Author Response to Reviewer Comments

1. Main Comments

Reviewer comment:

"This paper presents an approach to estimate soil bulk density (BD) from soil data, environmental data, and remote sensing data. Classical pedotransfer functions (PTF) are compared with ML methods that include also remote sensing data to predict soil bulk density. An important result of the study is that soil organic matter or soil organic carbon are the most important input variables of PTFs and PTFs that use only these input variables perform the best. As a consequence, these PTFs relate changes in BD over time only to changes in soil organic matter or carbon. ML approaches also include other variables that could be linked to land use and land management. This improves the prediction of BD compared to the classical PTFs. But, to what extent these extra variables influence the BD estimates depends on the type of method that is used. A difference in sensitivity to different variables affects how predictions of changes in BD respond to changes of input variables over time. Comparing the predicted distributions of BD in 2009 with those of 2004, it seems that both PTFs and ML methods predict similar changes, although there are some differences. The importance or relevance of these differences was not very clear. Furthermore, since no measurements of BD in 2009 were available, it was not possible to verify whether changes in BD were predicted more accurately using the ML method. Given this lack of validation, "the authors should give other evidence that demonstrates the additional value of the ML approach they propose. For example, can the differences between the changes in BD that are predicted by the ML and PTF approaches be related to independent information on management etc… ? What is the correlation between the changes in BD that are predicted by the two approaches? Since the change in BD is probably small compared to its spatial variability, it would be interesting to know whether the two approaches predict similar spatial patterns of the change and how these patterns of change are related to which input variables." I think this additional information is needed to give the paper more relevance."

Response:

We sincerely thank the reviewer for the insightful and constructive feedback, which has helped to improve the clarity and scientific depth of our manuscript. In response, we have carefully revised the manuscript to address the four key concerns as follows:

First, regarding the influence of input variable sensitivity on BD predictions over time, we expanded explanations in Sections 3.4 and 4.2. These revisions clarify how the ML model incorporates remote sensing and environmental variables (e.g., NDVI, BSI, slope, temperature), allowing it to capture BD variation associated with land surface dynamics, unlike classical PTFs which rely solely on organic carbon. **Second,** we acknowledge the limitation that in-situ BD measurements for 2009 are unavailable. This is addressed in Section 5 (Conclusion). To provide additional evidence of the added value of the ML approach, we revised Section 3.6 and Figure 8 to more clearly illustrate the prediction differences between the ANN model and PTFs. Notably, PTFs were unable to handle high organic carbon values, leading in some cases to negative or unrealistic BD estimates (Figure 8d). In contrast, the ANN model remained robust. We also included a correlation analysis of predicted BD between models (Figure 8c) to further visualize prediction differences. **Third,** as the 2004 and 2009 soil samples were collected at different geographic locations, it was not methodologically appropriate to compute point-wise BD changes (ΔBD) or correlations of change between models. Therefore, instead of direct spatial comparison, we evaluated the distributional characteristics of BD predictions in 2004 and 2009 (mean, SD, skewness) in Table 7 and Figure 7, and compared model robustness and sensitivity to input variability. These results confirmed that the ANN model yielded more stable and realistic outputs across years, while PTFs were more sensitive to extreme OC inputs. **Finally,** while spatial patterns of BD change between years could not be directly mapped due to the non-overlapping sample locations, we addressed the reviewer's concern by performing a temporal feature importance analysis (Section 3.6, Table 8). This comparison revealed shifts in predictor influence between 2004 and 2009, most notably, increases in the importance of rainfall and SAVI, and decreases in NDMI, slope, and MSAVI. This finding has also been incorporated into the abstract to highlight the ANN model's temporal sensitivity and broader interpretability relative to traditional PTFs.

