RESPONSE TO REVIEWER #3

Overall Comment

This manuscript presents a high-resolution (90 m) dataset of forest soil bulk density (BD) and pH across China. The authors have compiled a substantial database of 4,356 soil profiles from extensive field surveys and, using a machine learning model trained on these observations, generated multi-depth maps that outperform existing global products for Chinese forest ecosystems. While the compiled database is valuable and the authors have addressed part of the previous reviewer comments, I remain concerned about the novelty, methodology, and interpretation of results presented in this study. The most significant contribution is undoubtedly the compilation of 4,356 soil profile measurements; however, this dataset is not publicly available. Furthermore, the study offers no methodological or mechanistic advances, and the focus is limited to only two soil properties rather than a comprehensive soil database. These factors collectively diminish the overall scientific contribution of this work.

Response

Dear reviewer #3,

We sincerely thank the reviewer for this careful and constructive overall assessment. We acknowledge the concerns regarding data availability, the scope of the study, and the extent of methodological and mechanistic novelty. We agree that the compilation of 4,356 forest soil profiles is a major component of the present study. However, due to confidentiality requirements associated with the source project, the original point-based profile database cannot be made publicly available at this stage. We regret this limitation and have clarified it explicitly in the revised manuscript.

We also acknowledge that this study does not aim to propose a new machine-learning algorithm or a mechanistic soil process model. Rather, its main contribution lies in the development of a forest-specific, high-resolution, multi-depth dataset of BD and pH for China, based on an extensive and spatially representative forest soil profile database. In the revised manuscript, we have clarified this positioning more explicitly and reduced the methodological emphasis in the Introduction. We have also strengthened the scientific rationale for focusing on BD and pH by explaining their importance for forest carbon-stock estimation, soil acidification assessment, and evaluation of forest soil condition.

Regarding the scope of soil properties, we agree that a more comprehensive forest soil database would be highly valuable. The development of a broader forest soil digital soil mapping database is currently ongoing, and we appreciate this important suggestion. At the present stage, we focused on BD and pH because they represent foundational physical and chemical attributes of forest soils and support two important ecological applications at large spatial scales. We have revised the manuscript accordingly to better communicate this rationale and to more clearly define the scope and contribution of the present study.

We believe that these revisions substantially enhance the rigor, transparency, and interpretability of the manuscript. Our point-by-point responses to all comments are provided below, and all corresponding revisions are marked in **blue** in the revised manuscript.

Best regards,

Jizhen Chen

General Comments

General comments 1

The Introduction does not clearly articulate the scientific significance of this work. Excessive space is devoted to methodological details that properly belong in the Methods section. The authors fail to convincingly explain why a forest-specific dataset is scientifically urgent beyond improving accuracy metrics, or why BD and pH were prioritized over other key forest soil properties. I recommend refocusing the Introduction on the scientific gaps in forest soil research.

Response

Thank you for this insightful comment. We have substantially revised the Introduction to better articulate the scientific significance of this work and to refocus it on the ecological and pedological gaps in forest soil research ***(Line 38-54 and Line 70-90)***.

Specifically, we clarified why BD and pH were prioritized by emphasizing that these two properties represent the physical and chemical dimensions of forest soil condition, respectively, and are directly relevant to forest carbon-stock estimation, soil acidification assessment, and ecosystem functioning. We also strengthened the rationale for a forest-specific dataset by explaining that Chinese forest ecosystems span pronounced climatic gradients, complex topography, and highly diverse vegetation types, which together generate strong spatial heterogeneity in soil properties through vegetation–soil interactions, topography-driven variability, and distinct pedogenic processes. In addition, we revised the Introduction to highlight the broader scientific and practical value of large-scale BD and pH mapping for forest soil inventories, carbon accounting, and the identification of areas vulnerable to soil degradation and acidification. Finally, we reduced the methodological emphasis in the Introduction and retained only a concise description of the overall mapping framework, while removing detailed discussion that is more appropriate for the Methods section. As a result, the revised Introduction now places stronger emphasis on the scientific urgency and ecological significance of developing a high-resolution, forest-specific BD and pH dataset for China.

