

Comment on **essd-2021-362**

Anonymous Referee #2

Referee comment on "Daily soil moisture mapping at 1 km resolution based on SMAP data for desertification areas in northern China" by Pinzeng Rao et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-362-RC2>, 2022

This manuscript describes a daily 1-km soil moisture dataset for North China in 2015-2020. The dataset was generated with several machine learning methods for downscaling from the well-known 36-km SMAP data. A soil moisture dataset of high spatiotemporal resolution is definitely desirable to earth-science community. However, there are many significant issues should be addressed. Here list a number of comments for potentially improving the manuscript.

- 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.
- Line 18-19: are you sure "very sensitive to SM"?
- Line 32: provide references for GLDAS.
- 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.
- 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.
- Line 56: what does 'directly retrieve' mean?
- 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.
- Line 91-92: wrong description of the region with "monsoon climate". So is the desertification.
- Line 93: "water-vapor-ecosystem", what does it mean?
- Line 115: Give the full spelling for NDWI, LSW, ECMWF, EVI, geotiff and many others for their first appearance in text.
- The parameters used for ML are linearly correlated. Does it affect your results?
- Line 177: incomparable?
- Equations for RMSE (6) and (8) are wrongly expressed.
- 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.
- Line 251-260: The errors are large between the Maqu and the Bbaso network, which

need substantial investigation.

- Line 283: "due to spatial resolution" is a superficial reasoning. Insightful clarification should be given.
- 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.
- Line 295: "little variation"? change the words.
- There are too many "some" in text. Vague expression.
- Line 327: strange subtitle.
- Line 335: "influence of soil texture (sand, silt and clay) is relatively weak, but it cannot be completely ignored.". why?
- Line 347: IncNodePurity? What is it?
- Line 350: various noises? How many?
- Line 367: mainly significantly. Remain one.
- Line 382-383: delete it.
- Line 391: "a framework was proposed"? It does not make sense.