(Remarks to the Author)

1. What are the noteworthy results?

The study shows that high-latitude soils represent global hotspots of soil organic carbon vulnerability, with future SOC losses largely driven by particulate organic carbon (POC) rather than mineral-associated organic carbon (MAOC). Under future climate scenarios, particularly SSP585, POC exhibits a stronger sensitivity to warming and accounts for the majority of projected SOC losses. The work further proposes the fraction of POC relative to total SOC (fPOC) as a functional indicator of SOC vulnerability to climate change, supported by multiple modeling approaches, sensitivity analyses, and regional assessments. Importantly, the study quantifies the potential climatic impact of POC losses, suggesting that neglecting this fraction may lead to underestimation of future CO₂ emissions in global carbon cycle models. The consistency of these results across different modeling frameworks and scales adds robustness to the main conclusions.

2. Will the work be of significance to the field and related fields?

The work is of high significance for soil biogeochemistry, climate science, and carbon cycle modeling, as it provides an integrated assessment of the vulnerability of soil organic carbon fractions under climate change. While previous studies have recognized the functional distinction between particulate and mineral-associated SOC, most have focused on present-day distributions at regional scales or emphasized MAOC as the dominant stable carbon pool. This study advances the field by integrating an expanded global database, harmonizing diverse SOC fractionation methods within a common functional framework, and explicitly linking contemporary empirical controls to future projections under CMIP6 scenarios. By quantifying the contribution of POC to future SOC losses at the global scale and proposing its use as an indicator of climate vulnerability, the study substantially extends existing knowledge and offers clear implications for Earth system models and climate mitigation strategies.

3. Does the work support the conclusions and claims, or is additional evidence needed?

Overall, the conclusions and claims are well supported by the data and analyses presented, particularly following the revisions in response to reviewer comments. The robustness of the results is strengthened by extensive sensitivity analyses, including stratification by fractionation method, the use of regional models, and multiple independent algorithms, as well as by the explicit assessment of spatial and methodological uncertainties. The integration of biogeochemically informed modeling further supports the conceptual basis of the conclusions.

While both the reviewers and authors acknowledge limitations related to the availability of direct evidence on microbial mechanisms, POC subfractions, and deeper soil layers (>30 cm), especially in high-latitude regions, these gaps do not undermine the main findings. Rather, they highlight clear avenues for future research, which are appropriately recognized and discussed in the revised manuscript.

4. Are there any flaws in the data analysis, interpretation and conclusions? Do these prohibit publication or require revision?

No critical flaws were identified in the data analysis, interpretation, or conclusions that would preclude publication. The main weaknesses noted by the reviewers relate to methodological heterogeneity in POC and MAOC data, the relatively limited availability of observations in high-latitude regions, and assumptions associated with soil depth standardization and model stationarity. However, these issues are explicitly acknowledged, tested, and transparently discussed in the revised manuscript. Additional analyses indicate that the main patterns remain robust across different data subsets and methodological approaches, and the conclusions are carefully aligned with the scope and limitations of the available data, avoiding unwarranted extrapolation. I believe that the revised content adequately addresses and supports the discussion of the identified limitations.

5. Is the methodology sound? Does the work meet the expected standards in your field?

Yes, the methodology is sound and meets the current standards of the field. Key strengths include the explicit functional harmonization of soil organic carbon fractions, soil depth standardization based on widely accepted approaches, and the use of multiple independent models combined with cross-validation. The study also explicitly incorporates uncertainty analyses, sensitivity tests, and regional modeling, and consistently integrates empirical methods with biogeochemically informed approaches, strengthening the conceptual basis of the inferences. While some simplifications are inevitable in global-scale studies, they are methodologically justified and transparently discussed, and do not compromise the validity of the results or conclusions.

6. Is there enough detail provided in the methods for the work to be reproduced?

Yes. Following the revisions requested by the reviewers, the procedures for data harmonization, soil depth standardization, predictor selection, and model training are described clearly and in sufficient detail. Data quality control and exclusion

criteria, as well as modeling parameters and evaluation metrics, are explicitly reported. In addition, the data and code used in the analyses are publicly available, ensuring the reproducibility of the results. Overall, the study meets the expected standards of transparency and reproducibility for contemporary global syntheses.

Overall, the revised manuscript presents robust results and a sound methodological framework. While some limitations remain, they are adequately addressed and do not compromise the main conclusions, and the study appears suitable for publication.

Minor suggestion

In the Introduction, I suggest rephrasing the sentence in lines 274–278 to avoid redundancy, as the same idea is repeated in consecutive sentences.

(Remarks on code availability)

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Reviewer#1' comments

This article explores the crucial role of soil organic carbon (SOC) in the global carbon cycle and its uncertain response to climate change, with a focus on particulate organic carbon (POC) and mineral-associated organic carbon (MAOC). Based on extensive data collection and machine learning methods, it aims to reveal the global distribution, key drivers, and future trends of POC and MAOC under different climate scenarios. While the study is conceptually innovative and methodologically sound, offering new insights into SOC dynamics, it has significant issues affecting the accuracy and reliability of the results. Below are the major and minor questions.

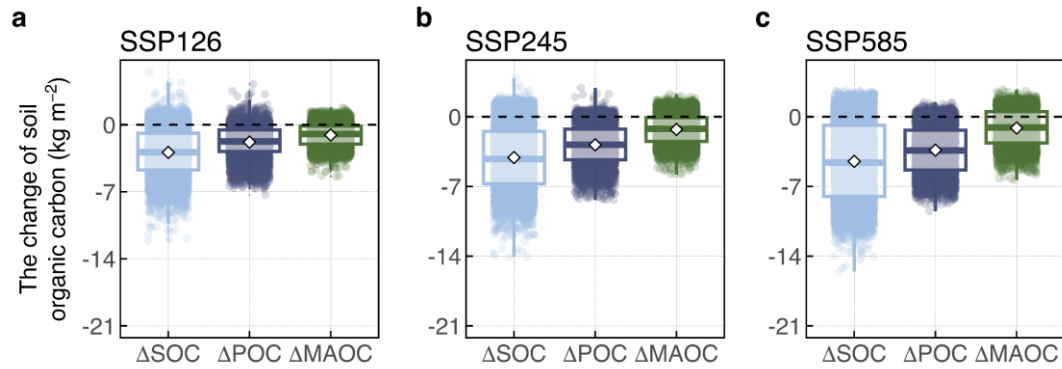
[Response] Thank you for your positive and constructive assessment. Please find below our point-by-point responses.

Major questions

1. Only 2% of observations are from polar regions, which may lead to an underestimate of SOC losses there. Polar regions are crucial in the global carbon cycle, and insufficient data makes it hard to ensure the global applicability of the conclusions.

[Response] We appreciate the reviewer's valuable comment. To address this concern, we have made every effort to include as much high-latitude data as possible in our dataset. However, the scarcity of soil carbon observations in polar regions remains a common challenge in this research field. To transparently reflect this issue, our analysis of predictive uncertainties shows higher uncertainty levels for POC and MAOC projections in high-latitude soils compared to lower latitudes (Supplementary Fig. 33, 37-39). This pattern highlights high-latitude soils where additional field data collection would be most beneficial for improving model reliability in future studies.

In addition, following Reviewer 2's suggestion, we trained a separate model using only the high-latitude soil data. The R^2 values from the high-latitude-only model were within 6% of the globally trained model, indicating comparable performance. The projections under future climate scenarios with this model indicate that POC losses in high-latitude soils predominantly drive the overall SOC losses. Notably, the magnitude of these POC losses is comparable to the estimates derived from the model trained with the global dataset, reinforcing the robustness of our projections (Supplementary Fig. 9-11, 21).



Supplementary Figure 21 The absolute changes of topsoil soil organic carbon (SOC) stock, particulate organic carbon (POC) stock, and mineral-associated organic carbon (MAOC) stock in high-latitude soils under SSP126, SSP245, and SSP585 scenarios from 2081 to 2100, based on models trained with high-latitude soils ($\geq 60^\circ \text{ N}$ or $\leq 60^\circ \text{ S}$). Box plots indicate the medians (horizontal lines), 1st and 3rd quartiles (boxes), $1.5 \times$ interquartile range (whiskers), and means (diamonds).

2. Random forest models, though good at handling large datasets and nonlinear relationships, cannot directly explain the complex biogeochemical mechanisms of SOC dynamics. A deeper understanding of these mechanisms is needed for more accurate predictions.

[Response] We appreciate the reviewer's insightful comment. The main objective in this study, however, was to identify areas where overall SOC losses are primarily driven by POC changes, rather than to fully resolve the underlying biogeochemical mechanisms. For this purpose, the random forest model is well-suited, as it effectively captures nonlinear interactions among multiple environmental drivers and enables spatially explicit predictions at the global scale.

To address your concern, we complemented the random forest approach with the BINN model, which incorporates 21 biogeochemical parameters and leverages a process-based framework for prediction. By incorporating key process-based variables (e.g., microbial efficiency, temperature sensitivity parameters, and substrate quality indices), the BINN model partially bridges the gap between empirical correlations and mechanistic understanding. While the model remains data-driven, its inclusion of biogeochemical constraints helps ensure that projections are physically plausible rather than purely statistical. As stated in the previous version of our manuscript, the BINN model projections indicate substantial POC losses at high latitudes, contributing up to $92 \pm 5\%$ of global total SOC losses under SSP585 (Supplementary Fig. 22). The details of the BINN model, including its structure, assumptions, and parameters, are described in the Supplementary Methods.

3. POC and MAOC interact in complex ways, such as during microbial decomposition. The article does not adequately explore these interactions and their impact on overall SOC dynamics, which is essential for a comprehensive understanding of SOC cycling and climate responses.

[Response] We thank the reviewer for this thoughtful and valuable comment. Indeed, POC and MAOC interact to jointly regulate SOC dynamics. In our study, we primarily emphasize that POC losses dominate SOC responses under future climate scenarios. However, in mid- and low-latitude soils, MAOC losses are relatively minor, and MAOC can even increase. Therefore, we highlight that while protecting POC, measures that promote the transformation of POC into MAOC are also important for jointly enhancing SOC stability (Lines 525-534).

Detailed questions

1. The data comes from published studies and unpublished sources, which may vary in sampling and measurement methods. How is the consistency and comparability of the data ensured? More details are needed on data standardization and quality control.

[Response] We appreciate the reviewer's thoughtful comment. Both published and unpublished POC and MAOC data were derived from topsoil (≤ 30 cm) and subsequently standardized to the 0-30 cm depth to ensure comparability across datasets. The published POC and MAOC data were obtained using size, density, or combined size–density fractionation methods, whereas the unpublished data were all measured using the size-based method. To evaluate potential methodological effects, we trained different models based on datasets derived from the various fractionation methods and conducted predictions. The predicted future trends of POC and MAOC remained consistent across these models, highlighting the robustness of our results (Supplementary Fig. 29-31). To further ensure data quality, both published and unpublished data were subjected to the same standardized quality control procedures (Lines 619-625). We have clarified this issue in the revised manuscript.

2. Land cover types are critical for the results. How were they determined for each observation point? The accuracy and reliability of this classification method, and whether it accounts for land cover changes affecting SOC, need to be addressed.

[Response] Thank you for this useful suggestion. The land cover type associated with each observation was taken directly from the original publications. For the spatial prediction of current global POC and MAOC stocks, we used the MODIS MCD12C1 product, which provides consistent, satellite-derived land cover data at the global scale. For future projections, land cover information was obtained from the global land-use and land-cover change (LUCC) simulation product (Lines 716-718). This approach accounts for projected land cover changes under different climate and socioeconomic scenarios, thereby capturing a key component of future SOC dynamics, as well as

ensuring that both current and future land cover effects on SOC fractions are considered in our analysis.

3. Environmental variables like MAT and MAP come from different databases. Do their temporal and spatial scales match the SOC observation data? Mismatches may introduce errors, which should be discussed.

[Response] Thank you for this insightful suggestion. To ensure consistency, all environmental variables used in this study were standardized to a 0.5° spatial resolution and extracted according to the latitude and longitude of each observation point. For variables such as MAT and MAP, we primarily used the values reported in the original publications. When such information was not available, we extracted the corresponding variables from publicly available global databases (Supplementary Table 2).