General Comments 2

The authors state that a variance inflation factor (VIF) threshold of <10 was applied to reduce multicollinearity among environmental predictors. However, substantial collinearity appears to persist among the final selected features, particularly climatic and biological variables. I recommend re-examining the collinearity diagnostics and providing more detailed technical information. Additionally, no vertical corrections appear to have been applied; this limitation should be acknowledged in the Discussion. Furthermore, the absence of depth-specific covariates (e.g., three-dimensional parent material data, groundwater information) may explain why climatic variables such as precipitation remain dominant drivers even in deeper soil layers (Fig. 7 & 8)—the model simply lacks better subsurface predictors.

Response

We sincerely thank the reviewer for this careful and constructive comment. In response, we have re-examined the collinearity diagnostics and clarified the procedure in the revised manuscript. Importantly, the VIF screening was conducted separately for each soil depth layer, rather than using a single pooled set of predictors across all layers. In other words, collinearity diagnostics and variable selection were performed independently within each depth interval to account for potential vertical differences in predictor relationships. We have now made this point explicit in the Methods and provide the VIF values of all retained predictors in ***Tables S8 and S9.*** All selected covariates satisfied the predefined criterion of VIF < 10.

We agree, however, that some climatic and biological variables may still show moderate correlations even after screening, especially because environmental predictors are often interrelated in large-scale soil mapping studies. We have therefore added more detailed technical information in the revised manuscript to improve transparency regarding the collinearity diagnostics and predictor selection procedure.

Regarding the reviewer's concern about vertical correction, we acknowledge that although predictor screening was conducted separately by soil layer, the model was still limited by the lack of depth-explicit subsurface covariates at the national scale. In particular, information such as three-dimensional parent material and groundwater-related variables was not available for inclusion. As the reviewer rightly noted, this limitation may have reduced the model's ability to capture deep-soil processes and may partly explain why climatic variables remained relatively important in deeper soil layers. We have now added this limitation to the Discussion ***(Section 4.3; Lines 604–606)***.

General Comments 3

The current evaluation scheme is biased in favor of the authors' dataset. It is unsurprising that estimates of BD and pH better reproduce observations used to train the model compared to global products interpolated from independent data. A fairer comparison would require an independent test dataset. I suggest compiling additional BD and pH measurements from published literature and re-evaluating model performance against this independent dataset.

Response

We sincerely thank the reviewer for this important and constructive suggestion. We fully agree that evaluation against a truly independent test dataset would provide a more rigorous and fairer basis for comparison.

Following this suggestion, we carefully screened the currently available public soil databases to identify an appropriate external dataset for validation. However, after detailed examination, we found that the currently accessible datasets were not ideal for this purpose. Specifically, after filtering the WoSIS database for forest bulk density observations in China, fewer than 20 usable samples remained, which we considered insufficient for a robust and representative evaluation. We also examined the recently published national soil dataset from China, but found substantial overlap between its source records and those used in previously published products such as CLDSv2 and ChinaSoilInfoGrids. As a result, using these datasets as an independent benchmark could still compromise the independence of the comparison.

Given these limitations, we considered that retaining the IV-based dataset comparison might be potentially misleading. Accordingly, we have removed the corresponding comparison and discussion from Section 4.1 in the revised manuscript. We have also noted this limitation in the manuscript and will seek to compile additional independent BD and pH observations from published literature and other reliable sources in future work to enable a more rigorous external evaluation.

General Comments 4

While the model interpretation is thorough, the Discussion would benefit from a deeper exploration of ecological drivers and consequences. Based on the SHAP analysis, the authors should explicitly discuss how driving factors of BD and pH in forest ecosystems differ from those reported in mixed-land-use studies. For example, how do key environmental features mechanistically regulate soil BD or pH in forests? Why might findings based on data from other land-use types not hold for forests? Emphasizing these forest-specific distinctions would substantially strengthen the scientific contribution. Moreover, Section 4.1 relies heavily on statistical metrics such as RMSE and MEC without analyzing the implications of observed differences. For instance, the finding that deep-soil (60–100 cm) BD (1.25 g/cm^3) is significantly lower than SoilGrids estimates (1.37 g/cm^3) is potentially important and should be discussed in terms of its impact on national-scale forest carbon stock estimates.

