Dear Editor and Reviewer # 3:

We appreciate your insightful comments on our paper. The comments offered have been immensely helpful. We have responded to every question, indicating exactly how we addressed each concern or problem and describing the changes we have made. The revisions have been approved by all authors. The point-to-point responses to your comments are listed below in **blue**.

This is an interesting effort in developing the SM product for crop dryland, which has potential for various applications. The paper is well written and organized. Taking the CIR as a predictor seems to be a useful way to predict SM in crop dryland. However, I have some concerns as following. Please pay more attention on the comments about line 174-175.

Point 1: Why only mapping SM for dryland, not rice?

Response: Rice is commonly grown in southern areas with plenty rainfall or northern areas well equipped by irrigation in China. Therefore, soil moisture is usually over saturated and keeps constant (near 100%) during the whole growing season (Zheng et al., 2000; Alhaj Hamoud et al., 2019). Considering the significant role of SM for maize and wheat planted in dryland across China, we mapped the SM for crop drylands, not including rice.

Reference:

Alhaj Hamoud, Y., Guo, X., Wang, Z., Shaghaleh, H., Chen, S., Hassan, A., and Bakour, A.: Effects of irrigation regime and soil clay content and their interaction on the biological yield, nitrogen uptake and nitrogen-use efficiency of rice grown in southern China, Agricultural Water Management, 213, 934–946, https://doi.org/10.1016/j.agwat.2018.12.017, 2019.

Zheng, X., Wang, M., Wang, Y., Shen, R., Gou, J., Li, J., Jin, J., and Li, L.: Impacts of soil moisture on nitrous oxide emission from croplands: a case study on the rice-based agro-ecosystem in Southeast China, Chemosphere - Global Change Science, 2, 207–224, https://doi.org/10.1016/S1465-9972(99)00056-2, 2000.

Point 2: Line 110-115: there are two sources of FC which one is used?

Response: Field capacity (fc) was obtained from OpenLandMap which included fc under 33kPa at 0cm (b0) and 10 cm (b10) depth. When predict ChinaCropSM$_{0\text{-}10cm}$, we used fc under 33kPa at 0cm (b0) depth. When predict ChinaCropSM$_{10\text{-}20cm}$, we used fc under 33kPa at 10cm (b10) depth.

Point 3: Line 120: the short name "AMS" is used only one time. Consider full name. In addition, what is R4, R5 and R16? And it should not be calculated only for AMS but for each cell, as a predictor.

Response: Many thanks for your advice. We have increased the full name of "AMS" in the revised paper.

Yes, R4, R5 and R16 is calculated for each cell, as a predictor. Actually, the R4, R5 and R16 are river network vector data at different levels in China. When training sample data, we calculate the distance for AMS. Additionally, we calculated the distance from each cell to river network vector data when predicting the ChinaCropSM.

Point 4: Line 171: Grammar error. Not a complete sentence.

Response: Thank you for your careful comments. We have modified it (Line 180).

"As for the response variable (Classified Irrigation CIR), it is calculated by irrigation threshold (Table 2) and in situ information, including crop type, phenology and soil depth.".

Point 5: Line 174-175: It should not be random splitting because SM of different time from the same site may be highly correlated. This will give a higher performance for the model. Instead, the splitting should be based on sites, i.e., data from a site should be all in the training set or all in test set. Note that the model is predicting unknown locations based on the observing sites, and the spatial interpolation ability should be evaluated by the site-based splitting.

Response: Thanks very much for your constructive comment.
According to your site-based splitting method, we re-optimized the hyper-parameters of the prediction model to reduce overfitting and evaluated the prediction results. We found the soil moisture predicted by your method agreed well with in situ SM observations (ubRMSE ranges from 0.046–0.057, and R^2 ranges from 0.642–0.761), although the model performance drops slightly (Figure 1).