We acknowledge that some degree of temporal mismatch may exist between the SOC fraction data and the environmental variables from databases. However, given that most SOC observations and climate variables represent multi-year averages, and the slow-changing nature of SOC stock, we expect the impact of temporal mismatch to be minor relative to spatial variability (Wieder et al., 2018). We have explicitly addressed this source of uncertainty in the discussion section of the revised manuscript (Lines 606-610).

Reference:

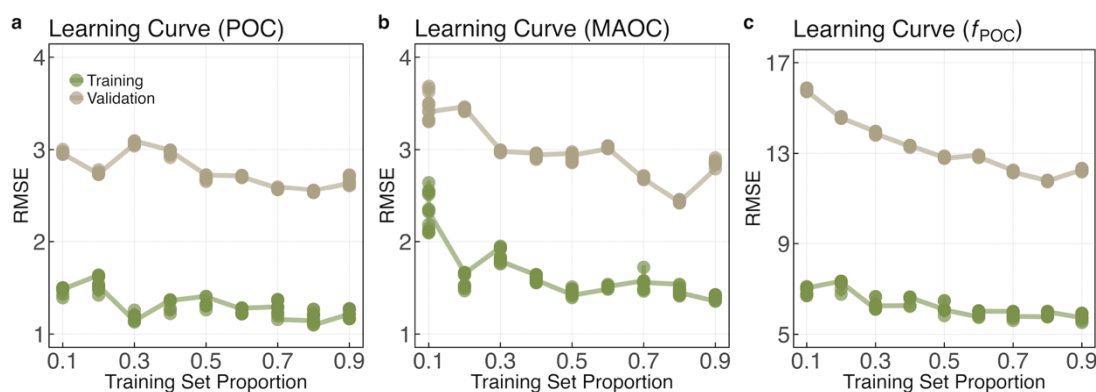
W.R. Wieder, M.D. Hartman, B.N. Sulman, et al. Carbon cycle confidence and uncertainty: Exploring variation among soil biogeochemical models. *Global Biogeochemical Cycles*, 2018, 24(4): 1563–1579.

4. With large datasets and many variables, random forest models may overfit. Overfitting could affect the reliability of future predictions.

[Response] Thank you for raising this important point. To evaluate the risk of overfitting, we generated learning curves, which show model performance as a function of training data proportion (Lines 660-663). Overall, the training root mean square error (RMSE) decreased and stabilized when the training proportion exceeded 0.5, while the validation RMSE reached its minimum at 0.8. Therefore, an 80/20 (training/validation) split was adopted for subsequent analyses. The curves indicate that model accuracy stabilizes with increasing training data, suggesting that the RF models are not overfitting and that the predictions remain reliable for future POC and MAOC projections (Supplementary Fig. 27).

In addition, to further ensure the robustness and generalizability of our random forest models, we employed a five-fold cross-validation approach using the caret R

package. Specifically, the dataset was randomly divided into five equal subsets, and in each iteration, four subsets were used for model training while the remaining one was used for validation. This process was repeated five times so that each subset served as the validation set once. The model performance metrics (e.g., RMSE and R^2) were averaged across all folds to provide a stable and unbiased estimate of prediction accuracy (Lines 678-681).



Supplementary Figure 27 Learning curves of random forest models for particulate organic carbon (POC), mineral-associated organic carbon (MAOC), and the proportion of POC relative to soil organic carbon (f_{POC}) predictions. RMSE, root mean square error.

5. The article uses three SSPs to predict SOC changes but lacks in-depth sensitivity analysis. Different SSPs imply varying social, economic, and policy factors that could differently impact SOC dynamics. The uncertainty of these scenarios on results needs more exploration.

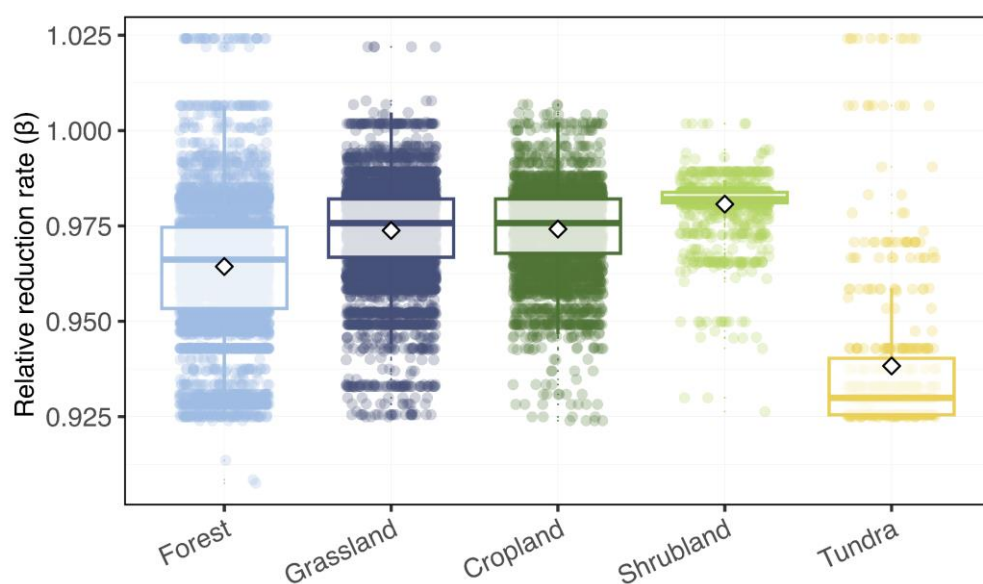
[Response] Thank you for this valuable comment. We quantified the uncertainties of predicted current and future global POC and MAOC under different SSP scenarios (Supplementary Figs. 37-39). In addition, we assessed the sensitivities of POC and MAOC to different climate scenarios based on the differences in their predicted values among the SSPs (Fig. 3; Supplementary Fig. 7-8, 34-36). The results show that high-latitude regions, which are projected to experience the largest POC and SOC losses, are also the most sensitive to variations in climate scenarios (Lines 359-376). These findings imply that mitigation policies leading to SSP126 could substantially reduce POC losses compared to the more business-as-usual SSP585 scenarios, highlighting the climatic sensitivity of SOC fractions.

6. The definitions of POC and MAOC are based on particle size, density, or both, which may be subjective. How do different separation methods affect the results? This should be discussed.

[Response] Thank you for this useful suggestion. To address differences in size vs. density separation methods, we divided the dataset into three sub-databases according to the separation method and analyzed them independently. Across all three methods, the results consistently highlighted high-latitude regions, boreal forests, and tundra as the main hotspots of future SOC losses (Supplementary Figs. 29-31). Thus, we conclude that methodological differences do not substantially affect the overall patterns and conclusions of our study.

7. The study standardizes soil depth data to 0-30 cm. Is this method suitable for all soil types and environmental conditions?

[Response] Thank you for this valuable comment. We acknowledge that a single parameter may not fully capture depth-related variations across all environmental conditions. Considering that land-use type has been identified as a major driver of POC and MAOC dynamics, we applied land-cover-specific β values for depth standardization, rather than using different β values for specific soil types or other environmental conditions (Supplementary Fig. 26). Although this approach may not fully account for all soil type-specific depth patterns, applying land-cover-specific β values better captures ecological variability in POC and MAOC vertical distribution while maintaining comparability across regions.



Supplementary Figure 26 The relative reduction rate (β) of topsoil soil organic carbon pool with increasing soil depth between different land covers.

8. Soil microorganisms play a key role in SOC decomposition and formation, especially for POC. Were changes in microbial communities and functions considered? Their omission could be a significant oversight.

[Response] Thank you for this insightful comment. We fully agree that microbial communities and functions play a key role in SOC decomposition and formation. Although our analysis in this study does not explicitly include microbial indicators, the model training incorporates predictors known to drive soil microbial community composition. Thus, the impact of community composition is accounted for to some extent.

To explore this issue in more detail, we analyzed a dataset from Chinese soils where POC and MAOC fractions were measured alongside microbial community data. These analyses indicated that fungal and bacterial biomass can directly influence the stocks of both POC and MAOC (Fig. R1).

However, only 6% of the studies in our dataset report POC and MAOC together with microbial community data. We acknowledge this issue as a limitation of our analysis. We indeed view this as a critical next step, and we are currently considering expanding our database to include microbial and enzymatic activity indicators to directly quantify biological mediation of POC-MAOC transformations.

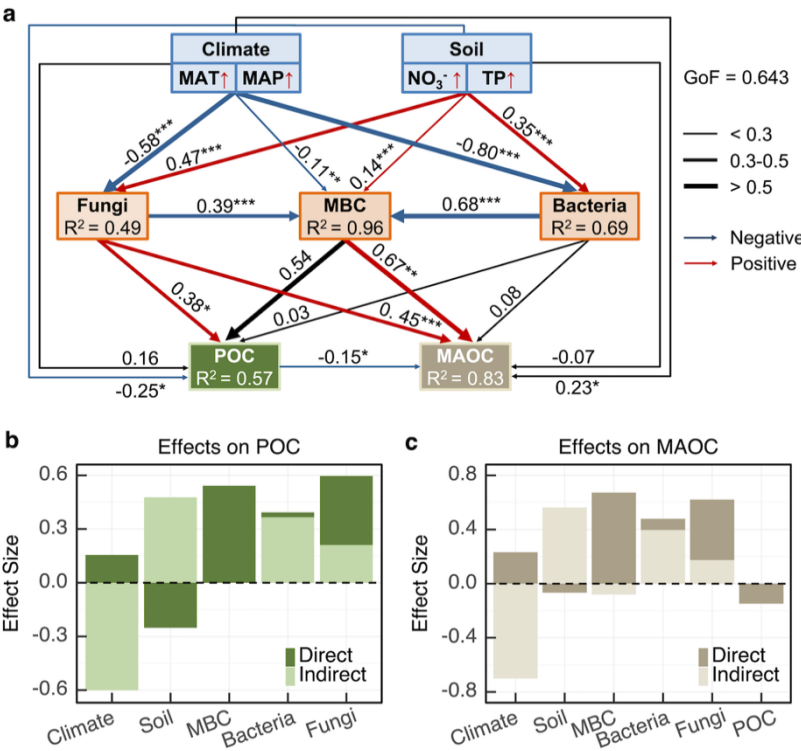


Figure R1 The result of partial least squares path models (PLSM). The red and blue line arrows denote significantly positive and negative correlations, respectively. Arrow width is proportional to the strength of the relationship, and the numbers adjacent to the arrows are standardized path coefficients. GoF: Goodness-of-fit. MAT: mean annual temperature. MAP: mean annual precipitation. NO₃⁻: soil nitrate nitrogen. TP: soil total phosphorus. MBC: soil microbial biomass carbon; Fungi: soil fungal biomass; Bacteria:

soil bacterial biomass; POC: Particulate organic carbon; MAOC: Mineral-associated organic carbon. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

We are grateful for Reviewer #1's detailed evaluation, which prompted substantial methodological refinement and additional mechanistic analyses to enhance the robustness of our conclusions. Collectively, these revisions reinforce the validity of our key findings while improving transparency and interpretability.

Reviewer#2' comments

Sun, Chen, and colleagues compiled a valuable dataset on soil organic carbon (SOC) fractions in various ecosystems. They drew attention to the vulnerability of particulate organic carbon (POC) in high-latitude regions, particularly in tundra and boreal forests. While this focus is timely and relevant to advancing the understanding of the terrestrial carbon cycle, the manuscript currently lacks sufficient empirical evidence to robustly support these claims. Significant additional analyses are necessary to substantiate the conclusions. Currently, the manuscript reads more like a perspective or hypothesis-driven commentary than a research article grounded in rigorous data analysis.

