Dear reviewer,

Thank you for your comments and suggestions. Based on your suggestion, I have made certain revisions to the manuscript.

(1) misuse of desertification, monsoon and other geographic terms throughout the manuscript. The study region, defined by the authors, is not “areas affected by desertification”, neither “monsoon climate region”. Please check in detail.

Thank you very much for this suggestion. First, for the "monsoon climate region". I found that my description was wrong after submitting the manuscript. It should be a temperate continental climate.

Northern China is mostly arid with an annual precipitation of less than 400 mm. The region belongs to the temperate continental ~~monsoon~~ climate and is subject to large-scale desertification. The desert areas of Northern China are susceptible

Second, for “areas affected by desertification”. Our study area is provided by the National Forestry and Grassland Administration (China). I checked with the appropriate manager recently, and they thought that it is more scientific to replace “areas affected by desertification” with “desertification areas”. All of these words have been changed in the manuscript.

Daily soil moisture mapping at 1 km resolution based on SMAP data for desertification areas~~areas affected by desertification~~ in Northern China

(2) Line 18-19: are you sure “very sensitive to SM”?

Thank you very much for this suggestion. The ultimate purpose of our study is vegetation restoration in the study area, where soil moisture is a key indicator. "very sensitive to SM " is a bit absolute, I have modified it to " sensitive to SM".

such as vegetation index and surface temperature. Areas affected by desertification in Northern China, which are ~~very~~ sensitive to SM, were selected as the study area, and the downscaled SM with a resolution of 1 km on a daily scale from 2015 to 2020

(3) Line 32: provide references for GLDAS.

Related literature has been added.

In the past, SM data were mainly obtained through ground measurements or the assimilation of products based on land surface models such as the Global Land Data Assimilation System (GLDAS) (Fang and Lakshmi, 2014; Zawadzki and Kędzior, 2016; Liu et al., 2021). Although most accurate SM data at different soil depths can be obtained, field measurements and in

(4) Line 36-40: data assimilation products may be produced with satellite data as inputs. Thus, it is not independent on remote sensing. Modify your statements.

I'm afraid I did not make it clearly. What I want to express in this sentence is: the accuracy of the soil moisture product of data assimilation is mainly affected by the land surface model and the original input data. The remote sensing data you mentioned is the original input data. It is not the same as the remote sensing SM product we mentioned above. In order to avoid ambiguity, I have made some modifications.

(Peng et al., 2021). Compared to ground measurements, remote sensing products can provide good spatial and temporal coverage of SM with a relatively low cost to the user (Zeng et al., 2015; Zhao et al., 2018; Meng et al., 2020). Data assimilation SM products largely depend on the accuracy of the land surface model and the original inputs ~~data~~ (Zawadzki and Kędzior,

(5) Line 44: what does the "very stable" mean here? Passive microwave radiometer data are sensitive to more influences, such as atmospheric effects and surface vegetation.

I'm afraid I did not make it clearly. I would have liked to express that passive remote sensing products are generally more stable compared to active remote sensing. In order to avoid ambiguity, I deleted it.

scattering and greatly affected by the surface roughness and vegetation types (Lievens et al., 2011; Wagner et al., 2013). Unlike active sensors, passive microwave radiometers or sensors are rarely affected by scattering ~~have almost no scattering and generate very stable SM products~~ (Abbaszadeh et al., 2019). Common passive microwave SM products are listed in Table 1

(6) Line 56: what does 'directly retrieve' mean?

I'm afraid I did not make it clearly. It has been modified.

proposed the temperature vegetation dryness index (TVDI) and used it to assess the SM status. ~~Despite their higher resolution, however, optical remote sensing data do not allow to directly retrieve true SM.~~ Relative SM indicators can be calculated using optical remote sensing data, however, reliable ground measurements or other data are still required to obtain the true value of SM.

(7) Each method produces a dataset. That does not mean the multiple machine learning methods produce the datasets following the normal distribution. In this sense, statistical mean may be biased, which is well-known to climate community.

Thank you very much for this suggestion. Our study uses a combination of multiple machine learning to select the best regression model for each period and not by taking an average for the SM result.

(8) Line 91-92: wrong description of the region with "monsoon climate". So is the desertification.

Thank you very much for your suggestions. I have made modifications. Refere to Question (1).

(9) Line 93: “water-vapor-ecosystem”, what does it mean?

I'm sorry for the non-standard expression. I made a simple modification.

to climate and hydrological changes and have fragile ecosystems. Soil water is a key parameter in land-atmosphere interactions~~of the water-vapor-ecosystem~~(Ma et al., 2019), and its change greatly affects the survival of vegetation and

(10) Line 115: Give the full spelling for NDWI, LSW, ECMWF, EVI, geotiff and many others for their first appearance in text.

