Dear commenter,

Thank you for your comments and suggestions. Based on your suggestion, I have made certain revisions to the manuscript. The first is to describe the XGB algorithm in more detail. The second is to discuss the deficiencies of the RF algorithm in more detail. The main revisions include the following:

(1) Introduction: Added description of shortcomings of RF algorithms

correlated with the dependent variable (Elshorbagy and Parasuraman, 2008; Ågren et al., 2021). Since RF algorithm uses random sampling with replacement, its simulation results will not exceed the range of training set and tend to ignore some extreme values when used as a regression model (Belgiu and Drăguţ, 2016). –Extreme gradient boosting (XGB), as a new ensemble learning method (Chen and Guestrin, 2016), performs well in some fields (Wang et al., 2020; Fan et al., 2021; Ma et al., 2021), but it has rarely been used for soil moisture downscaling. Compared with methods such as RF, XGB adopts the boosting weighted sampling method, which can better simulate the extreme values existing in the samples (Chen and Guestrin, 2016). The coarse-resolution remote sensed SM (>10 km) itself has ignored some maxima or minima with relatively finer-grid SMs, so a method that better simulates extreme values will obviously have certain theoretical advantages.

(2) Methods: Added the description of the main formula of the XGB algorithm

RF and XGB are tree based ensemble algorithms, which have prediction accuracy and good generalization ability, and are not prone to overfitting (Rao et al., 2018; Abbaszadeh et al., 2019). RF is a multiple-tree algorithm improved by ~~Bootstrap~~ bootstrap to reduce decision tree bias in determining the splits (Mohana et al., 2021). Many studies have used RF to build regression models of remotely sensed SM and related variables, and almost all achieved better results compared to other regression methods (Zhao et al., 2018; Qu et al., 2019; Hu et al., 2020). In contrast, the application of XGB, which applies a

8

regularized gradient boosting framework, is still very limited. However, XGB has incomparable advantages in generalization performance and accuracy (Wang et al., 2020).

The XGB algorithm is a boosting-type ensemble of multiple CART decision trees (Chen and Guestrin, 2016). The predicted result of the boosting-type tree ensemble model can be expressed as follows:

$$\hat{y}_i = \emptyset(x_i) = \sum_{k=1}^{K} f_k(x_i), f_k \in F \qquad (4)$$

where F is the space of regression tree, K is the total number of trees, which means the model uses K additive functions, $f_k(x_i)$ is the weighted score of the k-th tree on i-th input data (x_i).

XGB adopts a regularized learning objective to optimize the simulation results.

$$Obj(\emptyset) = \sum_{i=1}^{N} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \qquad (5)$$

where l is the loss function, N is the total number of input, Ω is the regularization term to penalize the model complexity and prevent overfitting.

(3) Result 1: Added the comparison results using boxplots for all SM products and in situ observed SM.

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.