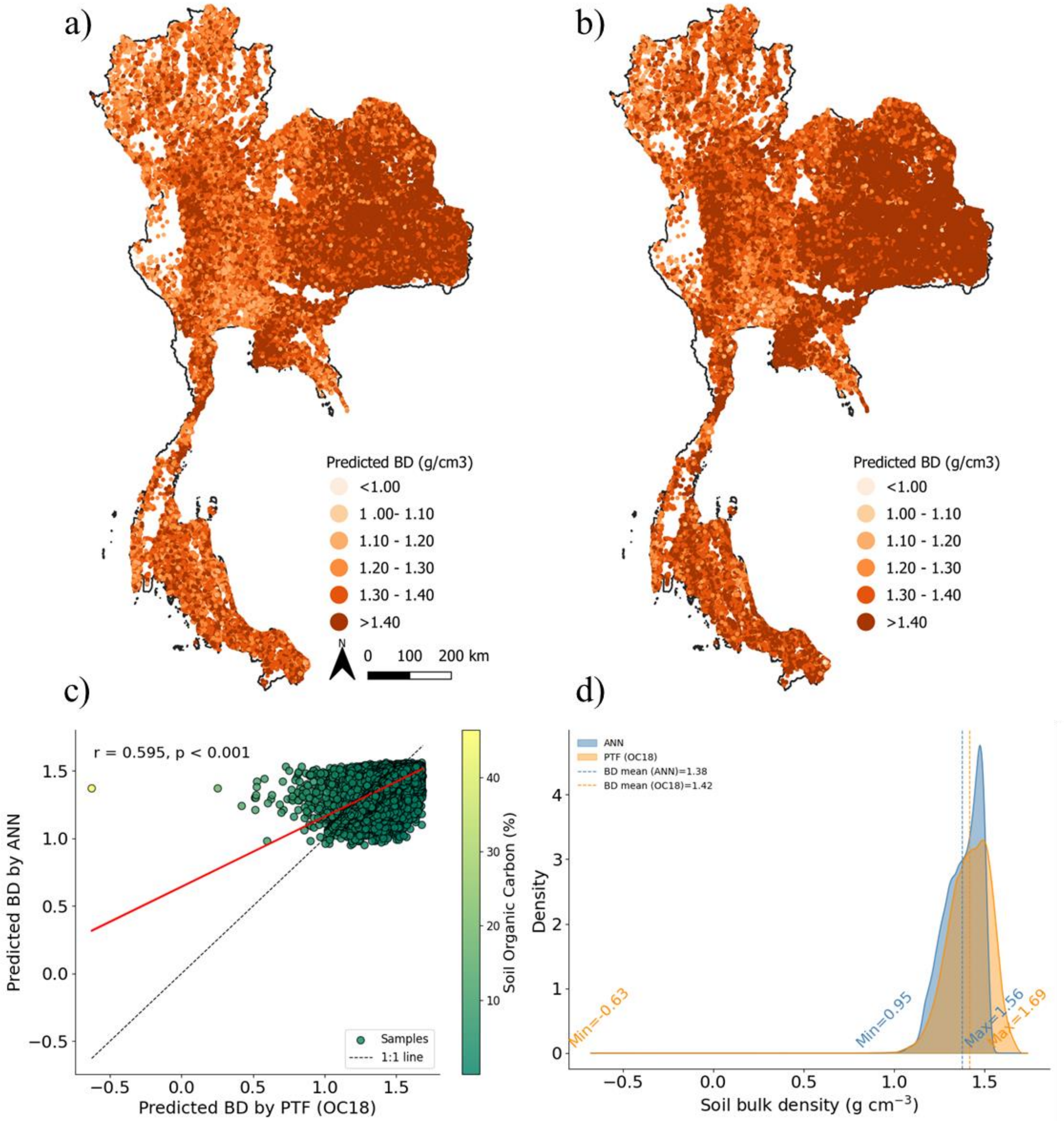


Figure 1 Comparison of soil bulk density (BD) predictions in 2009 using ANN with RS inputs (a) and PTF (OC18) (b), the Pearson correlation between ANN and PTF predictions (c), and the histogram of corresponding BD density distributions (d).

