



Comment on **essd-2022-177**

Anonymous Referee #2

Referee comment on "A 1-km daily soil moisture dataset of China based on in-situ measurement using machine learning" by Qingliang Li et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2022-177-RC2>, 2022

The paper employs multiple datasets, including in-situ observations, remote sensing retrievals, soil properties and reanalysis products, to generate a high-resolution multi-layer soil moisture product over China using the random forest model. The topic is of interest to wide audience and the data itself holds the potential to contribute to related research. However, the paper has several major defects and it should not be accepted unless they are reasonably resolved. The following lists my main and minor comments:

Major comments:

- The paper writing is extremely poor. Language errors and statement repeats or inconsistency are found through the whole manuscript. For example, "dataset of China" in the title should be "dataset over China" or "dataset for China"; in Lines 14-15, the authors have stated "high quality gridded soil moisture products" are "usually available from remote sensing... with coarse resolution" but then they raise that "high quality" is characterized by "high-resolution...", which is obviously contradictory; there are many other cases like the usage of "... is acted as". More importantly, the Results part contains many statements that are actually discussion while the Discussion parts contains too many results. Additionally, much discussion in the Results and Discussion sections actually lacks sufficient evidence support (e.g., Lines 361-362).
- The soil moisture product has a spatial resolution of 1 km while the input data, ERA5-Land product, has a resolution of 9 km. So how did the authors pre-reprocess the ERA5-Land data? They mentioned that in-situ observations were adjusted to ERA5-Land soil moisture but did not introduce the specific methodology.

Minor comments:

- The soil moisture product ranges from 2010-2020 but this time coverage is still too short for analysis in related fields, for example, the occurrence of droughts. I am wondering why the authors chose such a target period.
- In the text, the authors mentioned terms such as "Liaoning province", "Sichuan province" and "the plateau", which are not friendly to readers that have no the background knowledge.