Response

Thank you for this insightful and constructive comment. We agree that the previous version of Section 4.1 emphasized model performance more than the ecological interpretation of the observed forest-specific patterns and their implications.

In response, we substantially revised ***Section 4.1*** by adding two new discussion paragraphs ***(Line 467-487 and Line 504-518)***. First, we now explicitly discuss the ecological mechanisms that may underlie the forest-specific controls of BD and pH, and explain why relationships derived from mixed-land-use datasets may not be directly transferable to forest ecosystems. In the revised text, we clarify that mixed land-use products may partly reflect anthropogenic influences that are less prevalent in forests, such as soil compaction associated with tillage and traffic, and pH modification caused by fertilization or liming. By contrast, our forest-specific results are discussed in relation to ecological processes that are more characteristic of forests, including litter input, organic matter turnover, drying–rewetting dynamics, root activity, clay illuviation, base-cation leaching, and acid inputs from litter decomposition and root exudation. We also explain how these mechanisms vary with depth and may lead to different controls on BD and pH in surface and deep soils.

Second, we now discuss the ecological consequences of the lower BD predicted for the 60–100 cm layer in terms of national-scale forest soil carbon stock estimation. Specifically, we state that, assuming other terms remain constant, using the SoilGrids BD estimate instead of our forest-specific estimate would increase the estimated soil mass, and thus the SOC stock, of the standard 60–100 cm layer by approximately 9.6%. At the same time, we added an important qualification that this implication should be interpreted cautiously because the standardized 60–100 cm interval does not always correspond to a fully developed and ecologically equivalent soil body in forest ecosystems, especially in mountainous regions where effective soil depth is constrained by topography, parent material, erosion, and shallow bedrock. We therefore emphasize that uncertainty in deep-soil carbon estimates arises from both BD and the actual soil volume represented by the standard layer.

We believe these revisions substantially strengthen the ecological discussion of forest-specific drivers and consequences, clarify why findings from mixed-land-use datasets may not hold for forests, and better explain both the significance and the uncertainty of the deep-soil BD difference for forest carbon stock assessment.

Minor Concerns

Minor concerns 1.

Lines 108–110: The dataset was supplemented by "independently conducted regional forest soil surveys" conducted in intervening years. However, no references or specific sources are provided for these regional surveys. Please list the relevant citations or data sources to ensure transparency and reproducibility of the data compilation.

Response

Thank you for pointing this out. We clarify that the regional surveys mentioned here were not derived from separately published external datasets, but were additional forest soil observations collected internally by our research team during the intervening years between the two nationwide campaigns. Therefore, no separate citations are available for these records. In the revised manuscript, we have clarified this point and added ***Table S2***, which summarizes the source information for these internal survey campaigns, including project name, survey year, geographic coverage, and number of profiles. The relevant text in ***Section 2.1.1*** has been revised accordingly ***(Line 109-110)***.

Minor concerns 2.

Lines 112–114: The description of quality control is insufficient for a high-impact data journal. Screening for "physically implausible values" based on "established reference ranges" is mentioned, but these ranges are not defined. Please specify the exact numerical thresholds used to filter BD and pH outliers (e.g., minimum and maximum allowable values). Additionally, clarify the "consistency checks" applied to geographic coordinates: did the authors explicitly verify whether sampling coordinates fall within the forest extent as defined by remote sensing covariates?

Response

Thank you for this important comment. We agree that the original description of the quality-control procedure was not sufficiently explicit. In the revised manuscript, we clarified that potential outliers were identified using the standard interquartile range (IQR) method, with values below $Q1 - 1.5 \times IQR$ or above $Q3 + 1.5 \times IQR$ flagged for manual inspection. In addition, we clarified the coordinate verification procedure by stating that all sampling points were overlaid with the forest mask and the environmental covariate rasters used for modelling, and that points outside the forest extent or located on NoData pixels were excluded from further analysis ***(Line113-119)***.