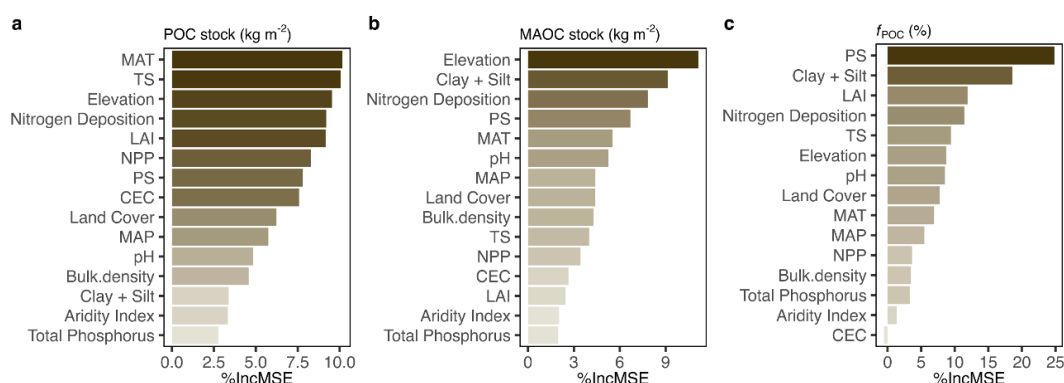
[Response] We thank the reviewer for the positive assessment. In the revised manuscript, we have added new analyses and evidence to substantiate our conclusions, ensuring the study is firmly grounded in data-driven results. Please find below point-by-point responses.

1. The study uses machine learning models that are trained and optimized with present-day spatial relationships between soil organic carbon (SOC), including the particulate (POC) and mineral-associated (MAOC) fractions, and environmental predictors. These models are then applied to predict changes in SOC under future scenarios. However, it is widely recognized that the spatial relationships between SOC and environmental drivers are likely to change under a changing climate or land use. The current modeling approach does not explicitly consider potential shifts in these relationships over time, raising concerns about the validity of future projections.

[Response] Thank you for raising this important point. We acknowledge that statistical models like random forest assume stationarity in the predictor–response relationships, which may not fully hold under future climatic regimes. To address your concern, we complemented the random forest models with the BINN model, which incorporates 21 biogeochemical parameters related to temperature sensitivity, substrate quality, and microbial turnover and is based on process-level representations of SOC formation and stabilization. Unlike purely statistical models, the BINN framework accounts for the mechanisms underlying SOC dynamics and thus provides an additional line of evidence for future projections. The BINN framework produced projections highly consistent with the RF results (Supplementary Fig. 22), especially regarding the vulnerability of high-latitude regions. This consistency across distinct models increases confidence that our conclusions are not an artifact of the method used to analyze our data. We have also added a note in the Discussion to explicitly acknowledge this potential limitation of random forest models and to highlight the complementary role of the BINN approach in addressing this issue (Lines 394-396).

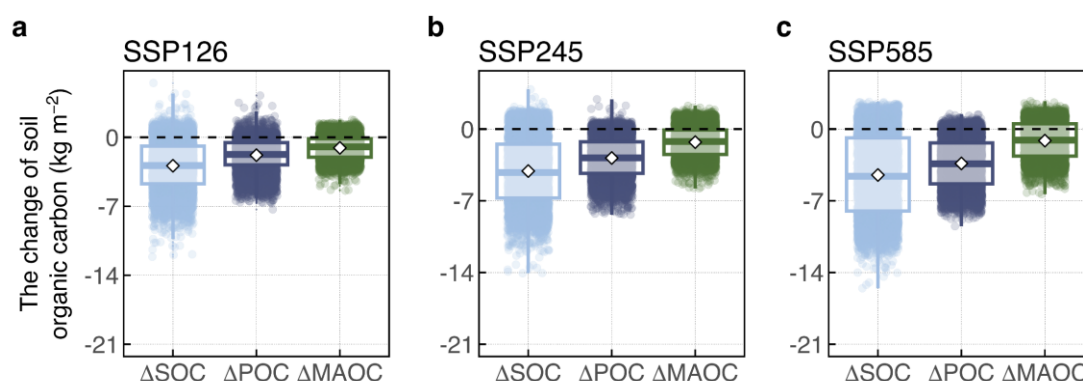
2. The manuscript places considerable emphasis on high-latitude ecosystems, particularly tundra and boreal forests. However, data from high-latitude regions account for only 2% of the total dataset. Specifically, observations from tundra and boreal forests comprise approximately 8% of the total dataset. In contrast, over 90% of the data used for model training and prediction are derived from low-latitude regions and ecosystems, such as forests, grasslands, and croplands. Given this imbalance, the conclusions regarding soil organic carbon (SOC) and its fractions in tundra and boreal systems may not be well supported by the underlying data. To strengthen the reliability of these region-specific interpretations, I recommend training and validating separate models for tundra and boreal forest ecosystems exclusively. Then, the performance and variable importance rankings from these targeted models should be compared to those obtained from models trained on the full dataset. This comparison would help determine whether general patterns derived from global data are applicable to high-latitude systems or if distinct mechanisms are at play in these vulnerable regions.

[Response] We thank the reviewer for this valuable suggestion. Following this recommendation, we conducted separate model training using high-latitude data only ($R^2 = 0.51$, RMSE = 4.83). The results showed that mean annual temperature was the key driver of POC distribution in high-latitude soils (Supplementary Fig. 1). Based on these predictors, our projections indicated that SOC losses in high-latitude soils are primarily driven by POC, which further supports our main conclusion (Supplementary Fig. 21). We have added these findings in more detail in the revised manuscript (Lines 330-332, 387-390). Due to the limited number of tundra samples, we were unable to train a robust model specifically for tundra. This consistency across distinct models increases confidence that our conclusions are not an artifact of the method used to analyze our data. We have revised the manuscript to emphasize that high-latitude soils are likely to represent global hotspots of POC-driven SOC losses, with tundra and boreal forests being the major land-use types contributing to future carbon loss.



Supplementary Figure 1 Variable importance of the machine learning random forest model trained exclusively on high-latitude soils for topsoil (a) particulate organic carbon (POC) stock, (b) mineral-associated organic carbon (MAOC) stock, and (c) the proportion of POC relative to soil organic carbon (f_{POC}). Mean annual temperature (MAT), mean annual precipitation (MAP), temperature seasonality

(TS), precipitation seasonality (PS), background nitrogen deposition (N deposition), aridity index, cation exchange capacity (CEC), percent of clay and silt (clay + silt), total phosphorus, net primary productivity (NPP), soil pH, and leaf area index (LAI) are continuous variables. Land cover is a categorical variable. Variable importance is ranked by the percent increase in mean square error (MSE).



Supplementary Figure 21 The absolute changes of topsoil soil organic carbon (SOC) stock, particulate organic carbon (POC) stock, and mineral-associated organic carbon (MAOC) stock between different forests under SSP126, SSP245, and SSP585 scenarios from 2081 to 2100, based on a model trained exclusively on high-latitude soils. Box plots indicate the medians (horizontal lines), 1st and 3rd quartiles (boxes), $1.5 \times$ interquartile range (whiskers), and means (diamonds).

3. Soil organic carbon (SOC) temperature sensitivity is influenced by the intrinsic kinetics of carbon compounds and by factors such as microbial accessibility, physical protection within soil aggregates, and substrate limitation. For example, free particulate organic carbon (POC) tends to be more sensitive to temperature, whereas POC encapsulated within soil aggregates is substantially more protected and thus less responsive to warming. Given these considerations, the assumption that warming will universally accelerate POC decomposition in high-latitude ecosystems due to POC's high temperature sensitivity warrants further scrutiny (see Introduction, Lines 271–274). This hypothesis oversimplifies the complex interactions that regulate SOC stability under changing thermal regimes and may require additional empirical support or mechanistic justification.

[Response] We appreciate the reviewer's insightful comment on the complexity of SOC temperature sensitivity. We now clarify that only the free POC fraction, which is weakly bound and microbially accessible, is expected to show strong temperature sensitivity. Occluded POC, protected within aggregates, may exhibit lower decomposition responses. We have discussed in the revised manuscript that soils with a higher proportion of free POC may indeed exhibit greater vulnerability to warming,

as this fraction is less physically protected compared to occluded POC within aggregates (Lines 459-463).

In addition, we now also acknowledge that currently available datasets reporting different POC fractions are very limited, which constrains our ability to distinguish between free POC and occluded POC in a global-scale analysis (Lines 546-549). We therefore highlight this as an important avenue for future work, emphasizing the need to integrate more comprehensive global datasets on POC subfractions to better disentangle their distinct responses to environmental change.

4. Numerous studies have investigated the environmental and biological drivers of particulate organic carbon (POC) in soils across different ecosystems and spatial scales. To provide appropriate context, I recommend that the authors summarize and cite relevant literature examining key determinants of POC dynamics. These determinants include factors such as vegetation type, climate variables, soil texture, and land use history. Given this existing body of research, the manuscript should clearly justify the necessity and novelty of identifying the key drivers of POC in the present study (see lines 282–283). A stronger rationale for this analysis will clarify how the study advances the current understanding of this topic.

[Response] We thank the reviewer for this comment. Indeed, numerous studies have examined the environmental and biological drivers of global POC, and previous work has identified factors such as vegetation type, climate variables, and soil texture as important determinants. However, while previous studies have characterized present-day POC controls, few have linked these empirical relationships to future climate projections or evaluated their implications for SOC vulnerability under CMIP6 scenarios. Furthermore, MAT emerges as one of the key drivers of global POC (Lines 288-290). To provide context, we have compiled a summary table of current literature on POC drivers (Supplementary Table 1), which highlights the key determinants identified in past studies.

The novelty of our study lies in linking these drivers to future projections of global POC under different climate scenarios, particularly in high-latitude soils. Whereas prior research mainly focused on understanding the current distribution, turnover, and controls of POC, our approach uses the established relationships between POC and environmental predictors to forecast future POC changes. Notably, our compilation integrates > 3,000 observations—over three times that of prior global efforts—allowing us to assess how established POC drivers translate into projected changes under different climate pathways. These projections can help improve Earth system models and provide guidance for soil management under climate change, thereby extending the current understanding from describing global POC patterns to predicting their responses to future environmental changes.

Supplementary Table 1 Summary of reported drivers of soil particulate organic carbon (POC).

Reference	Region	Main conclusions about POC
Hansen et al., 2023	Global	Mean annual precipitation and soil pH play key roles in global POC.
Guo et al., 2024	Global	Soil porosities are key drivers of global POC.
Zhou et al., 2024	Global	Mean annual temperature and net primary productivity are key drivers of global POC.
García-Palacios et al., 2024	Global cold regions	POC in cold regions are most vulnerable to warming.
Zhang et al., 2024	Global forests	Mean annual temperature and soil pH are the primary drivers of POC content in forest soils.
Viscarra Rossel et al., 2019	Australia	Climate factors are the main controlling factors of POC in Australia.

5. The manuscript reports the selection of 16 predictors for modeling and analysis but does not adequately explain the basis for this selection. Important variables, such as bulk density and elevation, which have been shown to significantly influence soil organic carbon (SOC) and its fractions, appear to have been omitted. The authors should clarify the criteria used to select the predictor variables and explain why certain potentially important factors were excluded. A transparent explanation would enhance the interpretability and reproducibility of the analysis.

