Dear commenter,

Thank you for your comments and suggestions. Firstly, uncertainty. In order to improve the simulation accuracy, I have tried many machine learning methods in the early stage, and chose a more representative method to describe in the manuscript. And I also found that tree classifiers, especially Xgboost, have significantly better performance and efficiency compared to other classifiers.

Secondly, in order to improve the accuracy, we set 16 days as a regression period. It reduces errors caused by poor timeliness of dynamic variables (such as NDVI and EVI) and little valid data for one day.

Lastly, in order to verify the accuracy of our data, in addition to the limited in-situ observed soil moisture data and precipitation data, we also compared the data with some mainly existing reliable gridded soil moisture product, such as SMAP L2 SM (1 km and 3 km), GCOMW/ASMR2 SM (0.1°), C3S SM (0.25°), ERA5 SM (0.1°) and FLDAS SM (0.1°). It turns out that the data we produced has obvious advantages, which are mainly reflected in three points. First, the value of our data is generally in the middle of these products, and it is also relatively close to the in-situ measured values (see Figure 8). The second is time series. The product we produced generally has more valid data compared to other products, and its variation range is more reasonable than several other products (see Figure 8). The third is the spatial distribution of these products. Our products present a better spatial pattern of soil moisture, which is close to the actual situation, and its high spatial resolution makes some information displayed more clearly than other products (see Figure 10). Of course, the description of uncertainty in the manuscript is not detailed enough. We will try to modify this part of the content.

DOY is the day of year. Since all MODIS products use this Julian date, this manuscript also names the data in this way for convenience. This dataset is freely available at https://doi.org/10.6084/M9.FIGSHARE.16430478.V5. I will describe it in detail in the manuscript.

Looking forward to your next suggestions. Thank you!