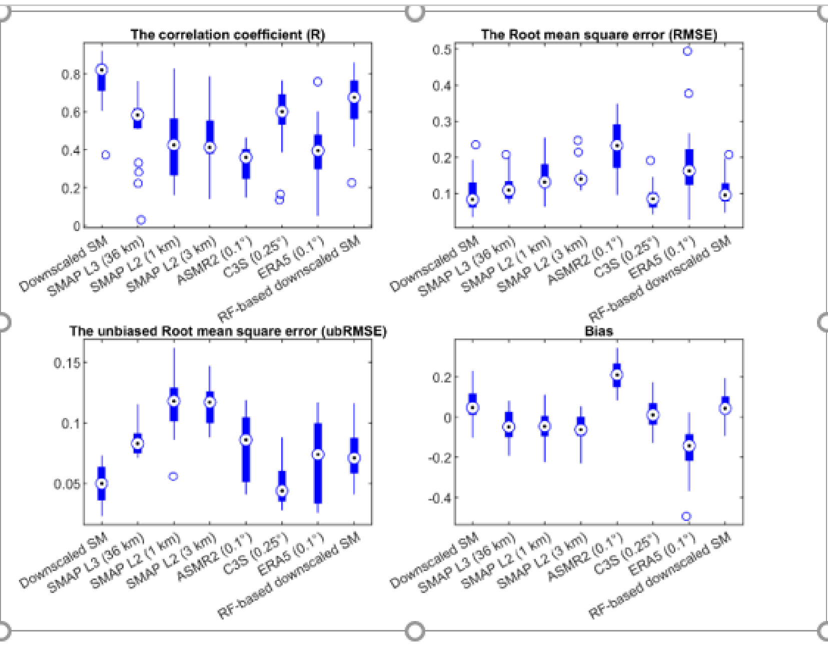
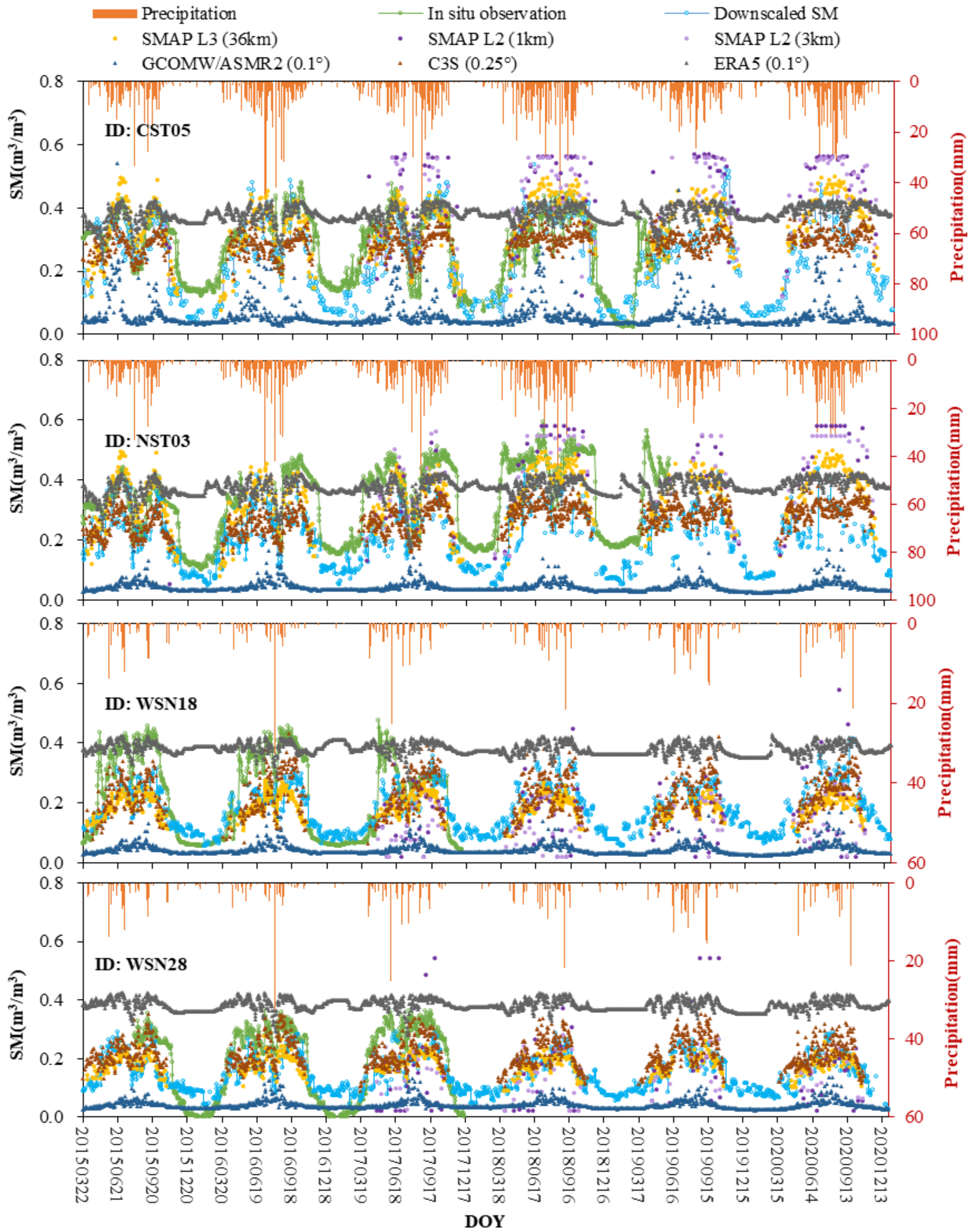
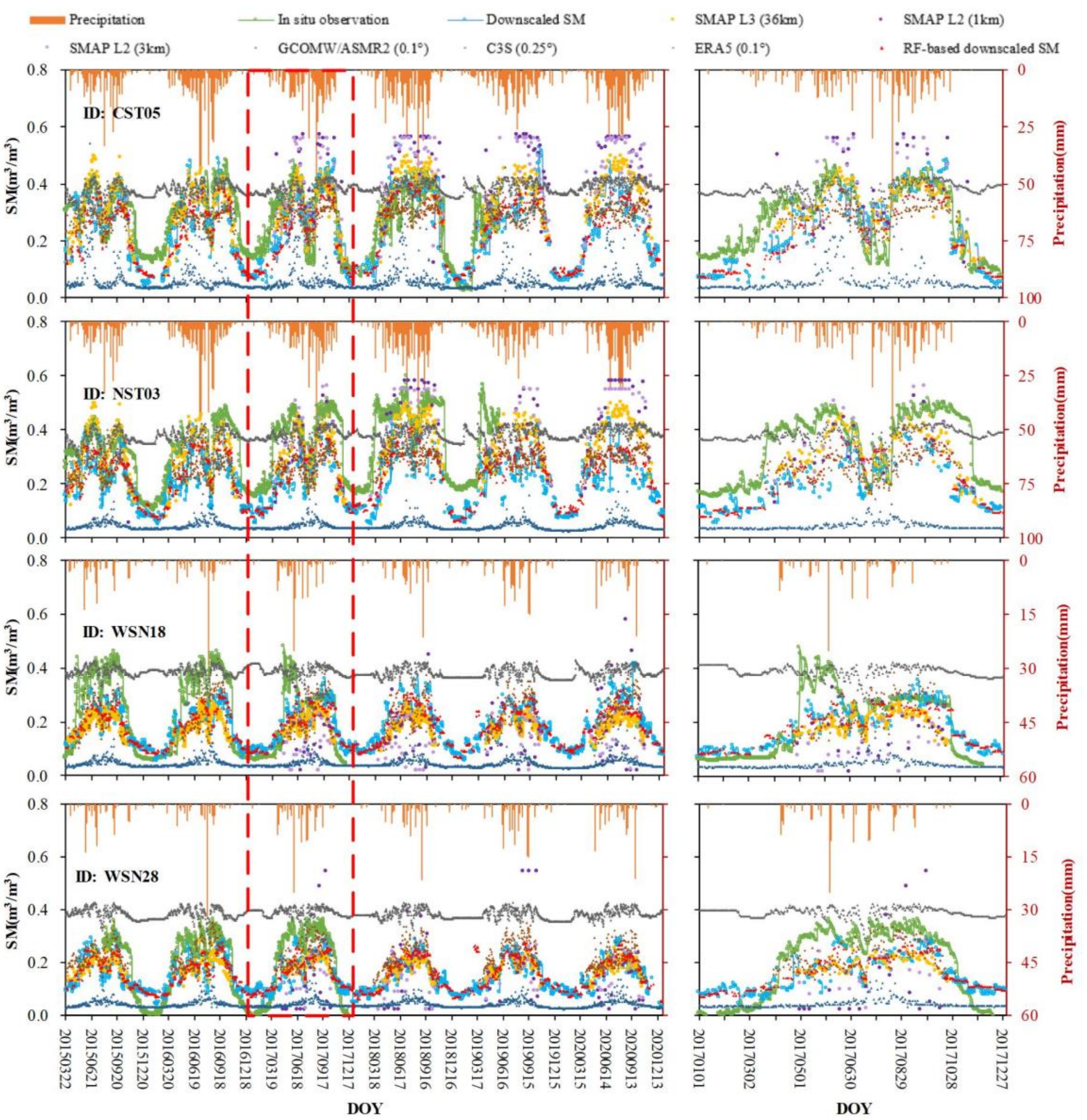