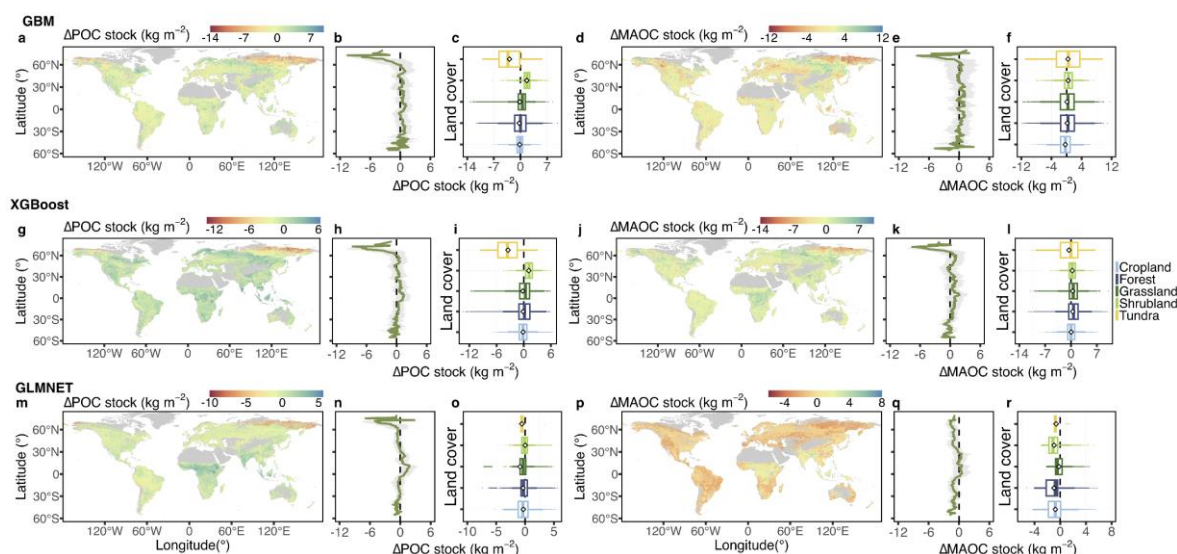
[Response] We thank the reviewer for this helpful suggestion. We selected predictors representing the major categories of drivers known to influence POC and MAOC dynamics, including climate (e.g., MAT and MAP), soil properties (e.g., pH and CEC), and vegetation (e.g., NPP and LAI). To address the reviewer’s concern, elevation was included as a predictor. Bulk density was not included as an independent predictor because it was already used in the calculation of POC and MAOC stocks. We also evaluated multicollinearity among all 16 predictors by calculating variance inflation factors (VIF), and all VIF values were below 5, indicating that multicollinearity is not a concern. The Methods section has been updated accordingly (Lines 659-660).

6. The manuscript repeatedly emphasizes that POC is more temperature-sensitive than MAOC, especially in tundra and boreal forest soils. However, Supplementary Figure 8 is the only piece of evidence supporting this claim. Since the differential temperature sensitivity between POC and MAOC is central to the study’s conclusions, it is essential to provide a more robust and comprehensive analysis (e.g., Lines 336–341, 428–430, 455–457, and 469–474). I strongly recommend incorporating multi-model comparisons to assess the robustness of the temperature sensitivity estimates, as well as clearly

describing these procedures in the Methods section. Additionally, a separate analysis focusing on high-latitude regions is necessary to substantiate the claim that POC is more temperature-sensitive than MAOC in these ecosystems. Without these additional analyses, the conclusions may be overstated or insufficiently supported by the available data.

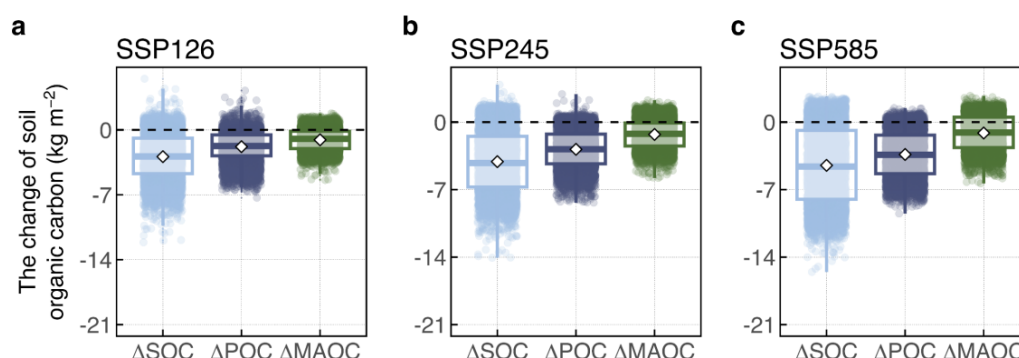
[Response] As suggested by the reviewer, we performed multi-model comparisons using random forest (RF), generalized boosted regression model (GBM), extreme gradient boosting model (XGBoost), and generalized linear model (GLMNET). All models consistently indicated that under future climate scenarios, POC losses exceed MAOC losses, confirming that POC dominates projected SOC declines (Supplementary Figs. 15-17).

In addition, we trained separate models using only high-latitude soil data. The results showed that under future climate scenarios, POC losses in high-latitude soils are greater than MAOC losses (Supplementary Fig. 1). These findings were largely consistent with those obtained from our global RF models, further demonstrating the robustness of our conclusions.



Supplementary Figure 17 Global distribution of the absolute change of topsoil (a, g, m) particulate organic carbon (POC) stock and (d, j, p) mineral-associated organic carbon (MAOC) stock under SSP585 scenario from 2081 to 2100, as predicted by generalized boosted regression models (GBM), extreme gradient boosting models (XGBoost), and generalized linear model (GLMNET). SSP, shared socioeconomic pathway. Δ POC stock and Δ MAOC stock are the differences between the future and present POC stock and MAOC stock. Here, SOC stock represents the sum of POC and MAOC stock. Δ SOC stock is the difference between the future and present SOC stock. The POC stock and MAOC stock from 2081 to 2100 were calculated using climatic factors of different models under the SSP585 scenario. The future mean annual temperature, mean annual precipitation, temperature seasonality,

precipitation seasonality, evapotranspiration, and leaf area index were the means of BCC-CSM2-MR, MPI-ESM1-2-HR, and IPSL-CM6A-LR. The future nitrogen deposition background was the mean of ACCESS-ESM1-5, NorESM2-LM, and NorESM2-MM. The future net primary productivity was the mean of IPSL-CM6A-LR, CMCC-ESM2, and CanESM5-1. All maps were at 0.5° resolution. **b, h, n, e, k, q**, Latitudinal profiles of POC stock and MAOC stock change at 0.5° latitudinal resolution. The green lines represent the absolute or relative change of POC stock and MAOC stock. The grey shading represents the standard deviation. **c, i, o, f, l, r**, The absolute change of POC stock and MAOC stock between land covers.



Supplementary Figure 21 The absolute changes of topsoil soil organic carbon (SOC) stock, particulate organic carbon (POC) stock, and mineral-associated organic carbon (MAOC) stock in latitude regions under SSP126, SSP245, and SSP585 scenarios from 2081 to 2100, based on models trained with high-latitude soils ($\geq 60^\circ \text{ N}$ or $\leq 60^\circ \text{ S}$). Box plots indicate the medians (horizontal lines), 1st and 3rd quartiles (boxes), $1.5 \times$ interquartile range (whiskers), and means (diamonds).

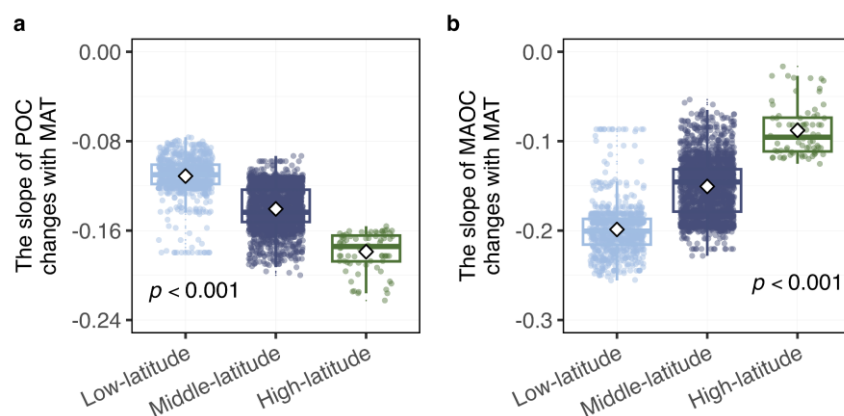
7. Although the authors introduce the Biogeochemistry-Informed Neural Network (BINN) model as supplementary analysis, it still does not resolve the fundamental inconsistency between the scope of model training and the focus of interpretation. Specifically, the BINN model is trained on a global dataset, yet the discussion focuses primarily on changes in soil organic carbon (SOC) fractions in specific regions (e.g., high-latitude ecosystems). This discrepancy raises concerns about the validity of region-specific conclusions derived from a globally trained model. To address this issue, regionally stratified modeling or the inclusion of region-specific interactions should be considered to better align the model's structure with the study's interpretive focus.

[Response] We sincerely thank the reviewer for this insightful comment. We used the Biogeochemistry-Informed Neural Network (BINN) to compensate for the limitation of the random forest model, which cannot directly represent the complex biogeochemical mechanisms underlying SOC dynamics. To address the reviewer's concern regarding the inconsistency between the training scope and interpretive focus,

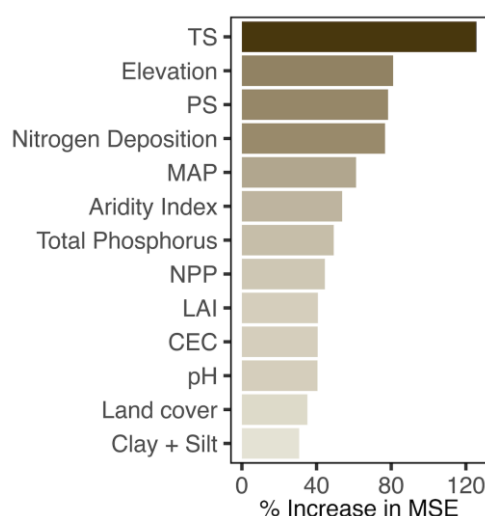
we additionally performed regionally stratified modeling. We trained and validated models using data subsets from high-latitude soils separately. The region-specific Random Forest models predicted that POC losses account for $75 \pm 3\%$ of total SOC losses under SSP585, closely matching the $81 \pm 10\%$ derived from the global model. This analysis supports the robustness of our regional conclusions and demonstrates that the global model captures consistent spatial patterns in SOC fraction dynamics. We have added this new analysis to the revised manuscript (Lines 365-368, 387-390; Supplementary Fig. 21) and believe it strengthens the validity of our region-specific interpretations.

8. The regional variation in the temperature sensitivity of POC and its underlying drivers has not been adequately explored. This spatial heterogeneity is essential to supporting the study's key conclusions, especially those discussed in lines 462–478. I recommend that the authors conduct a more detailed analysis of the spatial patterns in POC temperature sensitivity and identify the dominant environmental controls across different regions. This analysis would strengthen the manuscript's empirical basis and provide deeper insights into the mechanisms that regulate SOC responses to warming in diverse ecosystems.

[Response] We thank the reviewer for this valuable suggestion. In response, we conducted a detailed analysis of the spatial patterns of POC temperature sensitivity by calculating slopes of the POC-temperature relationship for each location. Specifically, quadratic polynomial models were fitted to capture potential non-linear patterns, and coefficients from these models were used to calculate point-specific slopes (first derivatives) at each site. These slopes quantify the rate of change in POC stocks with increasing mean annual temperature. Our results reveal that POC is more temperature-sensitive in high-latitude regions compared to middle- and low-latitude regions. Furthermore, we identified temperature and precipitation seasonality as well as elevation as the dominant environmental drivers of POC slope. These additional analyses have been incorporated into the revised manuscript (Lines 425-427, 504-508; Supplementary Figs. 23 and 25).



Supplementary Figure 23 The slope of (a) particulate organic carbon (POC) and (b) mineral-associated organic carbon (MAOC) among different latitude soils. MAT, mean annual temperature. Quadratic polynomial models were fitted to capture potential non-linear responses of POC and MAOC to MAT. Point-specific slopes (first derivatives) were calculated from the fitted models, representing the rate of change in POC and MAOC with increasing MAT. Box plots indicate the medians (horizontal lines), 1st and 3rd quartiles (boxes), $1.5 \times$ interquartile range (whiskers), and means (diamonds). The *p*-value indicates the statistical significance between different forests.



Supplementary Figure 25 Variable importance of the machine learning random forest model for the slope of particulate organic carbon (POC) stock. Mean annual temperature (MAT), mean annual precipitation (MAP), temperature seasonality (TS), precipitation seasonality (PS), background nitrogen deposition (N deposition), aridity index, cation exchange capacity (CEC), percent of clay and silt (clay + silt), total phosphorus, net primary productivity (NPP), soil pH, and leaf area index (LAI) are continuous variables. Land cover is a categorical variable. Variable importance is ranked by the percent increase in mean square error (MSE).