Thank you very much for this suggestion. I have carefully checked all abbreviations, and some abbreviations that do not appear are added with full spelling.

The soil wetness related indexes, including NDWI, NSDSI, and Land Surface Water Index (LSWI), were produced using bands of the MOD09A1 product. Their formulas are:

160 ~~ECMWF~~ reanalysis dataset (ERA5) produced by European Centre for Medium-Range Weather Forecasts (ECMWF) provides

(11) The parameters used for ML are linearly correlated. Does it affect your results?

Thank you very much for this suggestion. Collinearity between variables will affect the simulation results, which is not considered in the description process of this paper. We add some content in Section 4.2.

In general, except for ensemble algorithms (including RF and XGB), collinearity is more or less affected. Due to this advantage of the ensemble algorithm, many studies generally do not consider multicollinearity problems when using random forests for regression or classification.

389 **4.2 Advantages of model combination**

390 Both RF and ANN have been applied to downscale remote sensed SM so far, especially RF (Zhao et al., 2018; Qu et al.,
391 2019; Hu et al., 2020). This study showed that the simulation results of ANN have greater uncertainty, and the accuracy is
392 generally worse than that of RF (Figs. 4 and 5). The RF algorithm shows a good simulation ability, but in comparison, the
393 XGB algorithm also has a corresponding effect or even higher. We also compared our simulation results combining multiple
394 models and the RF-based simulation results. The results showed that the combined products have higher accuracy than the RF-
395 based products, which is mainly reflected in the relatively more reasonable simulation of peaks and valleys (Figs. 9 and 11).
396 MLR has the worst effect compared to the other four models, which is likely to be affected by variable collinearity. In fact,
397 many algorithms, especially linear ones, exhibit more or less poor robustness when there is high collinearity between variables
398 (Dormann et al., 2013; Cammarota and Pinto, 2021). However, several studies have shown that ensemble algorithms such as
399 RF and XGB are generally not affected by multicollinearity (Tomaschek et al., 2018; Chen et al., 2020; Feng, 2021).

(12) Line 177: incomparable?

This word I refer to the literature (Wang et al., 2020). The expression is a bit absolute, replaced by “prominent”.

Extreme Gradient Boosting (XGBoost), as a new ensemble learning method, was proposed by Dr. Chen Tianqi at the university of Washington in 2016 [25]. For now, this method has incomparable advantages in generalization performance, speed and accuracy compared with other ensemble learning algorithms [26–28]. For instance,

(13) Equations for RMSE (6) and (8) are wrongly expressed.

There should be nothing wrong with these two equations. The following is the equations from the literature (Hu et al., 2020).

$$\mathrm{RMSE} = \sqrt{E\left[(\theta_{SMAP} - \theta_{insitu})^2\right]} \quad (7)$$

$$\mathrm{ubRMSE} = \sqrt{E\{[(\theta_{SMAP} - E[\theta_{SMAP}]) - (\theta_{insitu} - E[\theta_{insitu}])]^2\}} \quad (8)$$

where $E[\bullet]$ represents the mean operator, θ_{isitu} is the in situ SM, θ_{SMAP} is the downscaled SM, σ_{SMAP} is the standard deviation of SMAP SM, and σ_{SMAP} is the standard deviation of the in situ SM.

(14) Figures 4 and 5: there are clearly seasonal variation in correlation coefficient and RMSE. It means significant systematic errors in the products. Give scientific explanation to the data reliability.

Thank you very much for your suggestion. Since the overall simulation effect is better, I have not considered this issue carefully before. My understanding is that the relatively poor effect in summer may be due to the effect of noise such as clouds. I added it in Section 4.1 (Discussion part).

While this study greatly improved the spatial resolution of SM data from 2015-2020 in the desertification areas of North China by downscaling SMAP SM products, it still presents some shortcomings. Although we chose high-quality MODIS images, variables such as daily LST and Albedo are still affected by clouds. It will have some impact on the simulation results, especially in the rainy season. Figures 4 and 5 show that the simulation results are generally better in the cold season with less clouds, and worse in summer with more clouds. ~~For example, D~~due to the image quality and coverage of SMAP ~~and the impact of noise from clouds on the MODIS products~~, the number of valid samples for a 16-day period may still be less than 100 points in cold seasons (Fig. 3). This study replaced the periods with less than 100 samples with the model of the previous periods.

(15) Line 251-260: The errors are large between the Maqu and the Bbaso network, which need substantial investigation.