Figure 8: Comparison of gridded products and in situ observation SM of the Maqu Network.

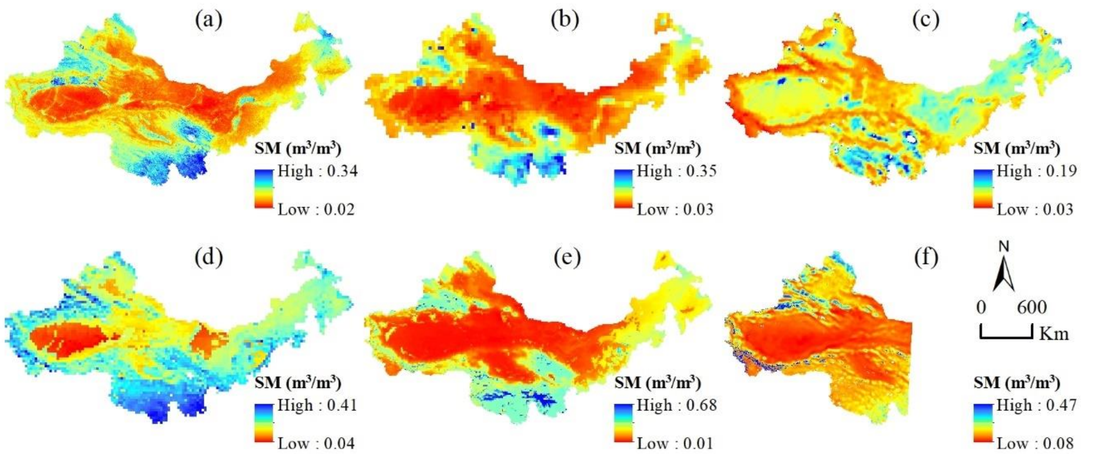
(4) Result 2: Given that random forests have proven good methods in some literature and our computational results, we added the RF-based downscaled SM in Figures 9 and 11 (Figures 8 and 10 of the original manuscript).