9. Selecting and optimizing the random forest model is a core component of the study's methodology. However, the manuscript lacks sufficient detail about this process. It is important to clarify whether the training–testing data split was performed using random sampling or a different strategy, such as spatial blocking, and how the hyperparameters were tuned, such as through grid search, random search, or cross-validation. The evaluation metric used to select the optimal model configuration should also be explicitly stated (e.g., RMSE, R^2 , or MAE). The method used to compute variable importance (e.g., impurity-based or permutation importance) also requires clarification. Due to the significance of this modeling step to the study's findings, I strongly recommend that the authors provide the complete code or a reproducible workflow (in the response letter) to ensure transparency and reproducibility.

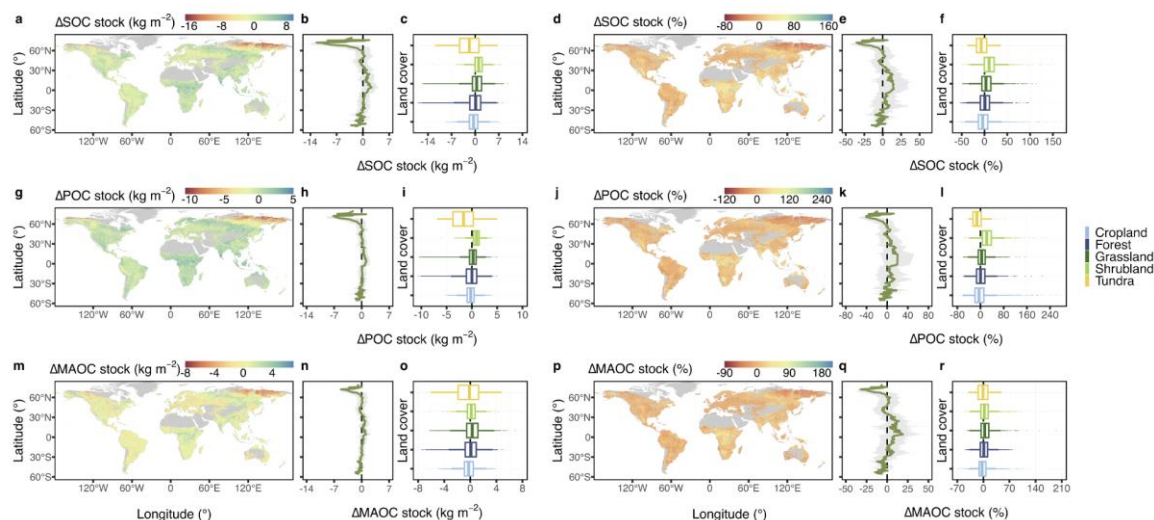
[Response] Thank you for this very useful suggestion. In the revised manuscript, we have provided a transparent and detailed description of the random forest modeling procedures (Lines 650-663). Specifically, we have now: (1) clarified that the dataset was randomly split into 80% training and 20% testing subsets; (2) detailed the hyperparameter tuning process. Random forest model optimization focused on two hyperparameters: the number of variables randomly sampled at each split (mtry) and the number of trees (ntree). We conducted a grid search across mtry ranging from 2 to 13 in steps of 2, and ntree ranging from 300 to 1500 in increments of 100. (3) specified that the evaluation metrics used to select the optimal model were RMSE and R^2 ; and (4) clarified that variable importance was quantified using permutation-based %IncMSE. To further ensure robustness and transparency, we added the parameter ranges used in the grid search, as well as learning curve analyses to assess potential overfitting (Supplementary Fig. 27). These clarifications ensure that our modeling workflow is fully transparent and reproducible, addressing the reviewer's concerns. All model scripts and parameter grids are available at the Figshare repository for full reproducibility (<https://figshare.com/s/501e92df94a15ae4ebfe>).

10. To account for inter-model variability and better capture structural uncertainty, it is important to incorporate multiple models instead of relying on a single modeling approach. I recommend that the authors apply a suite of complementary models (e.g., different machine learning algorithms or different parameter settings within the same algorithm) and report the multi-model mean as well as the range or standard deviation of the predictions across models. This would provide a more robust representation of prediction uncertainty and strengthen the credibility of the results, particularly when projecting future scenarios where model assumptions can substantially influence outcomes.

[Response] We appreciate the valuable suggestion to incorporate multiple models. Following this recommendation, in addition to the random forest (RF) model, we trained generalized boosted regression model (GBM), extreme gradient boosting model (XGBoost), and generalized linear model (GLMNET), with hyperparameter tuning for each algorithm to ensure optimal performance. These models were then applied to predict future changes in POC and MAOC (Supplementary Figs. 15-17).

The overall patterns across models were consistent, supporting the robustness of our findings. Importantly, our model comparison indicated that the RF model achieved the lowest RMSE and the highest R^2 among all tested algorithms (Supplementary Fig. 28). For this reason, we report RF-based projections as our main results, while also presenting outcomes from other models to provide a measure of structural uncertainty. In addition, we calculated the mean of multi-model results (Supplementary Figs. 18-20), which further highlights high-latitude regions as future hotspots of SOC loss and POC. This ensemble framework demonstrates that despite algorithmic differences, all

models converge on the same spatial pattern of high-latitude POC loss, reinforcing the robustness of our conclusions.



Supplementary Figure 20 Global distribution of the absolute and relative changes in topsoil (a, d) soil organic carbon (SOC), (g, j) particulate organic carbon (POC), and (m, p) mineral-associated organic carbon (MAOC) stocks under SSP585 scenario for 2081-2100, based on the multi-model mean of four machine learning models (random forest model, generalized boosted regression model, extreme gradient boosting model, and generalized linear model). SSP, shared socioeconomic pathway. Δ POC stock and Δ MAOC stock are the differences between the future and present POC stock and MAOC stock. Here, SOC stock represents the sum of POC and MAOC stock. Δ SOC stock is the difference between the future and present SOC stock. The POC stock and MAOC stock from 2081 to 2100 were calculated using climatic factors of different models under SSP585 scenario. The future mean annual temperature, mean annual precipitation, temperature seasonality, precipitation seasonality, evapotranspiration, and leaf area index were the means of BCC-CSM2-MR, MPI-ESM1-2-HR, and IPSL-CM6A-LR. The future nitrogen deposition background was the mean of ACCESS-ESM1-5, NorESM2-LM, and NorESM2-MM. The future net primary productivity was the mean of IPSL-CM6A-LR, CMCC-ESM2, and CanESM5-1. All maps were at 0.5° resolution. **b, e, h, k, n, q**, Latitudinal profiles of SOC stock, POC stock, and MAOC stock change at 0.5° latitudinal resolution. The green lines represent the absolute or relative change of SOC stock, POC stock, and MAOC stock. The grey shading represents the standard deviation. **c, f, i, l, o, r**, The absolute and relative change of SOC stock, POC stock, and MAOC stock between land covers.

Together, these additional analyses, regional validations, and expanded methodological details directly address Reviewer #2's concerns about empirical

robustness and modeling validity. They collectively confirm that high-latitude POC losses are a consistent and dominant driver of future global SOC decline.

We again thank Reviewer #2 for the thorough and thoughtful feedback. Addressing these comments led us to expand our analyses, clarify model assumptions, and strengthen the empirical foundations of our conclusions, which has greatly improved the manuscript.

Reviewer#3' comments

I thoroughly enjoyed reading the manuscript leading by Sun et al. This manuscript presents a well-compiled global dataset on particulate organic carbon (POC) and mineral-associated organic carbon (MAOC), drawing from both published and previously unpublished sources. The inclusion of data from underrepresented regions such as South America, Africa and Russia helps improve the spatial coverage compared to existing datasets. Using this dataset, the authors identify land cover and mean annual temperature as key predictors of global POC and MAOC stocks. Their projections under future climate scenarios indicate notable POC losses in tundra and boreal forest soils by 2100, contributing substantially to total soil organic carbon (SOC) decline in these regions. This focus on POC complements earlier studies that have largely emphasized MAOC.

Overall, the manuscript is well-prepared and, with moderate revisions, would be suitable for publication. While the manuscript is generally well-structured and informative, several aspects would benefit from further clarification and refinement. I outline specific suggestions for improvement below.

[Response] We sincerely thank the reviewer for the positive evaluation and insightful suggestions on our manuscript. In the revised version of the manuscript, we have carefully addressed all comments point by point, as detailed below.

1. While I appreciate the authors' efforts in harmonizing the dataset (e.g., standardizing to a common depth), the manuscript would benefit from a more detailed discussion of potential uncertainties. These include methodological differences in POC and MAOC quantification, variations in sampling depth, and potential disturbances from land use and agricultural management. A transparent acknowledgment of these uncertainties would enhance the interpretation of the results.

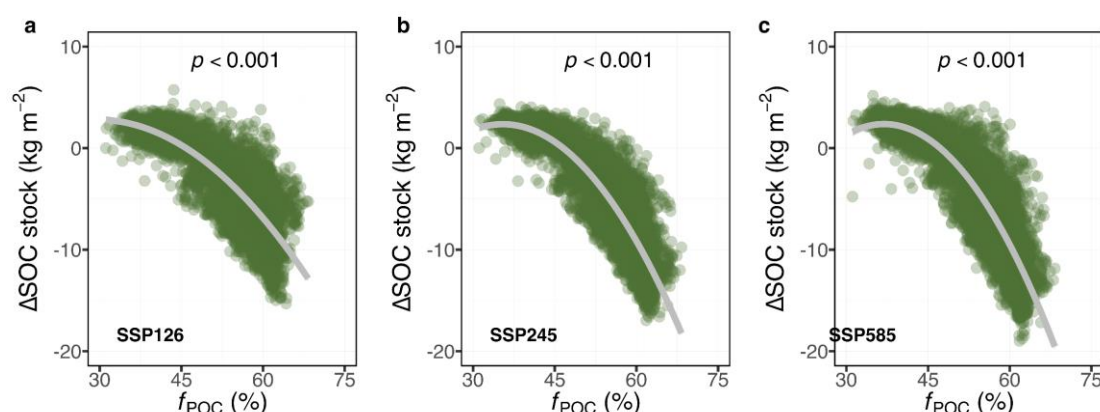
[Response] We thank the reviewer for this insightful comment. Methodological differences in the quantification of POC and MAOC may introduce bias. To minimize potential methodological biases, we grouped samples according to the measurement method and conducted separate model training and prediction. Despite these differences in methodological categories, the predicted future trends of POC and MAOC remained consistent, highlighting the robustness of our results (Supplementary Figs. 29-31).

Likewise, variations in reported sampling depth across studies can influence SOC fraction estimates. To address this, we standardized soil layers by land-use type and used the homogenized data for model training and prediction.

In addition, we explicitly discuss in the manuscript the uncertainties introduced by land-use disturbances. Agricultural management practices, such as tillage, fertilization, and residue retention, substantially alter SOC dynamics, especially in croplands. This is an important source of uncertainty, and future work should incorporate management-specific datasets to refine predictions of POC and MAOC under changing climate conditions (Lines 541-545).

2. The identification of f_{POC} as a potential indicator of SOC vulnerability to climate change is intriguing. If validated, it could serve as a valuable proxy for assessing SOC stability at global scales. I encourage the authors to expand on this point and evaluate whether f_{POC} can be reliably applied across various ecosystems and climate regimes.

[Response] We thank the reviewer for this valuable comment. Indeed, our analysis has shown that f_{POC} is closely linked to SOC vulnerability under climate change, which is further confirmed in high-latitude regions with high f_{POC} (Supplementary Fig. 24). Therefore, f_{POC} can be considered an effective indicator for identifying priority areas for SOC stabilization. This discussion has been incorporated into the revised manuscript (Lines 477-480).



Supplementary Figure 24 Relationship between the predicted proportion of particle organic carbon (POC) relative to soil organic carbon (f_{POC}) and the absolute change of (a-c) soil organic carbon (SOC) stock in high-latitude under different climate scenarios from 2081 to 2100. SSP, shared socioeconomic pathway. Δ SOC stock is the difference between the future and present SOC stock. Different colors of points represent the various land covers.