The evaluation accuracy of Babao network is generally lower than that of Maqu Network, the first reason is the measured soil depth. The soil depth measured by SMAP is 5 cm, the same as that measured by Maqu Network, while that measured by Babao Network is 4 cm. Another reason is that there is a bias between site measurements and remote sensing data itself. The relationship between the 1 km×1 km grid and the site itself will have a large error. In this study, even SM of some sites from Maqu network

did not match well the Remote sensing SM. To make the results more convincing, I added the following.

Further, all gridded SM products are compared with in situ SM. Figure. 8 shows a significantly higher correlation between the downscaled SM and in situ SM of the Maqu Network. The ubRMSE median of the downscaled SM is the smallest, and its RMSE is second only to the C3S (0.25°) product. The bias of the downscaled SM is higher than that of some products, even higher than the original SMAP L3 (36 km) data. Almost the same results can be obtained from in situ observations of Babao Network. (Fig. S2). The difference is that the bias of the downscaled SM is lower than the result of SMAP L3 (36 km). Compared with RF-based downscaled SM, the downscaled SM with multiple machine learning approaches performed better, especially R and RMSE.

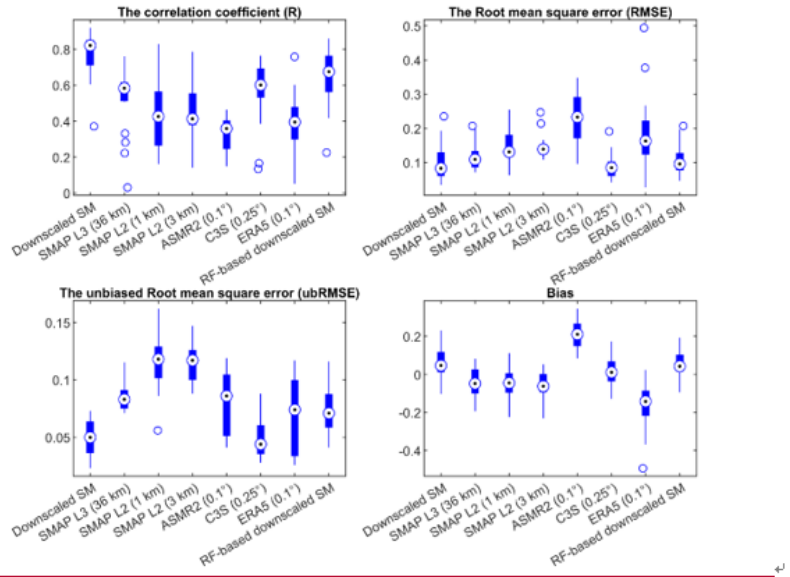

Figure 8: Comparison of gridded products and in situ observation SM of the Maqu Network.

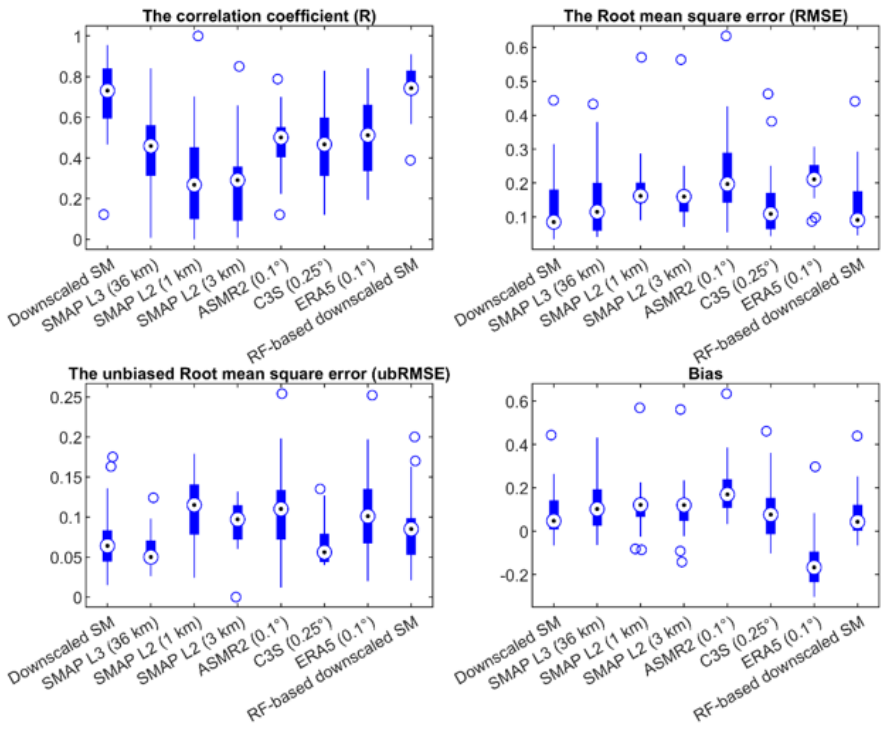

Figure S2: Comparison of gridded products and in situ observation SM of the Babao Network.

