Reply to Reviewer #2:

General comments

Based on global-scale data, this study considered the use of integrated machine learning models to investigate the estimation of daily soil moisture (SSM). The results of the study demonstrate that the integrated machine learning algorithm outperforms both the optimization and base machine learning algorithms in predicting SSM. The topic of SSM estimation is of significant importance. I have the following suggestions and questions to further enhance the current manuscript:

Thanks for the kind words! Your comments are immensely valuable in enhancing the manuscript's quality.

Specific comments

1. I would suggest including a literature review on data selection in the introduction, while reserving the Data section solely for describing the data used and the data preprocessing.

Reply 1: Thank you for bringing up this important point! The selection of features is an integral part of the ML methodology (relevance of the features/variables, length of data, structure of ML algorithms and their tuning), the data/features selection is based on previous works by others and our own insights (will be demonstrated in the feature importance figures). In this case, we will reorganize the section 2, by renaming "2. Data" to "2. Physical Features and Data". The secondary title will be:

- 2.1 Physical Features Selection
- 2.2 Data Source
- 2.3 Data Pre-processing
- 2.4 Data Split.

Then we will move the feature selection part from lines 109-112, 120-124, 131-138, 143-155, 163-171 to section "2.1 Physical Features Selection". In the beginning of section 2.1, we will add: "In order to predict SSM accurately, a multidimensional understanding of its complex dynamics requires a comprehensive integration of diverse environmental factors. While remote sensing techniques and advanced machine learning algorithms have revolutionized SSM estimation, the optimal selection of predictor variables remains a pivotal challenge. The dynamic interplay of precipitation, evaporation, land surface temperature (LST), vegetation index, soil properties, and topographic indices influences SSM patterns." to get an overview before explaining every predictor variable.

In the beginning of section 2.2, we will mention "Based on consideration of these physical features that influence that dynamics of SSM, we next describe the data used for training and testing the different algorithms."

2. This study investigates the utilization of integrated machine learning models. However, it raises the question of whether three base models in the integrated model are optimal. Have the authors considered the possibility of adjusting the number of models, such as exploring whether a model integrated with two machine learning models may yield better performance?

Reply 2: Thanks for your great suggestion. We built the ensemble models from three MLs to derive more robust and stable predictions than that of a single algorithm. It was an experimental design decision taken so that each ensemble structure is based on the same number of base models. The optimization of the number of the models included in the ensemble model is out of the scope of this manuscript but an interesting topic for a future study, thank you for highlighting it.

3. Page 3, Line 80: "(ii)justify … model". I would suggest adding experimental results or analyses in this area, such as manipulating the effect of a predictor on the results by increasing or decreasing its influence.

Reply 3: Thanks for your advice. In response to the comment regarding the justification of the model choice (ii) on page 3, line 80, we will further expand our analysis to provide additional insights into the rationale behind our selection. We plan to conduct a comprehensive feature importance analysis across different regression models, aiming to assess the impact of each predictor on the model outcomes. Specifically, the KNR+RFR+XB ensemble model performed the best, therefore we plan to compute the feature importance based on these three MLs, and then we calculate the mean feature importance of all features for these three MLs with permutation importance method (Li et al. 2021).

We believe by doing this experiment, we will observe the feature importance across different ML models and the overall feature importance. Then we can conclude that which input variables contribute to predicting the target variable. The result of this experiment will be added in the supplementary information.

4. Page 12, Line 255-271. The authors chose eight machine learning models based on their popularity and performance, but it appears that there is a significant amount of research applying artificial neural networks (ANN) or using them as a baseline (Uthayakumar et al., 2022; Senyurek et al., 2020; Liu et al., 2020...). Surprisingly, the authors did not include ANN in their selection.

Reply 4: Thank you for your perceptive observation and thoughtful suggestion regarding our choice of machine learning models. In our study, we selected eight machine learning models based on their established popularity and documented performance within the literature. While it is true that artificial neural networks (ANN) have garnered significant attention and application in various studies, we did incorporate the most simple ANN, namely the multilayer perceptron regressor (MLPR), into our study, in order to identify future lines of research. We appreciate your insights and the incorporation of additional ANN algorithms is an interesting topic for a future research.

5. The manuscript is too long and needs to be reduced, and the authors of the figures and tables need to be revised and optimized.

Reply 5: Thanks for your advice. In the revised manuscript, we already moved table 2 to supplementary information, it was changed to table S1. We also moved Figure 4 into supplementary information, it was changed to Figure S4. The r score column/columns were removed from Table 3,4,5, also decreased their size. In the end, all tables have been reformatted to increase readability.

References

Li, W., Migliavacca, M., Forkel, M., Walther, S., Reichstein, M., & Orth, R. Revisiting global vegetation controls using multi‐layer soil moisture. *Geophysical Research Letters, 48*, e2021GL092856, 2021