

Earth Syst. Sci. Data Discuss., referee comment RC2
<https://doi.org/10.5194/essd-2021-250-RC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on **essd-2021-250**

Anonymous Referee #2

Referee comment on "A global long-term (1981–2019) daily land surface radiation budget product from AVHRR satellite data using a residual convolutional neural network" by Jianglei Xu et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-250-RC2>, 2021

This paper produces a long-term dataset of net surface radiation from AVHRR TOA measurements and MERRA2 reanalysis spatial context using a neural network algorithm trained to surface observations from multiple global networks. One of the unique aspects of this dataset is that an extended triplet collocation method is used to determine when surface measurements should be representative of a 0.05 degree grid box and only reliable sites are used to train the model. The dataset performs very well, even better than CERES-SYN and GLASS satellite datasets, against surface observations except in cases with ice and snow. The method using spatial context is also quite effective at removing algorithm biases that impact other datasets due to incorrect assumptions about atmospheric conditions (e.g. clouds and aerosol).

The authors also identify the need for future work to improve the algorithm over ice/snow, fill in gaps when satellites were down, and suggest that temporal variability information could improve the algorithm further.

Major comments:

- One of the real advantages I see with this dataset is the long record—since the dataset starts in 1981 and has an accuracy equal to or exceeding other satellite-based estimates, this extends observation-based estimates of the surface radiation budget significantly. That could be of significant value for long-term climate studies. The authors could highlight this advantage more strongly in the abstract and conclusions.

- Evaluation and training are done against multiple networks, but some of these networks are interconnected, for example, some ARM and all SURFRAD sites are included in the BSRN. As I look at the list of sites in Table S1, it appears that some of these stations are included multiple times. For example, BSRN_DRA is the same station as SF_DRA because the SURFRAD Desert Rock stations is submitted to the BSRN global network. This is particularly a problem if any of the independent validation stations are also included in the training dataset. Please look into this duplication.
- I am curious whether the results shown in Figure 7 reflect the fact that some of these networks are included in training the AVHRR dataset. It isn't clear to me from the description whether training stations were also used in this analysis, or whether this only includes independent testing stations and stations that didn't meet the reliability requirements. But even if these validation stations are independent from training data, the network of measurements around the ARM Southern Great Plains sites, for example, may be more similar to each other than a site that is located in a much different climate regime (e.g. independent sites ARM_E06 and ARM_E41 sites). That could lead to overfitting. It would be helpful to understand how independent this validation dataset is.
- Figure 9 is really interesting! I found it quite compelling that the AVHRR dataset shows smaller biases and relationship to atmospheric factors than the GLASS dataset. That strongly suggests that the neural network approach is able to accurately capture variability related to the surface measurements in a more reliable way.
- Does Figure 14 show local time? Please label for clarity.

Minor comments:

Line 50: "RT-based physical methods show a great generalization" I am not sure what this phrase means, please revise for clarity

Line 310: should it be: "data was *then* removed"?

Line 346: "for in surface radiation estimations." Wording doesn't seem quite right here.

Lines 360-361: This sentence is awkwardly written and should be revised. Changing consistently to consistent, and site to sites would improve readability.

Line 483: should be very instead of "vary"

Line 545: should "produced" be replaced?

Line 563: GLASS is misspelled GALSS

Line 572: I think that 7.08 must be 0.78. Please check.

Line 690: I think "satellite replacement works" should be satellite replacement work if you are referring to times when there is no satellite data because it the satellites are being worked on.

Line 697: should be "covered surfaces".