Similarly, in the case of site-based splitting, all prediction accuracy of SM were consistently improved both for crops and depths with comparison of those without an irrigation module (e.g. R^2 increased by 9–41%, ubRMSE decreased by 21–26%) (Figure 2). Also, we further compared our ChinaCropSM1km with the two popular public global SM products (Table 1). All indexes of our ChinaCropSM were consistently indicated by the higher accuracy.

Different splitting methods during training and testing do affect model performance. Selecting which splitting method to improve the generalization performance is dependent on data. Generally, the larger size of data, the smaller effect of the splitting methods on the results (Birba, 2020). Therefore, the model performances of two splitting methods show no significant differences because of quantities of field observations available in our study. We have followed you to insert deeper and more extent discussions into our manuscript (Line 307~322 in the revised manuscript).

Reference:

Birba, D. E.: A Comparative study of data splitting algorithms for machine learning model selection, 2020.

The results are following:

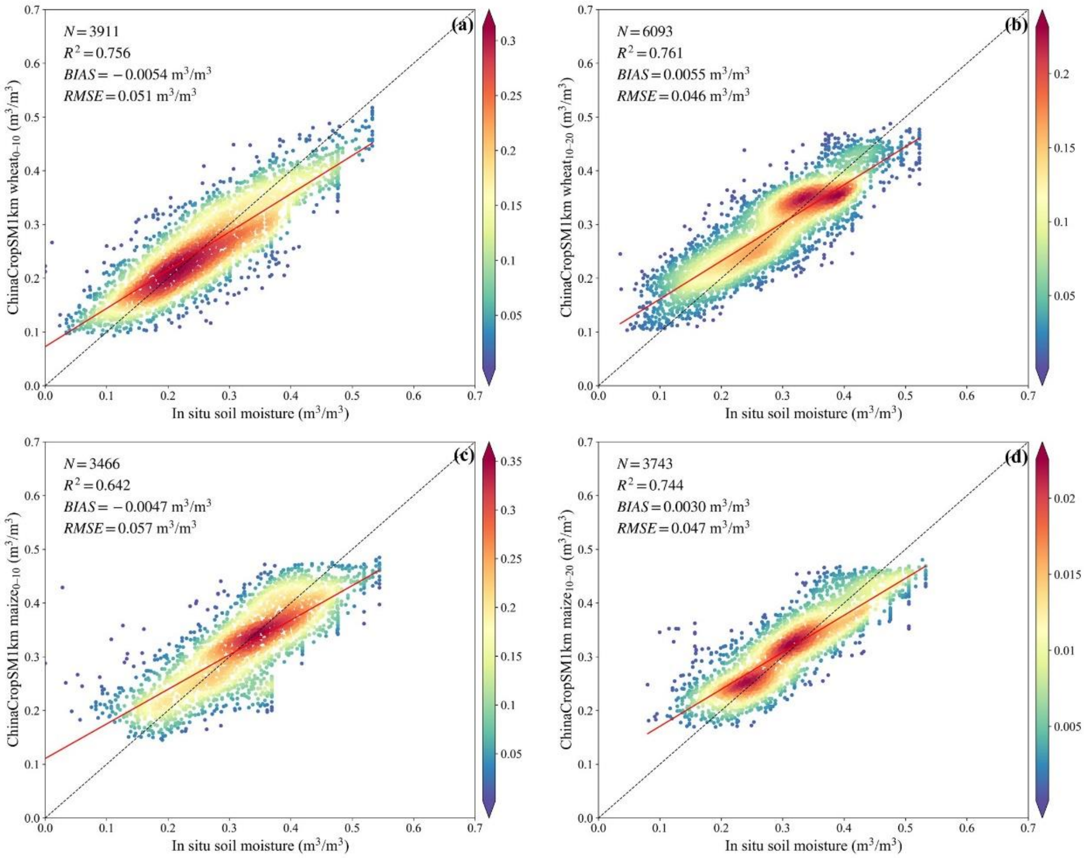

Figure 1 Comparison between the predicted soil moisture (ChinaCropSM1km) and in situ samples by crops and depths (cm) according to site-based splitting.