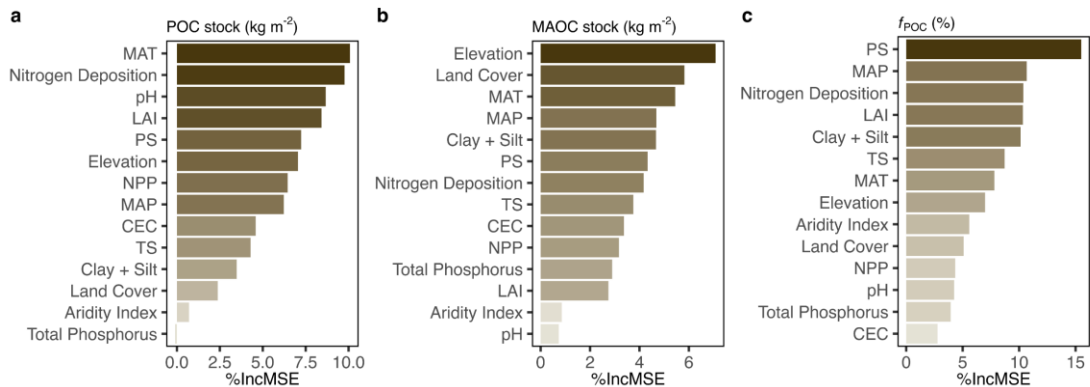
3. The authors estimate that projected POC losses could lead to a cumulative release of 224 Pg CO_{2e} by 2100. This is an important contribution, particularly given the historical underrepresentation of POC in global carbon models. However, the discussion of this figure remains somewhat superficial. The authors should clearly explain how excluding POC from carbon models could result in underestimating future CO₂ emissions.

[Response] We thank the reviewer for this valuable suggestion. Following the comment, we have expanded the discussion to explicitly explain how the exclusion of POC from global carbon models could lead to the underestimation of future CO₂ emissions (Lines 489–496). Specifically, POC is far more climate sensitive than SOC as a whole. Therefore, only including SOC as a whole pool might underestimate carbon losses with warming, especially because areas with the highest levels of predicted warming (high latitudes) generally have higher f_{POC} values. We write about this in the discussion.

4. The conclusion that boreal forests and tundra are hotspots for future POC and SOC losses is significant. It would be helpful to distinguish these two ecosystems more clearly. Are the mechanisms driving POC loss different between boreal forests and tundra? Are the underlying factors the same?

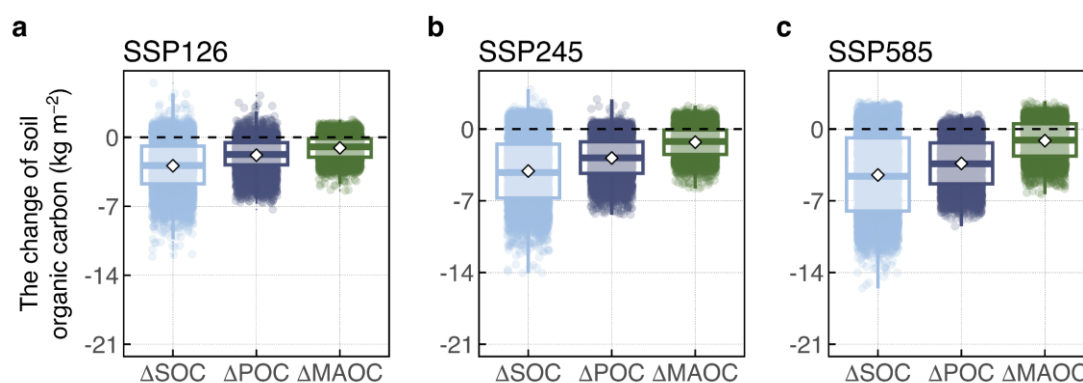
[Response] We agree that separating tundra and boreal systems would be ideal; however, the limited number of tundra sites precludes a statistically robust tundra-only model. Instead, we focused on high-latitude soil as a whole and conducted separate model training and predictions for this region. Nevertheless, our high-latitude model indicates that mean annual temperature remains a key driver of POC stocks in high-latitude soils, highlighting its central role in controlling carbon vulnerability across this region.

The results from our high-latitude models show that mean annual temperature is the key driver of POC distribution, and the projected POC losses under SSP585 scenarios account for approximately $75 \pm 3\%$ of the corresponding SOC losses (Supplementary Figs. 1 and 21). As further discussed in the revised manuscript (Lines 330-332, 387-390), we provide additional analysis and interpretation of these controls on POC dynamics in high-latitude soils.



Supplementary Figure 1 Variable importance of the machine learning random forest model trained exclusively on high-latitude soils for topsoil (a) particulate organic carbon (POC) stock, (b) mineral-associated organic carbon (MAOC)

stock, and (c) the proportion of POC relative to soil organic carbon (f_{POC}). Mean annual temperature (MAT), mean annual precipitation (MAP), temperature seasonality (TS), precipitation seasonality (PS), background nitrogen deposition (N deposition), aridity index, cation exchange capacity (CEC), percent of clay and silt (clay + silt), total phosphorus, net primary productivity (NPP), soil pH, and leaf area index (LAI) are continuous variables. Land cover is a categorical variable. Variable importance is ranked by the percent increase in mean square error (MSE).



Supplementary Figure 23 The slope of (a) particulate organic carbon (POC) and (b) mineral-associated organic carbon (MAOC) among different latitude soils. MAT, mean annual temperature. Quadratic polynomial models were fitted to capture potential non-linear responses of POC and MAOC to MAT. Point-specific slopes (first derivatives) were calculated from the fitted models, representing the rate of change in POC and MAOC with increasing MAT. Box plots indicate the medians (horizontal lines), 1st and 3rd quartiles (boxes), $1.5 \times$ interquartile range (whiskers), and means (diamonds). The p -value indicates the statistical significance between different forests.

5. The use of random forest (RF) models to predict future changes in POC and MAOC is appropriate given the data distribution. The authors also employed the BINN model for validation, which is commendable. However, additional information on the BINN model—its assumptions, structure, and limitations—should be included, at least in the Supplementary Information. Were any inconsistencies observed between the RF and BINN projections?

[Response] The BINN model integrates process-based parameters (e.g., microbial efficiency and temperature sensitivity parameters) into a neural-network framework, thereby bridging empirical and mechanistic approaches. We have added a comprehensive description of BINN, including its assumptions, structure, and limitations, in the Supplementary Information.

We found that the BINN predictions of spatial patterns were generally similar to those from the random forest models, with both approaches indicating that future SOC losses are mainly concentrated in high-latitude regions and are primarily driven by POC losses. However, minor discrepancies occurred across different land-use types: in the

BINN results, future changes in POC and MAOC showed smaller differences among land-use types compared to the RF model predictions. These differences between models likely reflect differing sensitivities to land-cover inputs rather than structural inconsistencies.

6. A statement regarding data and code availability should be included.

[Response] In the revised manuscript, we have added a Data and Code Availability Statement at the end of the Methods section. All data used in this study are publicly available or provided in the Supplementary Information. The code used for data analysis and model development is available at <https://figshare.com/s/501e92df94a15ae4ebfe>, ensuring transparency and reproducibility of our work.

Other Minor and Technical Comments

Line 232: Replace “SOC” with “soil organic carbon” for clarity.

[Response] We have revised the manuscript accordingly (Line 234).

Line 242: Clarify the meaning of “high latitudes”; the term appears abruptly.

[Response] We have revised the manuscript accordingly (Line 335).

Line 256: Use the full term “particulate organic carbon (POC).”

[Response] We have defined the full term “particulate organic carbon (POC)” at its first appearance in the manuscript (Line 263).

Line 268: Spell out “mineral-associated organic carbon (MAOC)” and ensure consistency throughout.

[Response] We have defined the full term “mineral-associated organic carbon (MAOC)” at its first appearance in the manuscript (Line 264).

Line 282: Clarify what is meant by “POC drivers.”

[Response] We have revised the manuscript to clarify the meaning of “POC drivers” (Lines 288-290).

Line 355: Justify the selection of climate scenarios. Are there significant differences across them?

[Response] We have added justification in the revised manuscript (Lines 696-706), clarifying that the selected scenarios (SSP126, SSP245, SSP585) represent contrasting climate pathways with significant differences in projected temperature and precipitation.

Lines 372-374: The observed global increase in MAOC is interesting—please provide a mechanistic explanation.

[Response] Thank you for this suggestion. While climate change accelerates the decomposition of POC, the relative stability or even increase in MAOC may partly result from microbial-mediated transformation of POC into MAOC, which can buffer potential MAOC losses.

Lines 377-388: Including model validation is good practice, but further methodological detail on the BINN model is needed.

[Response] Thank you for this suggestion. We have now provided additional methodological details of the BINN model in the supplementary material to clarify its assumptions, structure, and limitations (Supplementary material Lines 164-178).

Lines 397-398: The finding that POC loss may drive SOC decline in specific regions is noteworthy—consider emphasizing this more.

[Response] We have revised the last sentence to emphasize that POC losses are a key driver of SOC losses (Lines 408-415).

Line 417: Clarify why the focus is narrowed specifically to boreal forests at this point.

[Response] We narrowed the focus to boreal forests because, in addition to tundra, forests represent the largest contributors to future POC losses at the global scale, and

these losses are concentrated in boreal forests. For this reason, we highlighted boreal forests separately to better illustrate their critical role in global carbon vulnerability under climate change. We have clarified this in the manuscript (Lines 453-455).

Line 428: Clarify the overlap and distinctions among boreal forest, tundra, and high-latitude regions. Are the dominant drivers of SOC loss the same across them?

[Response] We clarified that tundra and boreal forests were identified as POC loss hotspots primarily from the perspective of land cover, and these land covers are predominantly located in high-latitude regions. Due to the limited number of tundra samples, it is currently not possible to train a robust model specifically for tundra ecosystems. Instead, we conducted separate model training and predictions for high-latitude soils as a whole. The results show that mean annual temperature is the dominant driver of POC distribution in high-latitude soils, and projected POC losses under SSP585 scenarios account for approximately $75 \pm 3\%$ of the corresponding SOC losses. We have clarified these points in the revised manuscript (Lines 330-332, 387-390).

Line 438: Provide complete information on the separate incubation study.

[Response] We have added more details on the referenced incubation study to clarify the experimental context (Lines 421-426). Qin et al. (2024) conducted large-scale topsoil sampling on the Tibetan Plateau and fractionated soils into POC and MAOC. Each fraction was separately incubated under controlled laboratory conditions for 300 days, and the temperature sensitivity of carbon decomposition (Q10) was measured. They found that POC exhibited higher Q10 and greater microbial diversity than MAOC, with distinct microbial community composition and co-occurrence patterns. These results indicate that POC is more labile and temperature-sensitive compared to MAOC, supporting its heightened vulnerability under climate warming.

Lines 472-478: Add relevant citations to support the discussion.

[Response] We have added the relevant citations to support the discussion (Lines 510-512).

Lines 492-507: The equal importance of POC and MAOC is effectively conveyed—consider providing concrete examples.

[Response] The original last sentence may have introduced some ambiguity. We revised it to emphasize the critical role of POC in SOC sequestration and stability

(Lines 525-534). The main purpose of this paragraph is to highlight the importance of POC and to stress that it should not be overlooked in SOC management, while MAOC remains important as discussed earlier.

Lines 509-523: Expand the discussion on data and methodological uncertainties to aid interpretation.

[Response] We have expanded the discussion on data and methodological uncertainties in Lines 600-610 and 625-629.

Lines 525-533: Reiterate more clearly that POC loss is a key driver of SOC loss in some regions—this is a central conclusion.

[Response] We have revised the text to clearly emphasize that POC loss is a key driver of SOC decline in certain regions (Lines 408-415).

Lines 546–550: Explain how the current dataset differs from previous global efforts.

[Response] We have clarified in the revised manuscript how our database differs from previous global efforts (Lines 308-311, 613-625). Specifically, our dataset includes a larger number of observations, incorporates a substantial amount of unpublished data, and improves spatial coverage across different land covers. Furthermore, we applied strict quality-control criteria, including exclusion of extreme POC and MAOC and verification of methodological consistency. These improvements enhance both the reliability and comparability of our dataset relative to prior global compilations.

Lines 553–557: Note any uncertainties or limitations associated with the earlier datasets.

[Response] Some previous datasets included unusually high soil carbon recovery rates. In addition, the coverage of these datasets was limited in regions such as Africa and Australia. In this study, we improved data quality and spatial coverage to address these limitations, providing a more reliable and consistent dataset (Lines 613-625).

