Review of the paper “Global hourly, 5 km, all-sky land surface temperature data from 2011 to 2021 based on integrating geostationary and polar-orbiting satellite data”

General Comment:

This manuscript describes a new all-sky Land Surface Temperature dataset that extends the regional all-sky LST developed by Jia, A. et al (2022). The new dataset provides global hourly coverage based on CGLS and MODIS LSTs. The improvement with respect to both these products is clear and was well explored, both in terms of validation statistics and spatial coverage. The methodology is reasonably well explained, although some details about the inputs of the scheme (i.e., radiative fluxes and remaining inputs of the SEB; the usage of CGLS LST uncertainty in the assimilation step) and remaining limitations (e.g., its use in NRT) could be more detailed for the sake of completeness. The manuscript is very well written and structured. I would like to congratulate the authors for the great work, both in terms of the development of the method itself and the extraordinary effort that was put into its validation, which unfortunately hasn’t been very common in manuscripts about this particular subject. However, I found some inconsistencies when playing with the dataset itself, which I think should be addressed and explained before the manuscript is published. They are related to data format, metadata and unexplained spatial inconsistencies.

Therefore, I recommend the publication of this manuscript, after some major issues are addressed, as detailed below.

Specific comments:

L53 - Jiang et al., 2015 reference missing in bibliography

L75 – ERA5-Land is missing in this list (and is cited below in L485): Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., ... & Thépaut, J. N. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. Earth System Science Data, 13(9), 4349-4383.

L259 – The meaning of δ is not explained

L269-271 – If I understood correctly this is a limitation for the implementation of this method in NRT (as data up to d+15 would be needed). This should be stated somehow.

L287 – This is incorrect. CGLS LST comes with an uncertainty estimate per pixel which could be used in this context. It accounts for uncertainties due to GSW regression model error (which generally increases towards higher viewing angles and moister atmospheres), sensor noise, surface emissivity and total column water vapour. Uncertainties are generally below 3 K, so you could be giving too little weight to the retrieval.

You should detail a bit better how you introduce the shortwave and longwave products into the model. Please provide more details about the CERES products and other SEB inputs for clarity. For instance, I would suggest including details on the retrieval frequency, spatial resolution and limitations / uncertainties of those products, and a few sentences briefly describing the algorithms. These are critical inputs for the surface energy balance model, so it is important to highlight their strengths and weaknesses.

Could you please provide a description of the validation statistics that were used? It's not clear if some kind of cloud filtering was used (I am guessing none was used). Please use the statistics recommended by the CEOS-LPV Land Surface Temperature Product Validation Best Practice Protocol https://lpvs.gsfc.nasa.gov/PDF/CEOS_LST_PROTOCOL_Feb2018_v1.1.0_light.pdf , or discuss why you decided not to use them.

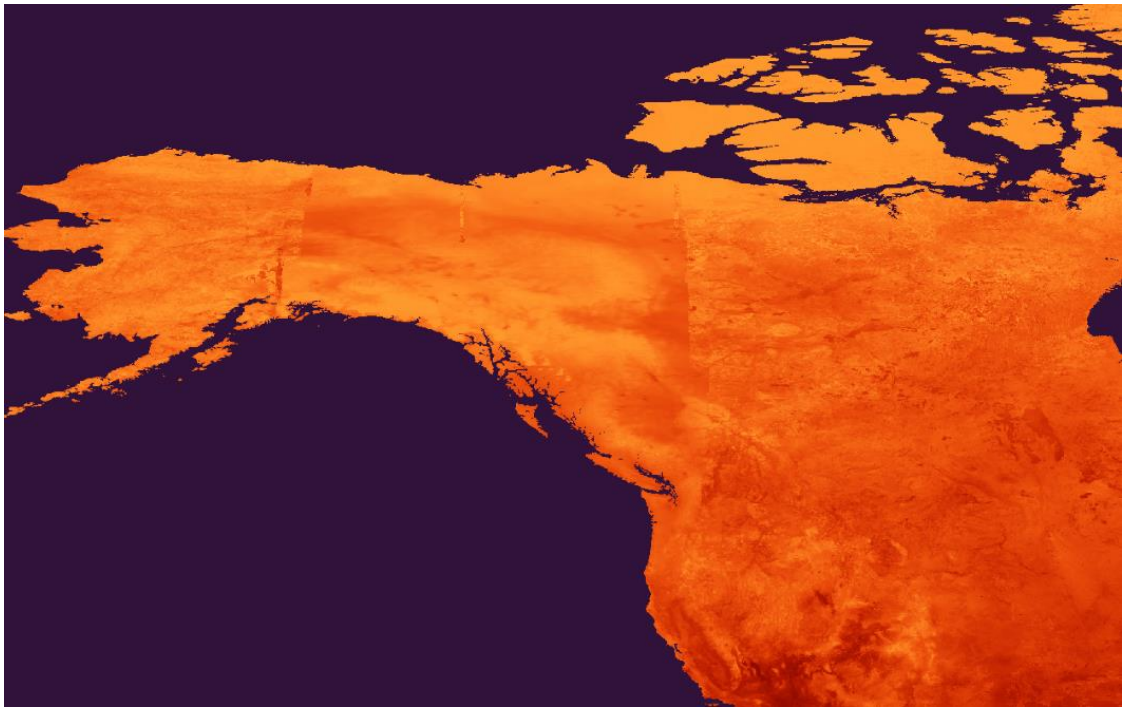
L353 - Using a nearest neighbour here introduces some additional error into the GHA-LST statistics, especially in times of day where the heating rate is higher. A more suitable way to minimize that is to interpolate LST values from consecutive hourly observations/ estimates to the MODIS acquisition time.