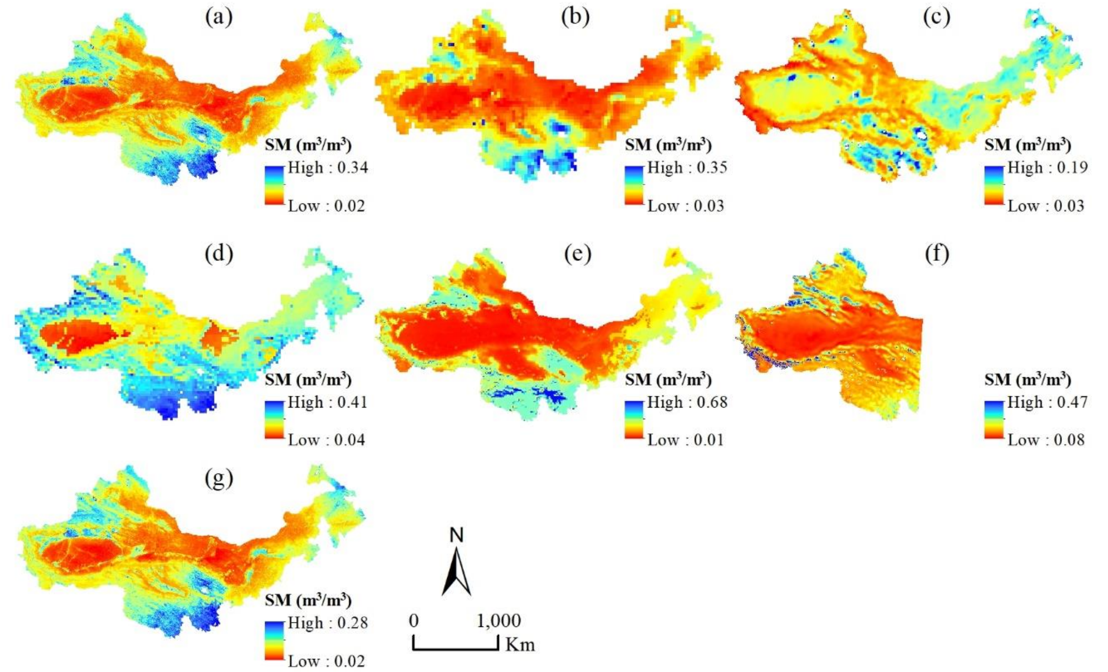
Precipitation
In situ observation
Downscaled SM
SMAP L3 (36km)
SMAP L2 (1km)
SMAP L2 (3km)
GCOMW/ASMR2 (0.1°)
C3S (0.25°)
ERA5 (0.1°)
ID: CST05
ID: NST03
ID: WSN18
ID: WSN28
SM(m3/m3)
Precipitation(mm)
20150322
20150621
20150920
20151220
20160320
20160619
20160918
20161218
20170319
20170618
20170917
20171217
20180318
20180617
20180916
20181216
20190317
20190616
20190915
20191215
20200315
20200614
20200913
20201213
DOY



Precipitation
In situ observation
Downscaled SM
SMAP L3 (36km)
SMAP L2 (1km)
SMAP L2 (3km)
GCOMW/ASMR2 (0.1°)
C3S (0.25°)
ERA5 (0.1°)
RF-based downscaled SM
ID: CST05
ID: NST03
ID: WSN18
ID: WSN28
SM(m³/m³)
Precipitation(mm)
DOY



(a)
(b)
(c)
(d)
(e)
(f)
SM (m³/m³)
High : 0.34
Low : 0.02
SM (m³/m³)
High : 0.35
Low : 0.03
SM (m³/m³)
High : 0.19
Low : 0.03
SM (m³/m³)
High : 0.41
Low : 0.04
SM (m³/m³)
High : 0.68
Low : 0.01
SM (m³/m³)
High : 0.47
Low : 0.08
N
0 600
Km



(5) Discussion: Based on the simulation results of random forests, the advantages of the joint approachs are discussed.

In order to reduce the error caused by the method, we also adopt a combination of multiple models to obtain the optimal results. Both RF and ANN have been applied to downscale remote sensed SM so far, especially RF (Zhao et al., 2018; Qu et al., 2019; Hu et al., 2020). This study showed that the simulation results of ANN have greater uncertainty, and the accuracy is generally worse than that of RF (Figs. 4 and 5). The RF algorithm shows a good simulation ability, but in comparison, the XGB algorithm also has a corresponding effect or even higher. We also compared our simulation results combining multiple models and the RF-based simulation results. The results showed that the combined products have higher accuracy than the RF-based products, which is mainly reflected in the relatively more reasonable simulation of peaks and valleys (Figs. 9 and 11).

Looking forward to your next suggestions. Thank you!