Minor concerns 3.

Lines 153–160: The description of environmental covariates relies excessively on URLs and organizational names (e.g., "National Tibetan Plateau Data Center," "Geological Map of China"), which is inadequate for a data journal. Formal bibliographic citations should be provided for every dataset used.

Response

Thank you for this comment. We agree that the description in the main text relied too heavily on URLs and organizational or platform names, which was not sufficiently clear for a data journal.

In the revised manuscript, we therefore streamlined this section by removing URL-style descriptions and replacing platform-based wording with formal dataset citations where appropriate. We also explicitly direct readers to Table S1, which provides the source datasets, spatial resolutions, and corresponding references for all environmental covariates used in this study. For variables derived from the same underlying datasets, we cite the original source datasets rather than repeating separate citations for each derived variable ***(Line 157-165)***.

Minor concerns 4.

Figure 3: The horizontal bars indicate significance using four asterisks (****), but the figure caption defines p-value thresholds only for one (*), two (**), and three (**) asterisks. Please update the caption or figure accordingly, and consider whether displaying such detailed significance levels is necessary.

Response

Thank you for pointing this out. We agree that the original significance annotation in Figure 3 was inconsistent with the caption and unnecessarily detailed. In the revised figure, we simplified the significance notation to three levels only (* $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$). The revised figure is now consistent with the caption (***Fig.3 in Line 275***).

Minor concerns 5.

Lines 283–289: The validation strategy employs random 10-fold cross-validation (CV). However, soil properties typically exhibit strong spatial autocorrelation. Random partitioning often results in training and validation samples being spatially adjacent, causing "information leakage" whereby the model overestimates performance by memorizing local spatial structures rather than learning relationships with environmental covariates. This concern is supported by the large discrepancy between CV results (MEC ~0.88) and independent validation (IV) results (MEC ~0.65), suggesting inflated CV accuracy. I strongly recommend implementing spatial cross-validation (e.g., spatial blocking or buffering) to obtain a more realistic assessment of model performance.

Response

We thank the reviewer for this important comment. We agree that spatial autocorrelation can potentially lead to optimistic estimates when random cross-validation is used, especially in spatial prediction studies.

In the present study, however, model performance was evaluated not only by 10-fold cross-validation, but also by an independent validation (IV) dataset. We consider the IV results to provide the more conservative and practically relevant assessment of predictive performance.

We acknowledge that spatial cross-validation (e.g., spatial blocking or buffering) could provide an additional perspective on model generalization under spatial dependence. However, implementing such a framework would require substantial restructuring of the current modeling and validation workflow and is beyond the scope of the present revision. Instead, we have added text in the Discussion to explicitly acknowledge that the random CV results may be somewhat optimistic due to potential spatial autocorrelation, and that future work should incorporate spatial cross-validation for a more rigorous assessment.

Accordingly, we respectfully did not implement spatial cross-validation in this revision, but we have clarified the limitation and strengthened the interpretation of the independent validation results in the revised manuscript ***(Line 565-568)***.

Minor concerns 6.

Lines 301–302: The predicted vertical pattern is described as "consistent with that observed in the field measurements," but no figure reference is provided. Please add a citation to "(Figure 3)" here so readers can readily verify the comparison between predictions and observations.

Response

Thank you for this helpful suggestion. We agree that figure citations should be provided where the predicted vertical patterns are compared with the observed field measurements. Accordingly, we revised the relevant sentence in the BD section by adding a reference to Fig. 3a ***(Line 305)***. We also made a similar revision in the pH section by adding a reference to Fig. 3b ***(Line 332)***, so that readers can more readily compare the predicted and observed vertical patterns..

Minor concerns 7.

Line 322: The order of the title is incorrect.