Line 555: Ensure consistent decimal formatting across text and figures.

[Response] We have revised this section accordingly and ensured consistent decimal formatting across the text and figures.

Line 562: Clarify whether all environmental variables used in modeling were at the same spatial resolution.

[Response] We have clarified in the text that although the environmental variables were obtained from different sources, they were all resampled to a uniform spatial resolution of 0.5° before being used in the modeling (Lines 586-588).

Lines 562-580: Justify the choice of specific open-access resources for environmental variables.

[Response] We selected these specific open-access resources because they are widely used and well-validated in global studies. Each of the chosen databases provides global coverage at relatively high spatial resolution and has been extensively applied in Earth system modeling and empirical analyses. We have clarified this rationale in the revised text (Lines 606-610).

Line 582: How many data points were excluded during quality control? Please specify.

[Response] During quality control, 489 observations were excluded because they did not meet the predefined criteria. We have added this information to the revised text (Lines 628-630).

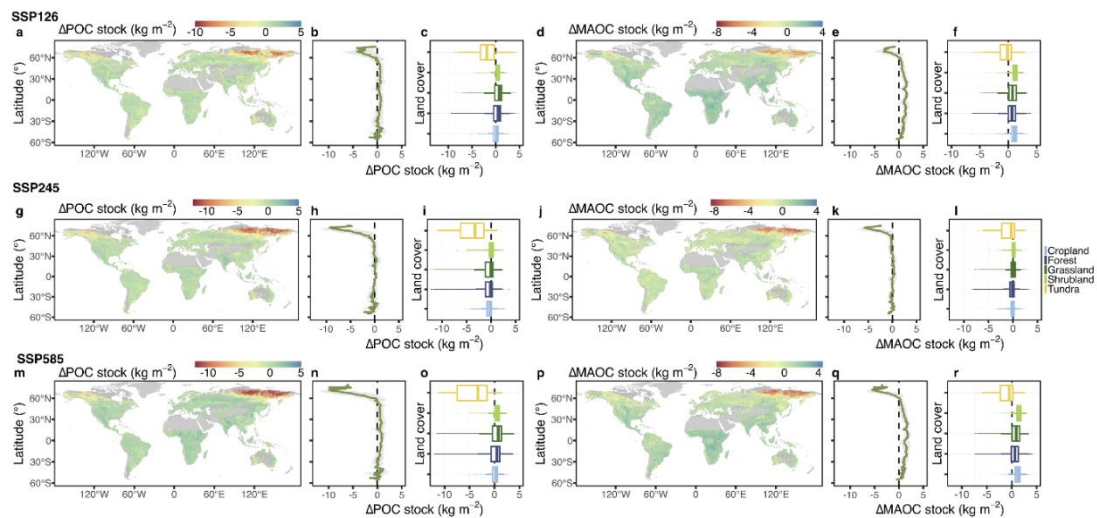
Line 589: Clarify how values can exceed 100%; explain the calculation method.

[Response] Soil organic carbon (SOC) recovery (%) was calculated as the sum of POC and MAOC stocks divided by the total SOC stock. Occasionally, the calculated SOC recovery exceeds 100%. This can be attributed to analytical uncertainties, cumulative errors from separately measuring POC and MAOC, and slight differences in extraction efficiencies compared with total SOC determination. Such deviations are common in SOC fractionation studies and are considered acceptable within the experimental error range (Lines 627-629).

Line 594: Discuss whether the results would change significantly if original sampling depths were retained.

[Response] We conducted additional analyses using the original, non-standardized sampling depths. The results were similar to those obtained after standardizing POC and MAOC to a uniform 0-30 cm depth (Supplementary Fig. 32). While standardization

does not affect the key conclusions, it allows for a more consistent and interpretable analysis across sites.



Supplementary Figure 32 Global distribution of the absolute change of topsoil (a, g, m) particulate organic carbon (POC) stock and (d, j, p) mineral-associated organic carbon (MAOC) stock under SSP126, SSP245, and SSP585 scenarios from 2081 to 2100, based on models trained with the non-standardized dataset. SSP, shared socioeconomic pathway. Δ POC stock and Δ MAOC stock are the differences between the future and present POC stock and MAOC stock. b, h, n, e, k, q Latitudinal profiles of POC stock and MAOC stock change at 0.5° latitudinal resolution. The green lines represent the absolute change of POC stock and MAOC stock. The grey shading represents the standard deviation. c, i, o, f, l, r, The absolute change of POC stock and MAOC stock between land covers.

Line 636: Justify the choice of the 2081–2100 period for projections.

[Response] This timeframe corresponds to the "long term" period defined in the IPCC Sixth Assessment Report (AR6), which assesses climate changes relative to the 1850–1900 baseline. Using this period enables a robust evaluation of carbon stock responses under stabilized future climate conditions and facilitates comparison with previous studies and IPCC projections. This is now mentioned in Lines 696–706.

Lines 652–654: Describe the method used to convert POC loss into CO₂-equivalent emissions.

[Response] We have clarified the method used to convert POC or MAOC losses into CO₂-equivalent (CO₂e) emissions in the revised manuscript. Specifically, the projected POC loss (in kg C) was multiplied by the molecular weight ratio between CO₂ and C

(44/12) to obtain the equivalent CO₂ emission. This clarification has been added to the revised manuscript (Lines 722-726).

Line 674: Include a data and code availability statement.

[Response] We have added a “Data and code availability” statement at the end of Lines 766-769 to provide information on access to all datasets and R scripts used in this study. The data and code are publicly available at <https://figshare.com/s/501e92df94a15ae4ebfe>.

Lines 680-701: Check for identical initials among authors and clarify as needed.

[Response] In the **Authors' contributions** (Lines 776-796), we carefully checked for identical initials among authors. Where initials were the same, we have clarified by writing the full surname while keeping the given name as an initial, in order to avoid confusion.

Lines 743-744: Spell out the full name of the journal. Review all references to ensure consistent formatting.

[Response] We have spelled out the full name of the journal for reference and carefully reviewed all references to ensure consistent formatting throughout the manuscript.

We sincerely thank Reviewer #3 for their positive evaluation and constructive suggestions. Their thoughtful comments helped us refine our interpretations, clarify methodological details, and better articulate the broader implications of our findings. Addressing these points has significantly improved the clarity, transparency, and overall quality of the manuscript.

Reviewer#1' comments

While the authors have compiled an impressive global dataset and addressed the reviewers' comments with considerable effort and additional analyses, fundamental concerns regarding the robustness of the core scientific claims and the validity of the extrapolations remain unresolved. The study's central conclusion—that high-latitude soils are a global hotspot for future SOC loss driven predominantly by POC—is not adequately supported by the underlying data and modeling approach. The issues of data representativeness, model mechanistic limitations, and over-extrapolation are too significant to warrant publication in its current form.

[Response] We thank the reviewer for this critical synthesis and concerns regarding our data and modeling framework. We fully acknowledge that global-scale assessments of POC and MAOC face inherent challenges regarding data sparsity in high-latitude regions, the limitations of spatial stationarity in empirical models, and the complexity of carbon-climate feedbacks. Nevertheless, our study integrates the largest global POC and MAOC dataset to date, harmonizes diverse methods, and links observations to future climate scenarios, providing novel insights into the vulnerability of high-latitude soils and the functional role of POC.

We have provided detailed, point-by-point responses to each of the specific concerns below, which we believe further demonstrate the reliability of our findings and their significance in identifying global soil carbon loss hotspots.

1. The manuscript's pivotal claim hinges on high-latitude systems, yet only ~2% of the observational data originate from these regions. The authors' attempt to validate this using a model trained exclusively on the sparse high-latitude data is commendable but ultimately circular and statistically tenuous. Basing a sweeping conclusion about a global vulnerability hotspot on a dataset that is profoundly underrepresented in that very region constitutes a critical flaw. This concern is echoed by Reviewer #2, who noted the "fundamental inconsistency between the scope of model training and the focus of interpretation."

[Response]

We thank the reviewer for this critical evaluation of our data distribution. While we acknowledge that observations from high-latitude regions constitute a small percentage of the total dataset, we contend that the robustness of our conclusion rests on the extensive environmental gradients captured by the 3,284 global observations. In machine learning and biogeochemical modeling, the breadth of climate and edaphic variables often outweighs the sheer number of samples from a specific geographic box. Consequently, our assessment of high-latitude vulnerability is not solely dependent on localized high-latitude observations.

Furthermore, to address the concern of "circular and statistically tenuous," we utilized a Biogeochemistry-Informed Neural Network (BINN). Unlike purely data-driven models, the BINN incorporates mechanistic constraints (e.g., carbon transformation pathways and biological kinetics). The fact that the BINN yields conclusions consistent with our other machine learning models suggests that the identified POC vulnerability in high-latitude soils is not a statistical artifact of sparse sampling, but a result of consistent biological signals found across the global gradient.

2. The reliance on Random Forest models, while useful for prediction, provides correlations rather than mechanistic understanding. The introduction of the BINN model, though a step towards process representation, does not fully resolve this issue. As Reviewer #2 points out, the BINN model is also trained globally and thus shares the same fundamental limitation in capturing region-specific mechanisms. The study identifies patterns but fails to provide a deeper, process-based explanation for why POC is so vulnerable in these systems, particularly without adequately accounting for critical factors like the distinction between free and occluded POC or explicit microbial community dynamics.

[Response] We appreciate the reviewer's insightful comment. While Random Forest models primarily capture empirical correlations, the BINN model provides process-based support for our conclusions by explicitly linking soil carbon fractions to biogeochemical mechanisms. We acknowledge that a more complete mechanistic

understanding of why POC is particularly vulnerable—especially regarding distinctions between free and occluded POC or microbial community dynamics—would require more detailed datasets. At present, global datasets on POC subfractions remain limited, which constrains our ability to make more fine-grained predictions. We have therefore acknowledged this limitation in the manuscript's discussion of uncertainties (Lines 526-529). Nevertheless, our combined modeling framework robustly supports the finding that POC is the dominant contributor to future SOC losses in high-latitude soils, and future studies incorporating free POC, occluded POC, and microbial data will further refine these mechanistic insights.

3. A key limitation of the spatial RF approach is the assumption of stationarity—that the relationships between predictors and SOC fractions will remain constant under future climate regimes. This is a strong and likely incorrect assumption for periods of rapid change. The authors acknowledge this but do not quantify how this limitation might impact their dramatic projections of POC loss and CO₂ release (e.g., 81 Pg CO₂e). The predictions of future carbon-climate feedbacks are therefore built on an unstable foundation.

[Response] We thank the reviewer for this important comment. We acknowledge that random forest models rely on the assumption of stationarity, which may introduce uncertainties when projecting responses under future climate conditions. However, we contend that the robustness of our projections is supported by the fact that our model is trained on a vast global dataset ($n = 3,284$) that captures extensive environmental gradients, including extreme temperatures and moisture levels. By training on these diverse contemporary analogues, the RF model effectively "learns" the responses of soil carbon to conditions that high-latitude regions are projected to encounter in the future, thereby mitigating the risks of temporal extrapolation. Consequently, we argue that while non-stationarity introduces uncertainty into the absolute magnitude of future CO₂ release, our model provides a robust estimation of SOC vulnerability.

In addition, our intention in converting projected POC losses to CO₂-equivalent values was to illustrate the potential vulnerability of POC under climate change. To

clarify this point, we have revised the manuscript and explicitly stated this interpretation in Lines 464-465, emphasizing that the calculation is intended to highlight the potential vulnerability associated with projected POC losses.

4. The manuscript repeatedly presents specific quantitative findings (e.g., "POC losses account for $81 \pm 10\%$ of SOC losses") with a degree of confidence that is not justified given the methodological caveats and data limitations. The analysis lumps together distinct high-latitude ecosystems (e.g., tundra vs. boreal forest), and as noted by Reviewer #3, fails to distinguish the potentially different mechanisms driving POC loss in each. This leads to overly broad and generalized conclusions.