(16) Line 283: "due to spatial resolution" is a superficial reasoning. Insightful clarification should be given.

I'm afraid I didn't express clearly. This sentence seems more like a discussion, so I deleted it. For a detailed explanation, refer to Section 4.4.

with the precipitation. ~~The difference between the downscaled SM and the in-situ measured SM is mainly reflected in the magnitude of the variation, which is probably due to the difference in spatial resolution~~. The variation trends of the RF-based

(17) Line 291: here appears 'process of vegetation growth'. SMAP SM data are subject to vegetation cover, which is known in the field, but the authors failed to address it.

This study does not consider the quality of SMAP data itself. Generally speaking, passive remote sensing is relatively less affected by surface roughness and vegetation. On the contrary, active remote sensing is more affected.

In fact, I add this sentence here mainly to express that the high SM from April to September is consistent with the seasonal law of vegetation growth. In order not to cause ambiguity, I removed it.

SM varies greatly in different months in desertified areas. ~~Figure 9~~Figure 10 shows the average SM in each month in the study area. The SM shows a monthly change pattern, and the values from June to September are bigger than in other months, especially in southern Qinghai Province, eastern Inner Mongolia Province, and western Xinjiang Province~~, which is consistent with the process of vegetation growth~~. The SM in some areas is low throughout the year, such as in the Tarim Basin of Xinjiang

(18) Line 295: “little variation”? change the words.

Thanks for your advice! It has been modified.

The annual average SM was also calculated (Fig. ~~S2~~S3). Compared with the monthly average SM, the annual average SM changed significantly less.~~Overall, there is little variation in SM in different years~~. Further, we compared the spatial

(19) There are too many “some” in text. Vague expression.

Thank you very much for your suggestion. The unnecessary "some" of the manuscript have been deleted. Some expressions are also further modified.

(20) Line 327: strange subtitle.

I referred to this writing form Zhao et al. (2018). Not very scientific, I modified it to “Variable importance assessment”.

4. Discussion

4.1 Variable importance assessment~~Regression variable importance~~

(21) Line 335: “influence of soil texture (sand, silt and clay) is relatively weak, but it cannot be completely ignored.”. why?

It caused some ambiguity and I deleted the latter part.

NDWI, which was demonstrated in Yue et al. (2019). Topographical factors also exhibite importance on SM, especially elevation. The influence of soil texture (sand, silt and clay) is relatively weak~~, but it cannot be completely ignored~~.

(22) Line 347: IncNodePurity? What is it?

It reflects an important indicator, and I added its full spelling.

~~Figure 12~~Figure 13: The average importance scores of variables for the RF based approach and XGB based approach. Note: The importance scores are presented by increase in Node Purity (IncNodePurity) where the sum value is normalized for the RF model; The XGB model uses Gain to reflect the weight of variables.

(23) Line 350: various noises? How many?

My expression was not clear, deleted "various".

The simulation results of long time series will inevitably suffer the interference of ~~various~~ noises. A combination of multiple methods can reduce overfitting and uncertainties (Zanotti et al., 2019; Yu et al., 2021). The five methods (MLR, SVR,

(24) Line 367: mainly significantly. Remain one.

I'm afraid I didn't express clearly. "mainly" was modified to "more"

SM is mainly positively correlated with precipitation and temperature, and a few regions are significantly negatively correlated with temperature. In terms of spatial distribution, SM of the sites in the eastern region (including Inner Mongolia Province, Hebei Province and Shanxi Province) is ~~mainly~~ more significantly affected by precipitation. Due to the influence of glaciers and snowmelt, the SM of the sites in the western region (Xinjiang Province and Gansu Province) is more affected by

(25) Line 382-383: delete it.

It has been removed.

(26) Line 391: "a framework was proposed"? It does not make sense.

I'm afraid I didn't express clearly. "a framework" was modified to "an approach".

In this study, ~~a framework~~an approach was proposed for downscaling 36 km SMAP SM products using MODIS optical products and other surface variables (mainly topographic data and soil data) based on multiple machine learning methods.

Finally, thanks a lot for your careful review and invaluable advices. 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!