Comment on essd-2021-362
Martin Schultz (Editor)

Editor comment on "Daily soil moisture mapping at 1 km resolution based on SMAP data for areas affected by desertification in Northern China" by Pinzeng Rao et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2021-362-EC1, 2022

Dear authors,

as handling editor for this article I had some difficulty making a judgement based on the two reviewers' comments. Since I was interested in the topic and methods you describe, I decided to conduct a review myself. Please find this below.

Kind regards,

Martin Schultz

This paper addresses an important topic and develops a promising methodology for the downscaling of soil moisture fields from a set of meteorological and soil input variables with the help of several machine learning models. While the manuscript is generally well written and understandable (also for me who is not directly an expert in soil science), I still have major concerns about publishing this manuscript in ESSD:

1.) I see a couple of methodological shortcomings related to the rigorousness of the machine learning methods applied (see detailed comments below). The methods description and critical discussion of results does not meet the current standards in the machine learning literature.

2.) this manuscript would fit much better into the sister journal Geophys Model Dev., because its focus is more on the method development and not so much on the data that is being produced. There are only few short statements about the relevance of the dataset produced, while the conclusion firstly emphasizes a new "framework".

3.) the description of some of the input data and of the target data product remains incomplete so that it is difficult to judge for a potential user of this dataset if it may be useful for her or his purposes.

In light of the above and given that reviewer 2 also had some critical remarks, I will reject this paper and encourage the authors to work on the issues described by reviewer 2 and below and resubmit to GMD instead.

Detailed comments:
I. 69 there is no such thing as "inefficient samples". Do you mean "insufficient samples"?

I. 83 unclear expression "Considering the role of SM in the ecological environment" -- "for ecology" or "for the environment"

I. 96 add reference to Figure 1. All figures must be referred to in the text. (this reference only appears in line 136)

general comment, page 4: it would be helpful to add a figure showing the precipitation climatology of the region and its variability rather than only showing the topography. This would allow readers to evaluate the geographic patterns of the SM results.

I. 144 following: to facilitate comparison of dataset resolutions, please convert all resolutions to km. Example: "It has a spatial resolution of 0.25 degrees (~28 km)". Also use the degree symbol consistently.

sections 2.2.2 to 2.2.4: please include more details about the datasets, such as spatial resolution, time span covered and hints about the evaluation of their quality.

Figure 2: what is a "set dataset"? I only know the terms "validation dataset" and "test dataset"

section 2.3.2: please define the target data product more more clearly. The section is named "downscaling", but the text describes a coarsening of resolution (to 36 km) and a functional mapping/regression from predictor variables to the predictant. What is the output resolution and how is this *downscaled*?

I. 202: how was the data split done? Given that there is substantial memory in the system and some data are used with a temporal resolution of 16 days, a simple random sampling approach is invalid as it will lead to overoptimistic validation results. This issue is discussed for example in Schultz et al., 2021, https://royalsocietypublishing.org/doi/10.1098/rsta.2020.0097

section 3.2: please add more quantitative evaluation results. There are many statements like "good" or "poor", but it remains unclear according to which target criterion these quality indicators are given. For example, what is an acceptable R2 value in your view and why?

I. 303: language. What does this mean "The average SM of the ERA5 products is polarised"?

section 4.1: be aware that there is a fine difference between the influence of a physical quantity on the target variable and the importance of input variables for the model regression. The latter can depend on the model architecture, data preprocessing and normalisation and other factors which have nothing to do with the cause-effect relationships in the real world. The text is ok as is, but it may help readers to assess the relevance of your results if you alert them to this point.

I. 350: The sentence "The simulation results of long time series will inevitably suffer the interference of various noises." is meaningless and not based on a sound mathematical concept from statistics.

I. 356 remove "even" before "smaller"

general comment on section 4.2 and also methods section: your statements about the quality of individual ML models are not supported by the evidence shown in the paper.
Information is missing about the detailed set-up of these ML models. Otherwise you might be comparing apples and oranges, for example if a very small ANN is compared to an RF with many trees and branches. Have you done any ablation studies to find the optimum set-up for each ML method? Furthermore, the discussion would benefit from a somewhat more elaborate reflection about why a certain ML model performs better or worse for certain seasons, groups of stations, etc.

section 4.3 How do these results relate to the variable importance analysis presented in Figure 12?