Table 1 Temporal comparison of relative importance (%) of predictor variables in ANN-Based BD estimation models in 2004 and 2009

No	Feature	Relative Importance in 2004 (%)	Relative Importance in 2009 (%)	Change (%)
1	BSI	7.16	6.81	-0.35
2	NDSoI	7.18	7.23	+0.05
3	NDVI	6.75	7.22	+0.47
4	NDMI	6.88	6.02	-0.86
5	SAVI	6.23	6.89	+0.66

6	OC	6.06	6.95	+0.89
7	DBSI	7.05	6.86	-0.19
8	temp	6.95	6.55	-0.40
9	elevation	6.43	6.51	+0.08
10	rainfall	5.95	7.51	+1.56
11	EVI	6.88	6.55	-0.33
12	MSAVI	6.98	6.44	-0.54
13	aspect	6.18	6.42	+0.24
14	CI	6.46	6.20	-0.26
15	slope	6.86	5.83	-1.03

2. Detailed Comments

Ln 26: 'Surface BD is dominated factor' Change to Surface BD is a dominating factor…
Response: Revised as suggested to "Surface BD is a dominating factor."

Ln 31: A reference would be needed here.
Response: We have added supporting references.

Ln 32 `Pedotransfer Functions (PTFs) have long been used to estimate BD by predicting soil properties based on readily available soil attributes.` This sentence has a strange structure. Skip: by predicting soil properties.
Response: We have revised.

Ln 50: 'In contrast, vis–NIR spectra from spectroscopy did not show significant differences in performance compared to PTFs-based models, but were still superior (Katuwal et al., 2020).' This is contradictory.
Response: We revised the sentence for clarity by deleting "but were still superior."

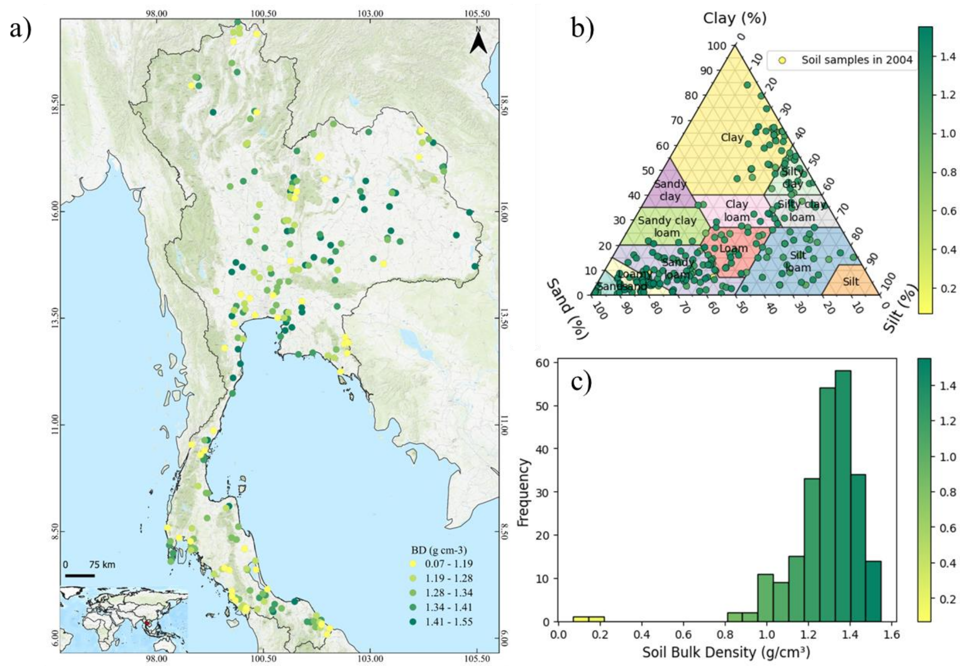
Ln 58: 'leading to issues such as overestimation' I think this is one specific issue but not a general problem.
Response: We agree with the reviewer and have revised the sentence to "…which in some cases may result in overestimation." for accuracy

Ln 83: 'Additionally, soil samples with OC data collected in 2009 were used for model implementation.' How many.
Response: We have clarified this in the manuscript by specifying that 76,089 soil samples with OC data were collected in 2009 and used for model implementation.