Response

Thank you for pointing this out. We agree that the title numbering was incorrect. We have corrected the section title from “3.5.2 Spatial patterns of pH” to “3.3.2 Spatial patterns of pH” in the revised manuscript.

Minor concerns 8.

Line 359: Non-English characters appear in the text.

Response

Thank you for pointing this out. We apologize for the unintended non-English/garbled characters introduced by cross-referencing in the previous manuscript version. In the revised manuscript, we have carefully checked the full text and corrected these issues throughout to ensure consistent and proper character encoding.

Minor concerns 9.

Lines 369–374: This section describes the SHAP methodology rather than presenting results, duplicating content from the Methods section. Please move this methodological description to Section 2.3, where SHAP is introduced, and focus the Results section directly on key findings derived from the SHAP plots.

Response

Thank you for this helpful comment. We agree that the original text in Lines 369–374 described the SHAP methodology rather than presenting results, and thus duplicated content from the Methods section. In response, we moved the methodological description of SHAP to Section 2.3 ***(Line 227-229)*** and revised the opening paragraph of Section 3.5 ***(Line 371-379)*** so that it now focuses directly on the key findings from the SHAP plots. The revised text highlights the dominant contribution of climate-related covariates and the depth-dependent differences in predictor importance for BD and pH.

Minor concerns 10.

Lines 455–515: While Section 4.1 quantifies statistical differences (RMSE reduction) between the authors' dataset and previous products (CSDLv2, SoilGrids), the explanation for these discrepancies remains largely qualitative and generic. The discussion attributes differences broadly to "forest-specific processes" (e.g., clay illuviation, acidification) without addressing where these differences are most pronounced or why. For example, does the 90 m resolution offer particular advantages in the complex terrain of Southwest China compared to coarser global products? The authors claim improved performance results from optimal covariate selection but provide no evidence to support this assertion. I recommend linking this claim to the SHAP analysis: are there specific environmental drivers critical for forest soils that are likely under-represented or smoothed out in general-purpose global models? The authors should deepen this section by analyzing the spatial pattern of differences, discussing environmental conditions (e.g., steep slopes, specific climate zones) where the model diverges most from others, and explaining likely causes based on feature importance analysis rather than relying solely on general pedological theories..

Response

Thank you for this insightful comment. We agree that the previous version of Section 4.1 relied too heavily on general pedological explanations and did not sufficiently connect the observed differences with their spatial expression or with the SHAP-based feature importance analysis.

In the revised manuscript, we substantially rewrote Section 4.1 to address this point in three ways ***(Line 470-487).*** First, we now explicitly link the discrepancy between our forest-specific product and existing general-purpose products to the SHAP results. The revised text emphasizes that BD and pH in forest soils are primarily associated with climatic and hydrological predictors, and that this pattern indicates a stronger role of climate- and vegetation-related controls in forest ecosystems than in broader mixed-land-use mapping frameworks. This provides direct support for our interpretation that forest-specific covariate selection contributed to the improved model performance.

Second, we added a more explicit discussion of where the divergence from existing datasets is most pronounced and under which environmental conditions it is more likely to occur. Based on the spatial comparison and regional statistics (Fig. 6 and Fig. S5), we now note that the divergence appears more evident in humid southern forests and mountainous regions of southwestern China, where strong precipitation gradients, steep terrain, and heterogeneous vegetation structure are likely to intensify local soil variability. We also clarified that, in such settings, the 90 m forest-specific product is better able to preserve local heterogeneity than coarser global products, which tend to smooth these patterns.

Third, we revised the ecological interpretation to move beyond generic statements about "forest-specific processes." The new discussion now explains that, in forests, soil properties are more tightly coupled to canopy-mediated water balance, litter inputs, root activity, leaching, and biologically mediated acidification, thereby making climate-, hydrology-, and vegetation-related controls more influential than the broader cross-ecosystem relationships captured in general-purpose products.

We believe these revisions directly address the reviewer's concern by linking the improved performance to SHAP-derived feature importance, by discussing the spatial and environmental contexts in which model divergence is most evident, and by providing a more forest-specific ecological interpretation of the observed differences.