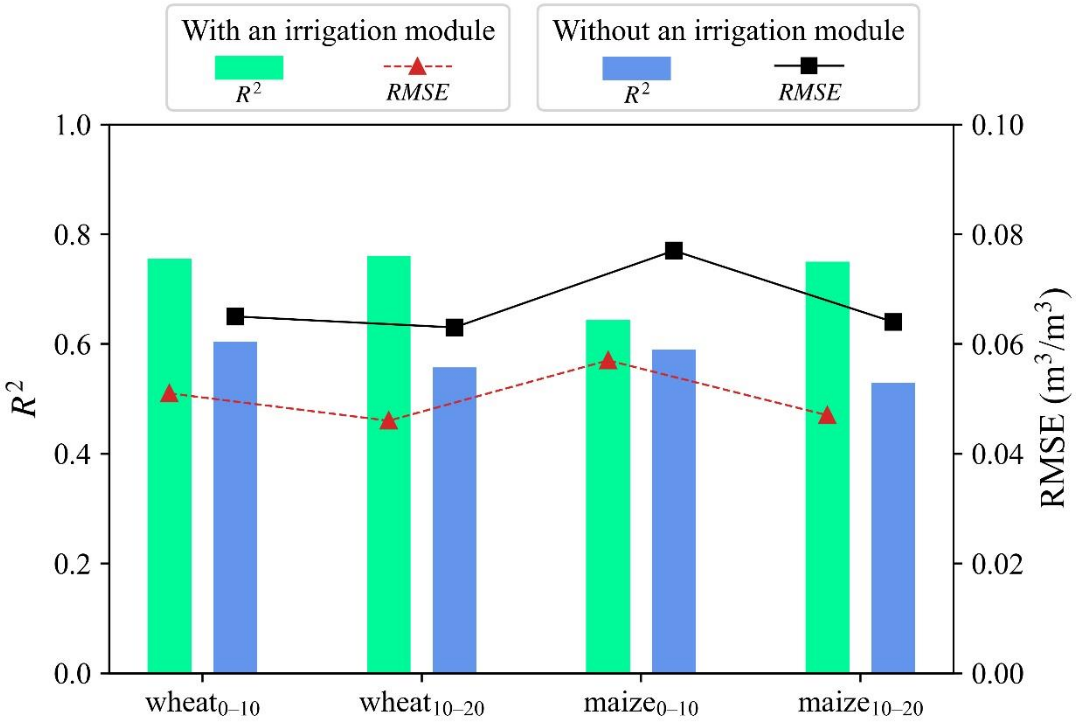

Figure 2 Comparison of soil moisture accuracy between with irrigation and without an irrigation module according to site-based splitting.

Table 1 Summary on means of evaluation indexes of three products (ChinaCropSM1km, RSSSM and ESA CCI SM).

Product	ChinaCropSM	RSSSM	ESA CCI SM
r	0.85	0.52	0.42
RMSE	0.054	0.144	0.120
bias	–0.005	–0.112	–0.066
ubRMSE	0.054	0.092	0.100

Point 6: Line 185: How many times do you run the model to get the importance, as the importance will be different each time. It should take the average importance of dozens of runs like 100.

Response: Yes, we did take the averages of dozens of runs. We ran each training model 50 times to get the importance and averaged the importance outcome.

Point 7: Fig.6 and 7: what are the different boxes stand for?

Response: The boxes in Fig.6 and Fig.7 actually stand for different results, with spatial pattern in Fig.6 and temporal one in Fig. 7. Both patterns were conducted between ChinaCropSM1km and the in situ SM observations.

The horizontal line within each box stands for median, the white dot for mean, the box bottom for first quantile, the top for third quantile, and black dots for outliers.

Point 8: Section 3.5: I do not think this comparison is fare. The evaluation using the test data for Cropland should be used instead of all in situ data because the model used them to establish leading to an independent evaluation.

Response: Actually, we only used the testing data for evaluating, not including all in situ data. We agreed well with you that using all observations will lead to an independent evaluation.