Note: The modifications are shown in green. The responses to comments are blue colored.

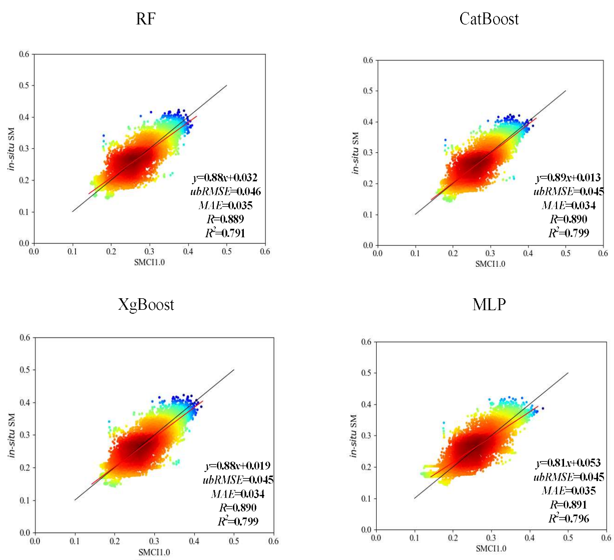
We are very grateful to the reviewer for reviewing the paper so carefully. These comments are very helpful to improve the quality of the manuscript. Please find our itemized responses below and our revisions will be in the revised manuscript.

Comment#1: In this study, only the Random Forest method was applied to derive the upscaled soil moisture data. Is it possible to try more MLs, such as CatBoost, XgBoost, and NeuralNetwork, to test how consistent or different the resulted products are?

Responds：

According to previous studies introduced in the introduction, RF models have proved to be successful in producing SM data and its computation time is acceptable. So, we choose this model in our studies.

According to this comment, we have also tested three more ML models (CatBoost, XgBoost, and Multilayer Perception) and found that the performance of these models is very close to RF. We applied the *scikit-learn* tools to evaluate them. We optimized the model parameters as follows. For the CatBoost model, we used the default parameters in catboost library. For the XgBoost model, we optimize the XgBoost by tuning the parameters through the *selection.GridSearchCV* function which is provided in the *scikit-learn tools*, we set *learning_rate* being 5.0e-3, *max_depth* being 6, *randam_state* being 42, and *estimators* being 100. For the Multilayer Perception, we set *random_state* being 1 and *max_iter* being 500.Take the year-to-year experiment at 10 cm soil depth as an example (see the figure below), RF achieved *ubRMSE* being 0.046, *MAE* being 0.035, *R* being 0.889, R^2 being 0.791. The *ubRMSE*, *MAE*, *R*, R^2 of CatBoost were 0.045, 0.034, 0.890 and 0.799, respectively. XgBoost achieved *ubRMSE* being 0.045, *MAE* being 0.034, *R* being 0.890, R^2 being 0.799. The *ubRMSE*, *MAE*, *R*, R^2 of ANN-based model were 0.045, 0.035, 0.891 and 0.796, respectively. According to these results, we don't think it is necessary to use other models instead of RF to produce the high-resolution SM as they cann't outperform it much.



In the modified manuscript, we have added the explanation why only RF model was shown in the discussion. The new expression is as follows:

It was necessary to note that we also compared the RF model with other ML models, including

CatBoost (Dorogush et al. 2018), XgBoost (Chen et al. 2016), and Neural Network (Rosenblatt et al. 1958) based models. We found that the performance of these models is very similar to RF models with a R2 around 0.79. In addition, RF has been widely applied and recognized in SM prediction and many other fields (Carranza et al. 2021, Lin et al. 2022, Ly et al. 2021) and it does not take too much computing time to make the predictions for the whole China. Hence, we only took RF model to produce the high-resolution SM data.

Comment#2: Most of the source datasets cover the period before year 2010. Is there any special reason why the new soil moisture only covers the period 2010-2020? Is it possible to extend the present time period to year 2000-2020?

Responds：

Thanks for your kind comments and helpful suggestions, although most of the applied covariates cover the period before year 2010, we do not access to the *in-situ* measurements before 2010, currently. However, the in-situ measurements before 2010 may be available from China Meteorological Administration (not open to us) and the number of stations is less than 800. If we produce the SM data set without any in-situ data (or only a few hundred stations), the quality of the data may be poorer as it will be extrapolation in time. However, we agree that it is still possible to extend the present time period to year 2000-2020 or even before. So, we list it as a future work in the conclusion as follows:
It is also possible to extent the time coverage of this data set before 2010, even though we do not have access to the in-situ data before 2010 now and the available in-situ stations may be less than 800, which will lead to poorer quality if not enough in-situ data are used.

Comment#3: The "Materials and Methods" read too long, and the authors may try to shorten the text and put some figures into the Supplementary Material.

Responds：

According to this comment, we have put the Figure 2 and related text into the Supplementary Material. We also shortened the text in this section as follows.