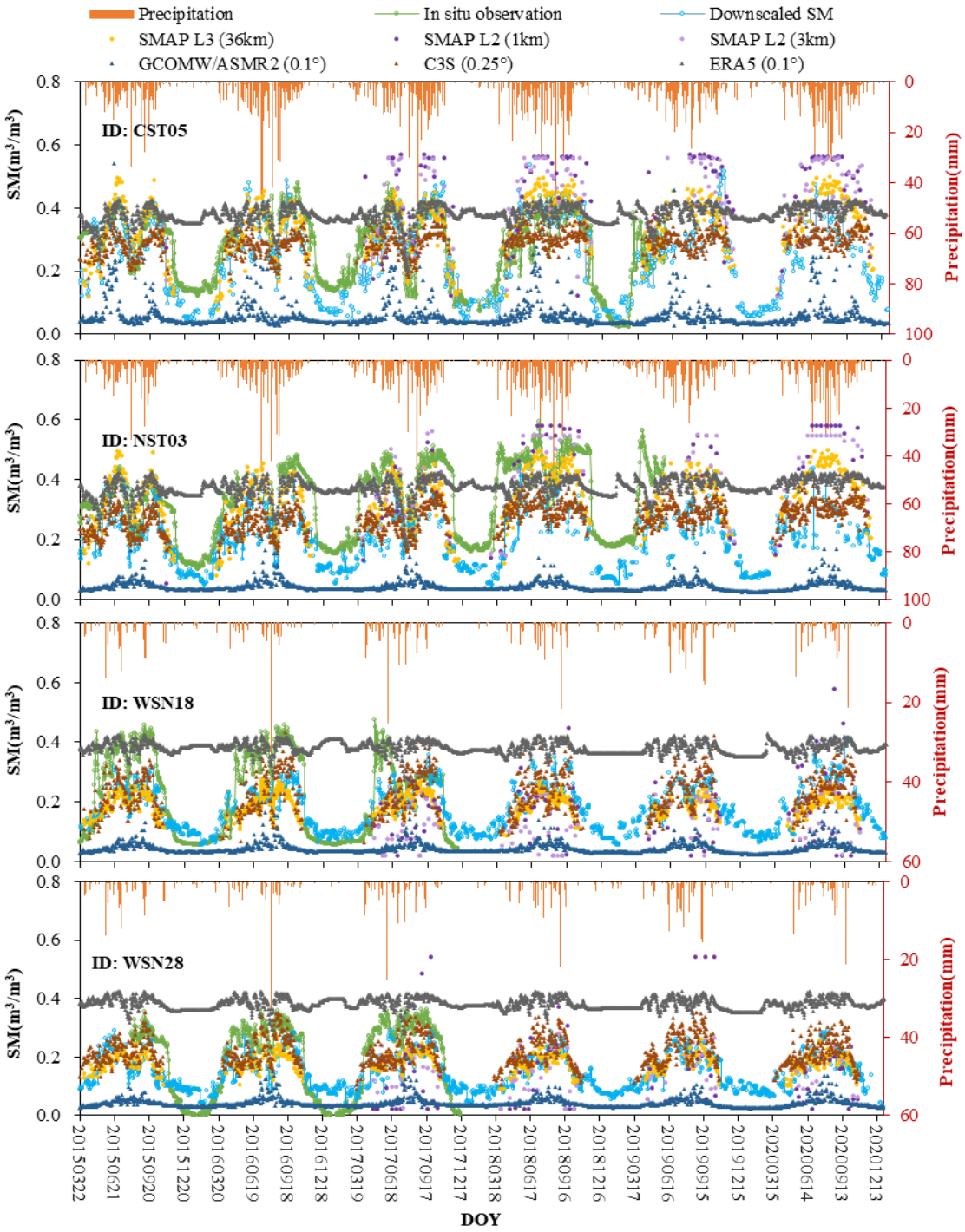

Figure 8: Time series of the in situ observed SM, the downscaled SM, the acquired gridded SM products and daily precipitation at the four selected SM sites (From Maqu Network and Babao Network, respectively).

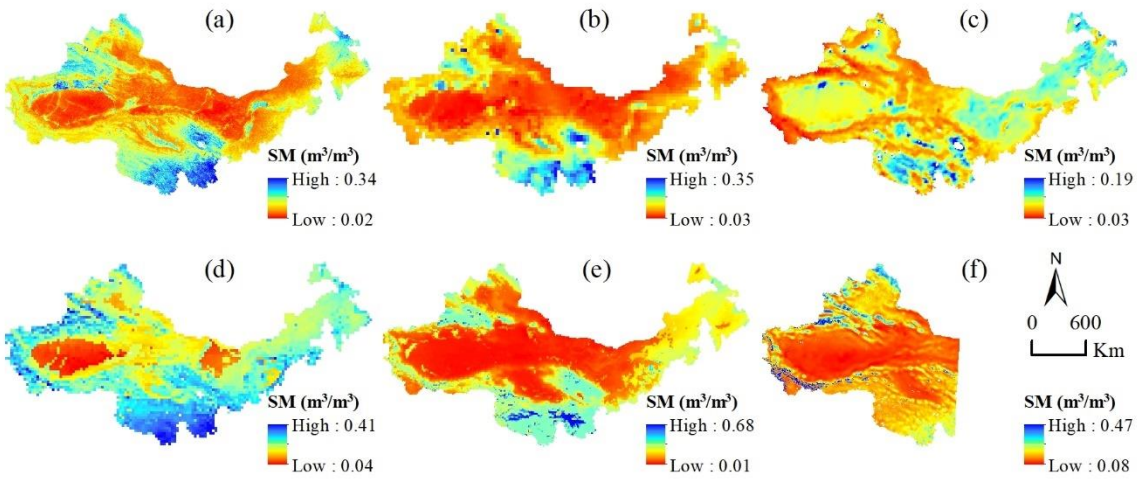

Figure 10: Daily average SM from 2015-2020 in the study area. (a)-(f) are the downscaled SM (1 km), SMAP L3 SM (36 km), GCOMW/ASMR2 SM (0.1°), C3S SM (0.25°), ERA5 SM (0.1°) and FLDAS SM (0.1°), respectively.

model. The in situ observed SM data obtained in this paper are relatively limited, and their spatial distribution is concentrated in a certain part of of the study area, which is not representative. It increases the uncertainty of the simulation results. In order to verify the accuracy of the data as much as possible, we compared this product with the existing more reliable gridded SM products, and the results showed that our product showed certain advantages in both time series and spatial distribution (Figs .8 and 10).

The Chinese government focuses on desertification reduction through afforestation and the establishment of grasslands. SM data with high temporal and spatial resolution can provide a reference for the next steps of revegetation.

5 Code and data availability