Earth Syst. Sci. Data Discuss., referee comment RC1
https://doi.org/10.5194/essd-2021-61-RC1, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on essd-2021-61

Anonymous Referee #1

Referee comment on "A harmonized global land evaporation dataset from model-based products covering 1980–2017" by Jiao Lu et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2021-61-RC1, 2021

General:

This is a highly interesting and relevant dataset that might prove useful for numerous studies. Given that it claims to be a benchmark dataset more detail and discussion is necessary.

I would expect more detail on the validation part and the role of GLEAM as an independent reference dataset. The motivation for the latter needs to be discussed more honestly including the shortfalls of this approach.

A more detailed validation might reveal that the merged product does not necessarily outperform the other datasets at all sites/time periods (which is okay). At the moment only average global validation metrics are discussed.

Language:

Please revise the language before publication, best with a native speaker, e.g. “Land evaporation (ET) plays a crucial role in the hydrological and energy cycle.”, the first sentence of the abstract. Next sentence: What are numerical products? The lack of …

“were distributed across the east-west direction banding manner” hard to understand.

Please also check the language of the data described on the data portal.

Major:
Please add a lot more details on the validation across different Fluxnet sites, ecosystems, years etc. How does data availability of in situ data in different years affect the validation? Can a significance test be included? Where (e.g. ecosystems, latitude, longitude) / When does the merged product perform better than the individual model output, where/when worse or similar?

The role of GLEAM needs to be clarified! What is the motivation of using a fourth model as an independent reference dataset? What is the influence of GLEAM on the final product? The weight map is I assume largely influenced by how the three models agree with GLEAM, but why is this objective? Could the merging have been done using in-situ data to determine which model performs best too?

GLEAM (a) actually uses ERA-Interim as input and will therefore be highly correlated with ERA5, also all reanalysis products will be quite correlated and are definitely not independent as they will assimilate millions of the same observations even if the models are different.

Again, this really needs to be clarified in the manuscript. One might argue that the resulting dataset uses the well-performing GLEAM dataset as a reference to compute the weights, however, one can not argue at all that GLEAM will be correct, or even superior to the other datasets across all pixels/time periods!

How in the end does this method compare to other merging methods, at least qualitatively? Add this to the discussion/conclusions.

The conclusion interestingly mentions the temporal nature of the errors, this is however not really mentioned in the main part unless I missed it. This is a very interesting aspect.

Minor:

L41: "from in-situ observations and satellite inversion" Which inverse satellite retrievals exist for evaporation? I'm wasn't aware of any and am curious. Add examples/citations.

L48: Can there ever be a global benchmark dataset if there is no in-situ data available everywhere? In the end, any dataset will be a proxy in these areas.

L62: “However, the practical application of maximized R method is usually found limited due to its use of only two most relevant in given data sources.” I couldn’t understand this sentence

Can paragraphs from 51 and 58 be combined? They seem to partly repeat the same issue.

L64: Is MSWEP a good example of a merging method? It’s the one I read about the most, but I’m not an expert in precipitation merging.

L69: Diversified methods?

L73: “determine the conditional density” difficult to understand if not familiar with the method, is there an easier way to say this?

L74: Is computational effort really a problem for these methods in the age of HPC and cloud computing? Just a thought.

L81: I think TC also requires strictly independent datasets and normally distributed errors which is seldom the case? Please double-check.

L88: Compared to which “simple method”?

L90: Which standards?

L91: “The REA method also produces a quantitative measure of reliability, which increases the overall reliability of simulation changes.” Didn’t understand this.

L99: “However, studies on the application of land evaporation are few. Which ones?

L109: Why is GLEAM the only independent dataset? Actually, the “a” version of GLEAM uses reanalysis data as inputs and therefore will be highly correlated with the others, especially ERA.

L105 contradicts L111: Is GLEAM merged or not?

L116: “the parameterized physical process” Which ones? A few more lines might be helpful here. Which extensive remote sensing observations? Actually, the “a” version uses a lot of model input data. There is an updated v35 version available just for information.

L144: Why monthly data? This would have a significant impact but is not really mentioned anywhere else.

L155: Why only monthly? How does this affect the method, already mentioned here?

L179: “based on GLEAM” What is meant by this?

2.2.1 Should this be "Coefficient of Variation"?

L201: So far this raises the biggest question mark for me: Why is GLEAM used as an independent reference dataset. Firstly, it will be highly correlated with the other datasets as it relies on similar inputs, especially the (a) version used here which is forced by ERA-Interim. Although it is a model specifically designed to estimate evaporation, whereas the other models are optimised towards a plethora of variables, it is, in the end, a similar, albeit more simple and specific, model.

L222: “and ensemble mean” Clarify that ensemble mean refers to the mean of the three products (it’s quite a small ensemble).

L223: “GLDAS Noah 2 ET is more than 20% higher than the ensemble mean, while ERA5 ET is almost the same with it, even MERRA2 Et more than 20% lower.” Please rephrase the sentence.

L228: “In order to reduce the risk of ...”. Which risk?

I’d better justify the choice of excluding some areas in the text. In the end, shouldn’t merging especially be important in areas with high discrepancies? As an extreme example: Where the datasets all agree merging is really necessary.

Are the differences actually very big in absolute terms? Especially in North Africa E is close to zero anyhow.

L273/276: significantly in a statistical sense? Did you compute significance? Could be good to add this.

L295: Which NDVI product was used for this analysis?

L322: The other models include soil moisture too, no? Please double-check.

L348: Why wasn't the dataset taken into account for this period? This could strongly affect the entire methodology and should be communicated more upfront. Maybe I missed something …

L405: 0.5 degrees?