In section 2.1, we combined the following text with the last paragraph of this section:
"The other aspect is the bias and standard deviation correction of in-situ SM, which is vital for our study to allow the ML model to achieve the high-quality SM product. We applied the same correcting method with that of Sungmin et al. (2020), who adjusted the raw in-situ SM observations to match means and standard deviation of the ERA5-Land gridded SM data at the corresponding time periods, grid cells and layers." (deleted)
The last paragraph of this section:
After the above data processing, we started to perform the correction of deviation and variance for in-situ SM, which is vital for our study to allow the ML model to achieve the high-quality SM product. In-situ SM data was obtained by various sensor types, which had different calibrations. Hence, to overcome the artifacts during the RF model training, we adjusted the observations to match means and standard deviation of the ERA5-Land SM at the corresponding time periods, grid cells and layers using the same method with that of Sungmin and Orth (2020). This method made

the target in-situ SM resemble the mean and standard deviation of ERA5-Land SM, and kept daily temporal variations which follow the original in-situ SM time series. As the soil depth of each soil layer of ERA5-Land SM was inconsistent with that of in-situ SM, we mapped the soil layer of ERA5-Land SM to the corresponding soil layers of in-situ SM. Hence, the in-situ SM from 10 cm to 30 cm were adjusted based on the gridded SM at layer2 from ERA5-Land dataset (7-28 cm), and the in-situ SM from 30 cm to 100 cm were adjusted based on the gridded SM at layer3 from ERA5-Land dataset (28-100 cm).

We have alsoput the Table 2 into the Supplementary Material and shortened the related paragraph as follows:

The number of random selected candidate variables from all the covariates (max_features) and the value for the minimum node size (min_samples_leaf) in RF model are the vital hyper-parameters which affect the performance. Other hyper-parameters, such as number of trees (n_estimators), were not tuned but simply determined based on RF's own training. Meanwhile, we applied the 10-fold cross-validation method to tune the values of max_features and min_samples_leaf, and they were selected from range [1,25] with a single interval and [5,30] with 5 intervals via grid hyper-parameters method for preventing RF model over-fitting. The accuracy of RF models with all hyper-parameters based on grid hyper-parameters method at 10 cm soil depth were shown in Table 1S. We could see that the root means square error (RMSE) obtained based on all the hyper-parameters ranged from 0.601 to 0.637 and the best accuracy (RMSE=0.601) can be achieved when max_features and min_samples_leaf set to be 1 and 20, respectively, which were used for further research.

We also revised shortened other contents. This will be shown in the revised manuscript.

Comment#4: For the "Results", they seem to be a combination of results analysis and short discussion. Please move relevant discussion content to the "Discussion" part.

Responds：

Thanks for your kind comments and helpful suggestions, we have moved Section 3.4 in the old manuscript to the discussion as Section 4.1, and put some short discussions in the results of the old manuscript into the "Discussion" part. The new section in the discussion is as follows:

Figure 2 and 2s shows that the result at 70 cm and 90 cm were significant worse than those at other depths. The reason may be that RF model is difficult to estimate accurate SM for only a few in-situ SM stations. From Fig. S1 (b), we can see that the total numbers of data at 70 cm and 90 cm soil depths are quite small. In other words, more abundant of data were expected to help RF model 'learn' complete relationship between covariates and in-situ SM and further improve the quality of high-resolution SM in China. Meanwhile, compared with the previous study of Sungmin et al. (2020), our SMCI1.0 showed the superior quality (Figure 4-6), because the larger numbers of in-situ SM data in China were applied for RF modelling.

From Figure 5, during the rainfall near 91th day across the Tropical Monsoon Climate zone (Am) and near 1st day across the Snow climate with dry winter zone (Dw), the in-situ SM did not increase with high precipitation, but the SMCI1.0 product could capture the increase in SM (denoted in the light blue rectangle). The reason may be that the applied covariates had bias with in-situ measurement and further affected estimation by RF model. Meanwhile, we also found the RF model could overcome much bias in dry conditions, except for that from 196th to 305th days in the snow climate, fully humid zone (shown in the light red rectangle). In the case of 30 cm soil depth (Fig. S5), we could see an agreement between several peak events, it could be attributed to the soil texture homogeneity at the 10 and 30 cm soil depths. Almost all climatic regions had lower dynamic ranges at 30 cm soil depth than that at 10 cm, this may be attributed to the persistent behaviour of SM at 30 cm soil depth. In the case of 30 cm soil depth in Fig. S6, the SMCI1.0 product had higher accuracy than that at 10 cm soil depth (Figure 6), especially in terms of ubRMSE and MAE metrics. The reason may be the background aridity led to low variability of SM in the deeper layers (Karthikeyan and Mishra 2021) and the RF model can capture the variation in SM easier.