Ln 84: ‚These samples included measurements of‘ which samples? The ones collected in 2004 and in 2009?
Response: We have clarified this in the manuscript to specify that the measurements refer to the 236 soil samples collected in 2004 for model development.

Eqs 1 and 2 are nearly identical. Eq 1 can be skipped. Eq 3 is trivial and can be skipped as well.
Response: We have removed Eq. 1 and Eq. 3 from the manuscript, keeping only Eq. 2 for clarity.

Figure 1: the color scale of the histogram does not match with that in the figure.
Response: We have revised Figure 1 to ensure a consistent color scale across all panels (map, texture triangle, and histogram) and added a clear color bar.



Ln 83: temperate climate: shouldn’t it be tropical climate?
Response: We corrected to “tropical climate”

Ln 111: weighted median. Which weights were used?
Response: We clarified that no weighting was applied and used a pixel-wise median composite of all cloud-free images (January–December). The text has been revised by replacing “weighted median” with “median” for clarity.

Ln 135: 'Root Mean Square Error (RMSE)' with respect to what? The 2004 BD measurements?

Response: We confirm that RMSE was calculated with respect to the observed BD measurements from 2004 in order to validate the PTFs. The PTF with the lowest RMSE was then selected and applied to the 2009 dataset for comparison with the ML+RS model. To avoid confusion, we have revised the manuscript text to explicitly state that RMSE was calculated against the 2004 BD measurements.

Ln 270 'as no ground-truth BD measurements were available for validation in that year' The main purpose of the study was to investigate if the change in BD over time could be derived using the PTFs. To my understanding, that would require sampling of BD over time.

Response: We agree with the reviewer that the absence of ground-truth BD measurements in 2009 is a limitation of our study, as it prevents direct validation of temporal changes. To acknowledge this, we have revised the manuscript and added one sentence in the Conclusion section before the future work paragraph: "A key limitation of this study is the absence of ground-truth BD data in 2009, which restricted direct validation of temporal predictions." This addition makes the limitation explicit while emphasizing that future BD sampling will be necessary for validating long-term model predictions.

Ln 272: The 2009 dataset comprised 76,089 soil samples, containing OC percentages at a depth of 30 cm. Were sites where samples were taken in 2004 revisited in the 2009 campaign?

Response: We clarify that the 2004 and 2009 soil datasets were collected under different sampling campaigns and therefore do not correspond to exactly the same locations. However, both datasets cover the same soil series, soil texture groups, and land-use types, and were collected from sites located in close proximity wherever possible. This ensures a high level of comparability between the two campaigns despite the absence of exact site revisits.

Ln 286: If you want to investigate changes of a variable in time, you best observe the parameter at the same location. Then you do a paired t-test.

Response: We agree with the reviewer that paired sampling would have been the most robust approach for detecting temporal changes in BD. However, because the 2004 and 2009 datasets were collected from different locations under separate campaigns, a paired t-test was not possible. Instead, we used Welch's t-test (two independent samples with unequal variances), which is more appropriate for independent datasets with unequal sample sizes. We have clarified this point in the revised manuscript by updating Section 2.8 to emphasize why Welch's t-test was selected.

Ln 287 $\mu 2009$ and $\sigma 2004$ should be $\sigma 2009$ and $\sigma 2004$

Response: We have corrected.

Ln 321: In contrast, the poorest-performing model, PSOC8, exhibited an RMSE of 6.273 g cm^{-3}, highlighting significant predictive errors (Fig. 4). The RMSE is far beyond the maximal value of BD of soils. Can it be that wrong units for in- or output variables were used?

Response: we recheck unit again and we found that united correct but PSOC8 had OC and percentage of clay (cl) as predictor in the equation, and this cannot handle very high OC it will predict BD so high such as in inorganic soil type.

Ln 398: 'This increase may be attributed to factors such as intensified land management practices, reduced soil organic matter, or increased soil compaction over time.' It would be important to discuss how intensified land management practices and soil compaction are related to variables that are used as input in the ANN.

