

Earth Syst. Sci. Data Discuss., referee comment RC3
<https://doi.org/10.5194/essd-2022-254-RC3>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on **essd-2022-254**

Anonymous Referee #3

Referee comment on "ChinaCropSM1 km: a fine 1□km daily soil moisture dataset for dryland wheat and maize across China during 1993–2018" by Fei Cheng et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2022-254-RC3>, 2022

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.

Why only mapping SM for dryland, not rice?

Line 110-115: there are two sources of $FC_{i\frac{1}{4}}$ which one is used?

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.

Line 171: Grammar error. Not a complete sentence.

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.

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.

Fig.6 and 7 ¹/₄ □ what are the different boxes stand for?

Section 3.5 ¹/₄ □ 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.