Interestingly, it was inconsistent for the results of R, ubRMSE, and MAE in Fig. 2 and Fig. 4, which had the same phenomenon with the previous study (Sungmin and Orth 2020) (represented in their Fig. 4 and Fig. 5). For example, SMCI1.0 product had the ubRMSE, MAE and R being 0.046, 0.035 and 0.889 at 10 cm soil depth in Fig. 2. However, in Fig. 4, the box-plot represented the lowest ubRMSE, MAE and highest R of SMCI1.0 product were nearly 0.03, 0.02, and 0.7, respectively. The reason may be that the same metrics were calculated in different ways, the one in Fig. 2 was to count the results of all stations and temporal period, and the one in Fig. 4 was to count the results of only temporal period at one station.

It was necessary to note that we also compared the RF model with other ML models, including CatBoost (Dorogush et al. 2018), XgBoost (Chen et al. 2016), and Neural Network (Rosenblatt et al. 1958) based models. We found that the performance of these models is very similar to RF models with a R2 around 0.79. In addition, RF has been widely applied and recognized in SM prediction and many other fields (Carranza et al. 2021, Lin et al. 2022, Ly et al. 2021) and it does not take too much computing time to make the predictions for the whole China. Hence, we only took RF model to produce the high-resolution SM data.

Comment#5: The "Discussion" really needs to be reorganized and improved; the current one does not provide deep thoughts on the new soil moisture products, in terms of their differences/similarities/uniqueness compared to previous products/work and implications for the soil moisture modeling and detection and attribution.

Responds:

Thanks for your kind comments and helpful suggestions, we first removed the Section 4.1 in the old manuscript and put the related text into the "Conclusions" part. The new expression is as follows:
In this study, the gridded soil moisture was estimated through RF method in China based on the ERA5-Land reanalysis, USGS land cover type and DEM, reprocessed LAI and soil properties from CSDL, which included soil depths from 10cm to 100cm and had 1km spatial and daily temporal resolution over the period from 1 January 2010 to 31 December 2020.

We set "Sensitivity to precipitation, air temperature and radiation" as Section 4.2, as it is close to the new Section 4.1. We set Section 4.3 as "Factors affecting the quality of SMCI1.0". We combined the original Section 4.3 and 4.4 as the new Section 4.4 "Requirement of further validations and improvements".

In addition, we have added the Section 4.5 providing some thoughts on our product about implications for the soil moisture modeling and attribution, meanwhile, in this section, we have also added the discussion about comparison between our product and previous products. The new expression is as follows:

In this section, we mainly discussed the comparison between SMCI1.0 and previous products, and the implications for the soil moisture modeling and attribution. From the previous results in Section 3, we can see that SMCI1.0 generally outperforms the existing SM products (ERA5-Land, SoMo.ml and SMAP-L4) at most cases. The most important uniqueness of SMCI1.0 is taking the in-situ SM data as the training target with abundant sample size. Even though we used the ERA5-land to correct their means and standard deviation at each site, the temporal variation still came from the observations. We have also tested to train the RF model with the original SM observations and found that the performance of the model decreased dramatically with a R2 of 0.67 compared to the model with correction (a R2 of 0.79). And more importantly, the resulting SM maps demonstrated unreasonable noisy spatial distribution. These indicates that the in-situ SM in China have essential data inconsistency and the correction according to ERA5-Land is necessary which has physical consistency. Furthermore, SMCI1.0 is provided with relatively high spatial and temporal resolution (1-km and daily) for ten soil depths, which makes it possible for wider applications at finer scales and deep soils for the whole China, while reanalysis and remote sensing SM data are often at coarser resolution and remote sensing SM data are only for the surface soil.

However, SMCI1.0 estimated by machine learning model cannot always reflect the variation of SM well, especially for some extreme events or so called "tipping points" (Bury et al. 2021). From Fig,5, we can see that SMCI1.0 deviated from the in situ SM in some cases, though this also happed to the other three SM products. For example, from 35th day to 61th day across the Snow climate, fully humid (Df), SMCI1.0 and SoMo.ml overestimated, while SMAP_L4 underestimated. "Tipping points" denoted that slowly changing SM sparks a sudden shift to a new (Bury et al. 2021). This is a huge challenge for estimating in-situ SM by ML models, because "tipping points" make the dynamics of complex system simplify down to the limited number of possible "normal forms" (Bury et al. 2021). ML models cannot accurately capture such extreme events. Hence, for these extreme events, we hope ML models trained on a sufficiently diverse database of possible SM variation, so that complex relationship between SM and predictors will be captured better and "tipping points" will be approached. In the future work, a possible solution is to apply a Land surface model, such as Common Land Model (Dai et al. 2003), to simulate large numbers of SM data and select the local bifurcations in SM variation as supplementary samples.

Comment#6: Grammar mistakes can be noticed in many places, for example, for the sentences between lines 82-87, 91-06, and 112-114 among others. The authors are suggested to get help from native English speakers and thoroughly check the whole manuscript before the next submission.

Sorry for the grammar mistakes. We have carefully checked the whole manuscript and revised the inaccurate description. We will also ask a native English speaker to help us for English revision before the next submission.