Response: Since Section 3.5 presents results, we kept it unchanged, but we revised Section 4.2 (second paragraph) to clarify the connection between land management practices and ANN input variables. Specifically, we added an explanation of how predictors such as NDVI, BSI, NDMI, slope, and temperature act as proxies for management-driven processes, including vegetation removal, bare soil exposure, soil compaction, and organic matter decomposition. This revision strengthens the link between remote sensing predictors and land management impacts on BD, making it clear how the ANN model captures management-related effects that PTFs relying only on OC cannot represent.

Ln 400: 'The minimum BD values also showed a substantial increase, rising from 0.12 g cm^{-3} in 2004 to 0.95 g cm^{-3} in 2009´ if not the same sites were visited, a comparison between the extremes is not very informative.

Response: We thank the reviewer for this valuable comment. We acknowledge the limitation that the soil samples from 2004 and 2009 were not collected at the exact same sites. However, both campaigns covered all soil types and were taken from nearby locations, ensuring broad comparability. While the results should be interpreted with caution, they still provide useful guidance for future research and allow us to observe general trends. We have noted this limitation in the conclusion section for clarity.

Ln 407: This transformation suggests a reduction in the occurrence of extreme BD values and a more balanced distribution of BD by 2009. See my comment above.

Response: We thank the reviewer for this follow-up. As noted in our response to Ln 400, we recognize that comparisons of extreme values should be interpreted with caution since the sampling sites in 2004 and 2009 were not identical. To address this, we have acknowledged this limitation in the conclusion section. We therefore retained the description in Section 3.5 as part of the reported results but clarified its interpretation in conclusion.

Ln 447 'Our findings show that OC-based PTFs, while exhibiting strong alignment with the RS-ANN model in mean BD values, displayed higher variability and greater prediction uncertainty, particularly in regions with fluctuating organic matter content' Where is that shown?
Response: We have clarified the text in Section 4.1 by explicitly pointing to Figure 7 and Table 7, which demonstrate the higher variability and uncertainty of OC-based PTFs compared to the ANN model.

Ln 458 where extreme BD values can lead. Do you mean extreme OM?
Response: We have corrected the text to state that the issue arises from extreme OM values leading to unreliable BD estimates, not extreme BD values.

Ln551: 4) 'Skewness and kurtosis analyses revealed that the RS-ANN model improved from a highly skewed distribution in 2004 (skewness = -2.81, kurtosis = 15.37) to a more balanced distribution in 2009 (skewness = -0.58, kurtosis = -0.41).' If I understand it correctly, this is the skewness of the distribution of predicted BD, and not of the distribution of the difference between observed and predicted BD. It is interesting to note that the ANN and PTFs predict a different distribution
Response: We confirm that the reported skewness and kurtosis values refer to the distribution of predicted BD values, not residuals. To avoid confusion, we revised the text to explicitly state that these are the skewness and kurtosis of predicted BD.

Ln 553: 'In contrast, PTFs continued to show high skewness and kurtosis, indicating persistent prediction errors for outliers.' This statement is not in line with the results shown in table 7.
Response: We revised the text to more accurately reflect Table 7. The revised sentence now states that PTFs showed higher skewness and kurtosis than ANN in 2004, but that the distributions became more comparable in 2009, avoiding the previous overstatement about persistent prediction errors.

Ln 555: ´5) The RS-ANN model demonstrated broader applicability across diverse soil types and land uses compared to traditional PTFs and OC-dominant ML models.´ Where is this shown?
Response: We have revised the sentence to explicitly reference the supporting evidence. The broader robustness of the ANN model is demonstrated by the feature importance analysis (Figure 6), which shows its use of multiple predictors (NDVI, NDMI, slope, temperature) compared to the OC-dominant PTFs.

We look forward to hearing from you in due time regarding our submission and to respond to any further questions and comments you may have.

Sincerely,
Sunantha Ousaha
2 October 2025