

Dear editor,

Thank you for your comments and suggestions. First of all, we are so sorry for taking so long time to reply to you. We are also aware of many problems in our manuscript. We have tried our best to revise our manuscript according to the comments. Attached please find the revised version, which we would like to submit for your kind consideration. We have written a point-by-point response letter for all reviewers. To make the reply more visible, Q represents questions raised by reviewers, and A are our answers for these questions. Below are your questions and suggestions:

(1) 69 there is no such thing as "inefficient samples". Do you mean "insufficient samples"?

A1: Thank you for your suggestion. It should be "inefficient sampling". It mainly refers to less scientific or reasonable sample sets. (Line 74)

(2) 83 unclear expression "Considering the role of SM in the ecological environment" -- "for ecology" or "for the environment"

A2: Sorry for the unclear expression. The "ecological environment" was changed as "terrestrial ecosystems". (Line 95)

(3) 1. 96 add reference to Figure 1. All figures must be referred to in the text. (this reference only appears in line 136)

A3: Thank you very much for your suggestions. We have added it in the Section 2.1 (Lines 109 and 110). Please see the revised version.

(4) general comment, page 4: it would be helpful to add a figure showing the precipitation climatology of the region and its variability rather than only showing the topography. This would allow readers to evaluate the geographic patterns of the SM results.

A4: Thank you very much for your suggestions. A figure showing the annual mean precipitation is added in Figure 1. Please see the revised version.

(5) 144 following: to facilitate comparison of dataset resolutions, please convert all resolutions to km. Example: "It has a spatial resolution of 0.25 degrees (~28 km)". Also use the degree symbol consistently.

A5: Thank you very much for your suggestions. This modification is implemented in the full text. Please see the revised version.

(6) sections 2.2.2 to 2.2.4: please include more details about the datasets, such as spatial resolution, time span covered and hints about the evaluation of their quality.

A6: Thank you very much for your suggestions. We added some content about these products, and the temporal and spatial resolutions of these products are described in detail. Please see the revised version.

(7) Figure 2: what is a "set dataset"? I only know the terms "validation dataset" and "test dataset"

A7: Thank you very much for your suggestions. It is described wrong, and it should be training set and test set. We have modified it. Please see the revised version.

(8) section 2.3.2: please define the target data product more more clearly. The section is named "downscaling", but the text describes a coarsening of resolution (to 36 km) and a functional mapping/regression from predictor variables to the predictant. What is the output resolution and how is this \*downscaled\*?

A8: Thank you very much for your suggestions. The downscaling process is described in more detail in the revised version. The process is divided into: the construction of 16-day regression model, prediction of 1-km daily SM product. Please refer to the revised version.

(9) 202: how was the data split done? Given that there is substantial memory in the system and some data are used with a temporal resolution of 16 days, a simple random sampling approach is invalid as it will lead to overoptimistic validation results. This issue is discussed for example in Schultz et al., 2021, <https://royalsocietypublishing.org/doi/10.1098/rsta.2020.0097>.

A9: Thank you very much for your suggestions. Based on the previous research (<https://doi.org/10.1016/j.rse.2014.02.015>; <https://doi.org/10.1016/j.rse.2019.111261>; <https://doi.org/10.1016/j.rse.2021.112646> ), we have improved this work. Stratified sampling considering the date of sample acquisition within 16 days was applied to split the training and test sets which means the number of training and test samples taken each day is equal. In addition, to ensure that the training and test sets of stratified sampling do not differ too much, the Kolmogorov-Smirnov (KS) test is adopted to test the distribution consistency of them. Based on this new samples, the model was re-run, and the results throughout the article were subject to change. Please see the revised manuscript for all revisions.

(10) section 3.2: please add more quantitative evaluation results. There are many statements like "good" or "poor", but it remains unclear according to which target criterion these quality indicators are given. For example, what is an acceptable R2 value in your view and why?

A10: Thank you very much for your suggestions. We have made a lot of modifications to Section 3.2. First, more quantitative evaluation results was added (Line 317). Second, in order to verify more scientifically, a figure (Figure 8: Comparison of gridded products and in situ observation SM of the Maqu Network) was added, and it shows that our simulation results have better accuracy compared to other products. Third, about R, we added the content "R greater than 0.6 and 0.8 indicates there is a reliable and strong correlation (Akoglu, 2018)" in Section 3.1 (Line 322). Fourth, there are some problems with the Figure 5, and The R values of the test set are not all greater than 0.8. The new results are all from new simulation results.

(11) l. 303: language. What does this mean "The average SM of the ERA5 products is polarised"?

A11: Sorry for my unscientific expression. What we mainly want to express is that: The average SM is significantly overestimated in the southern part of the study area, and underestimated in some areas in the northern of the study area. We have made some modifications (Line 431).

(12) section 4.1: be aware that there is a fine difference between the influence of a physical quantity on the target variable and the importance of input variables for the model regression. The latter can depend on the model architecture, data preprocessing and normalisation and other factors which have nothing to do with the cause-effect relationships in the real world. The text is ok as is, but it may help readers to assess the relevance of your results if you alert them to this point.

A12: Thank you very much for your suggestions. Due to the variables obtained in this study come from multiple data sources, data preprocessing can make errors, which I have added it in section 4.1 (Line 473).

(13) l. 350: The sentence "The simulation results of long time series will inevitably suffer the interference of various noises." is meaningless and not based on a sound mathematical concept from statistics.

A13: Thank you very much for your suggestions. It has been modified.

(14) l. 356 remove "even" before "smaller"

A14: Thank you very much for your suggestions. It has been deleted.

(15) general comment on section 4.2 and also methods section: your statements about the quality of individual ML models are not supported by the evidence shown in the paper.

A15: Thank you very much for your suggestions. We have made some modifications. The results of RF and XGB were compared with the combined result (Table 4 and Figure 8), and the combined result is better compared to a single model. Please see the revised manuscript for all revisions.

(16) Information is missing about the detailed set-up of these ML models. Otherwise you might be comparing apples and oranges, for example if a very small ANN is compared to an RF with many trees and branches. Have you done any ablation studies to find the optimum set-up for each ML method? Furthermore, the discussion would benefit from a somewhat more elaborate reflection about why a certain ML model performs better or worse for certain seasons, groups of stations, etc.

A16: Thank you very much for your suggestions. We tuned the hyperparameters in the process of doing this work, but this manuscript did not describe in detail. In the new version, we describe the parameter tuning process in more detail. Please refer to Attachment (Part 2: Hyperparameter Optimization Results).

In addition, regarding the difference in seasons, we think that it is mainly affected by the sample size (Line 344), and we also added discussion in the new version (Line 529).

(17) section 4.3 How do these results relate to the variable importance analysis presented in Figure 12?

The variable importance analysis is mainly for the predictors, and their can reflect the status of SM. Meteorological factors are the driving factors for SM. We have expressions in the new version. (Line 513)

Finally, thanks a lot for your careful review and invaluable advices. The version I trust again can meet your requirements. If that is not enough, seven of your suggestions; if possible, I am eager for this manuscript to be re-reviewed. Looking forward to the opportunity to learn from you! I also made some other revisions, please refer to other review results.

Looking forward to your next suggestions. Thank you!

Sincerely!

Pinzeng Rao

Address: Beijing, China.

Email: 578264905@qq.com