Comment on essd-2022-83

Anonymous Referee #2

Referee comment on "A global dataset of spatiotemporally seamless daily mean land surface temperatures: generation, validation, and analysis" by Falu Hong et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2022-83-RC2, 2022

This paper describes an improved annual and diurnal temperature cycle-based framework method to generate global spatiotemporally seamless daily mean LST products from MODIS data with the support of reanalysis data. The developed dataset performs very well against global in-situ surface observations. Overall, this new method produces a 0.5 --degree daily product of daily mean LST over the globe. Given that this data has high spatial resolution at a daily time scale, it should be a useful tool for climate studies after its flaws are addressed.

Major comments:

1. The developed GADTC product has a spatial resolution of 0.5-degree, how to deal with the scale mismatch between the in-situ measurements and the product, the validation can be carried out at a higher spatial resolution, such as MODIS original resolution. Maybe, the authors can classify the in-situ sites to different levels according to the spatial heterogeneity of the site, to further analyze the errors at different sites.

2. The Surfrad site only has 7 sites, Why not merge the data from the Surfrad and Fluxnet networks when validating the Tdm product. Also, in section 5.1, the ΔDTR can be obtained using the Surfrad and Fluxnet data together.

3. The authors used MAE and bias, why not use the RMSE, which is typically used in the LST validation.

Minor comments:

Line 67, some latest papers about the C6 MODIS LST accuracy can be added, such as DOI: 10.1109/TGRS.2020.2998945, https://doi.org/10.1016/j.jag.2018.04.006

Line 104, the MxD11C1 was derived using the day/night algorithm and giving a reference

Line 139, how to get the hourly values?

Line 319, Scenarios #1 and #3, How many sites per scenario, the results can be analyzed by scenario, not by Surfrad and Fluxnet.

Line 360, Fig.8, combines data from the two networks.

Line 373, how to prove the large errors at these sites are related to the high spatial heterogeneity