[Response] We thank the reviewer for this important comment. We acknowledge that the high-latitude region encompasses diverse ecosystems, which are governed by distinct biophysical processes and carbon stabilization mechanisms. To move beyond broad generalizations, we conducted a granular analysis by separately evaluating the projected carbon dynamics for these two biomes (Lines 357–365). Our results demonstrate that POC losses remain the dominant driver of total SOC loss in both tundra ($89 \pm 5\%$ under SSP 585) and boreal forests ($85 \pm 2\%$ under SSP 585).

Our study also distinguishes the underlying mechanisms contributing to these losses (Lines 427-434 and 439-446). In tundra soils, POC accumulation is primarily driven by cold temperatures and oxygen limitations that inhibit the decomposition of plant detritus, creating a high- f_{POC} environment that is highly vulnerable to microbial stimulation upon warming. In boreal forest soils, the high POC fraction is linked to the chemical recalcitrance of plant inputs (e.g., tannins and lignin) and a lack of aggregate protection, making this carbon pool particularly accessible to microbes as climate conditions shift. By identifying these biome-specific drivers while demonstrating the consistency of POC's dominant contribution across both systems, we provide a more nuanced and mechanically grounded foundation for our conclusions. The revised manuscript clarified that while the specific ecological drivers vary, the systemic vulnerability of POC remains a robust emergent property of high-latitude soils.

5. The authors correctly identify "Land Cover" as the most important predictor for POC and MAOC (Figure 2). However, their handling of future land cover is critically flawed. Their use of a global LUCC simulation product ignores a crucial feedback loop: climate change itself will directly and drastically alter high-latitude land cover. For instance, Arctic warming is causing boreal forest northward expansion ("greening"), shrub encroachment, wetland formation and degradation ("browning"), and other rapid biome shifts. These changes are not primarily driven by socioeconomic scenarios (SSPs) but are direct biophysical processes forced by climate. The LUCC product used likely reflects anthropogenic land-use change rather than capturing these climate-driven ecosystem transformations. Therefore, the "future tundra" upon which the model predictions are based may cease to exist or be replaced by other ecosystems by the end of the century. Using a static (or non-biophysically driven) future land cover map to predict soil carbon dynamics introduces potentially the largest unquantified error in the very regions identified as hotspots.

[Response] We thank the reviewer for this insightful critique regarding the biophysical feedbacks of land cover under climate change. We agree that climate-driven biome shifts—such as "greening" or shrub encroachment—introduce complexities not fully captured by socioeconomic-driven LUCC products. However, we contend that our core finding remains robust regardless of these specific land-cover transitions. While our projections are derived from a unified global modeling framework, our analysis reveals that POC remains the dominant contributor to SOC loss in both tundra soils ($89 \pm 5\%$) and boreal forest soils ($85 \pm 2\%$). This consistency across the two primary biomes involved in high-latitude shifts suggests that the high vulnerability of the POC pool is a systemic property of these cold-climate soils. Consequently, even if climate change triggers a "greening", the fundamental driver of soil carbon loss—the destabilization of weakly protected POC—remains unchanged. By providing this estimation of the potential carbon debt, we identify the risk that persists across shifting ecological boundaries.

6. The authors' conversion of projected POC loss to CO₂ emissions by simply applying the 44/12 ratio is an overly simplistic assumption that likely leads to a severe overestimation of the actual climate feedback. This calculation implicitly assumes that 100% of the "lost" POC is mineralized to CO₂. In reality, the decomposition of POC involves several competing pathways: A portion can be transformed into MAOC (the "microbial carbon pump" process). A portion can be lost as dissolved organic carbon (DOC) through leaching. Decomposition products can be utilized for microbial anabolism, forming new microbial necromass carbon. The authors briefly mention the importance of promoting POC transformation into MAOC in the discussion, but this admission highlights the fundamental flaw in their CO₂ calculation. If a significant fraction of POC is not directly converted to CO₂, their proclaimed figure of "equivalent to 2-3 times current annual emissions" is highly misleading. A responsible estimate must discuss the partitioning among these different fates and conduct sensitivity analyses, rather than adopting the most extreme and direct conversion assumption.

[Response] We thank the reviewer for this critique of the biogeochemical pathways following POC decomposition. We fully agree that the fate of "lost" POC is not solely mineralization to CO₂, but involves complex pathways, including transformation into stable MAOC, leaching as dissolved organic carbon, and incorporation into microbial biomass. Indeed, our modeling results already reflect these biogeochemical pathways. While we project substantial POC losses, we simultaneously observed a slight overall increase in MAOC ($0.12 \pm 0.01 \text{ kg m}^{-2}$ under SSP585; Fig. 3), suggesting that a portion of the carbon lost from the POC may indeed be sequestered into MAOC.

To ensure a responsible interpretation, we have refined the manuscript (Lines 464-467) to explicitly clarify that our CO₂-equivalent values are not intended as a definitive forecast of net annual greenhouse gas flux. Instead, they serve as an estimation of the "maximum potential carbon debt" and the inherent vulnerability of POC.

Reviewer#2' comments

The authors very well addressed my questions. I have no further comments.

[Response] We thank the reviewer for the positive feedback.

Reviewer#3' comments

After thoroughly reviewing the revised manuscript, I find that the authors have provided adequate revisions to fully address the comments from me and other reviewers. This is an interesting work and I'm happy to see it can be published soon.

There are some minor comments with the aim to further improve the manuscript.

[Response] We appreciate the reviewer's positive evaluation and helpful feedback. We have addressed the following minor points to further improve the manuscript.

Line 238: should be “our results identify high-latitude soils as global hotspots of...”

[Response] We thank the reviewer for the helpful suggestion and have revised the sentence accordingly.

Line 239: under what kinds of multiple climate scenarios? I think it should be more specific.

[Response] To improve clarity, we have specified the climate scenarios in the revised manuscript. The sentence has been revised as follows (Lines 238-239): “*We identify high-latitude soils as global hotspots of SOC vulnerability under shared socioeconomic pathway scenarios (SSP126, SSP245, and SSP585).*”

Line 241: should be “accounting for about $81 \pm 10\%$ of...”

[Response] The sentence has been revised accordingly.

Line 274-278: These two statements contain overlapping content and may be

streamlined for clarity.

[Response] The text has been revised to (Lines 265-264): *“Recent experimental evidence indicates that POC, particularly free POC, is considerably more sensitive to temperature changes than MAOC^{23, 24}.”*

Line 330: What does it mean high-latitude model?

[Response] To clarify, the sentence now reads (Lines 314-315): *“A random Forest model trained exclusively with high-latitude data indicates that”*

Line 417: controlling POC and MAOC stocks or changes?

[Response] We have clarified that “stock” is the correct term. The sentence now reads (Line 402): *“...controlling POC and MAOC stocks...”*

Line 497-498: “climatic sensitivity of POC losses” not clear

[Response] To improve clarity, we have revised the sentence to (Lines 474-475): *“Our results reveal substantial regional variation in projected POC losses under future climate scenarios.”*

Reviewer#4' comments

1. What are the noteworthy results?

The study shows that high-latitude soils represent global hotspots of soil organic carbon vulnerability, with future SOC losses largely driven by particulate organic carbon (POC) rather than mineral-associated organic carbon (MAOC). Under future climate scenarios, particularly SSP585, POC exhibits a stronger sensitivity to warming and accounts for the majority of projected SOC losses. The work further proposes the fraction of POC relative to total SOC (fPOC) as a functional indicator of SOC vulnerability to climate change, supported by multiple modeling approaches, sensitivity analyses, and regional assessments. Importantly, the study quantifies the potential climatic impact of POC losses, suggesting that neglecting this fraction may lead to underestimation of future CO₂ emissions in global carbon cycle models. The consistency of these results across different modeling frameworks and scales adds robustness to the main conclusions.

[Response] Thank you for your thoughtful comments. We appreciate your recognition of our main findings and your support for emphasizing the role of POC in soil carbon vulnerability.

2. Will the work be of significance to the field and related fields?

The work is of high significance for soil biogeochemistry, climate science, and carbon cycle modeling, as it provides an integrated assessment of the vulnerability of soil organic carbon fractions under climate change. While previous studies have recognized the functional distinction between particulate and mineral-associated SOC, most have focused on present-day distributions at regional scales or emphasized MAOC as the dominant stable carbon pool. This study advances the field by integrating an expanded global database, harmonizing diverse SOC fractionation methods within a common functional framework, and explicitly linking contemporary empirical controls to future projections under CMIP6 scenarios. By quantifying the contribution of POC to future SOC losses at the global scale and proposing its use as an indicator of climate vulnerability, the study substantially extends existing knowledge and offers clear implications for Earth system models and climate mitigation strategies.

[Response] Thank you for your thoughtful comments and for recognizing the significance of our work.

3. Does the work support the conclusions and claims, or is additional evidence needed? Overall, the conclusions and claims are well supported by the data and analyses presented, particularly following the revisions in response to reviewer comments. The robustness of the results is strengthened by extensive sensitivity analyses, including stratification by fractionation method, the use of regional models, and multiple independent algorithms, as well as by the explicit assessment of spatial and methodological uncertainties. The integration of biogeochemically informed modeling further supports the conceptual basis of the conclusions.

While both the reviewers and authors acknowledge limitations related to the availability of direct evidence on microbial mechanisms, POC subfractions, and deeper soil layers (>30 cm), especially in high-latitude regions, these gaps do not undermine the main findings. Rather, they highlight clear avenues for future research, which are appropriately recognized and discussed in the revised manuscript.

[Response] Thank you for your careful evaluation. We appreciate your recognition that our data and analyses support the conclusions, and we acknowledge the identified limitations as valuable directions for future research.

4. Are there any flaws in the data analysis, interpretation and conclusions? Do these prohibit publication or require revision?

No critical flaws were identified in the data analysis, interpretation, or conclusions that would preclude publication. The main weaknesses noted by the reviewers relate to methodological heterogeneity in POC and MAOC data, the relatively limited availability of observations in high-latitude regions, and assumptions associated with soil depth standardization and model stationarity. However, these issues are explicitly acknowledged, tested, and transparently discussed in the revised manuscript. Additional analyses indicate that the main patterns remain robust across different data subsets and methodological approaches, and the conclusions are carefully aligned with

the scope and limitations of the available data, avoiding unwarranted extrapolation. I believe that the revised content adequately addresses and supports the discussion of the identified limitations.

[Response] We sincerely appreciate your careful evaluation.

5. Is the methodology sound? Does the work meet the expected standards in your field?

Yes, the methodology is sound and meets the current standards of the field. Key strengths include the explicit functional harmonization of soil organic carbon fractions, soil depth standardization based on widely accepted approaches, and the use of multiple independent models combined with cross-validation. The study also explicitly incorporates uncertainty analyses, sensitivity tests, and regional modeling, and consistently integrates empirical methods with biogeochemically informed approaches, strengthening the conceptual basis of the inferences. While some simplifications are inevitable in global-scale studies, they are methodologically justified and transparently discussed, and do not compromise the validity of the results or conclusions.

[Response] Thank you for your positive assessment. We appreciate your recognition of the methodological rigor and the care taken to ensure robustness and transparency in our analyses.

6. Is there enough detail provided in the methods for the work to be reproduced?

Yes. Following the revisions requested by the reviewers, the procedures for data harmonization, soil depth standardization, predictor selection, and model training are described clearly and in sufficient detail. Data quality control and exclusion criteria, as well as modeling parameters and evaluation metrics, are explicitly reported. In addition, the data and code used in the analyses are publicly available, ensuring the reproducibility of the results. Overall, the study meets the expected standards of transparency and reproducibility for contemporary global syntheses.

[Response] We appreciate your acknowledgment of the clarity and reproducibility of our methods.

Overall, the revised manuscript presents robust results and a sound methodological framework. While some limitations remain, they are adequately addressed and do not compromise the main conclusions, and the study appears suitable for publication.

[Response] Thank you for your positive assessment and support of our manuscript.

Minor suggestion

In the Introduction, I suggest rephrasing the sentence in lines 274–278 to avoid redundancy, as the same idea is repeated in consecutive sentences.

[Response] To improve clarity, the overlapping statements have been streamlined. The text has been revised to (Lines 265-264): “*Recent experimental evidence indicates that POC, particularly free POC, is considerably more sensitive to temperature changes than MAOC^{23, 24}.*”