Fig5 – please homogenize the x and y scales in the plots

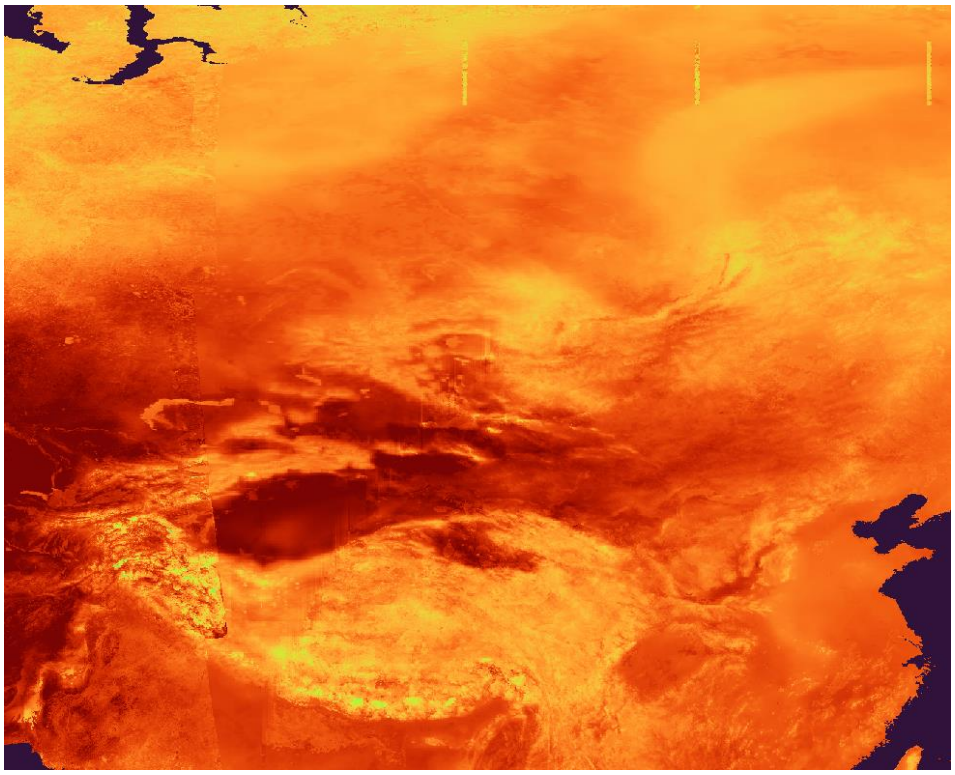
In section 3.2 I would have like to see some more information extracted from the large number of in situ sites used for validation, for instance the results discriminated by land cover type or even Koppen-Geiger climate zones. Its's also useful to have access to the individual in situ statistics, which could be provided in an annex.

(page 21) Regarding the issue of spatial consistency, as required by the Review Guidelines of the Journal, I've randomly selected a few timeslots to inspect the data.

- In the case of the GHA-LST_2018230_UTC09.tif, I found some strange vertical stripes and some changes in the detail of the spatial texture over Alaska:



- And a zoom over Central Asia, for the same timelot:



- This kind of features is common across other timeslots: there are areas with lot of finer scale structure and then abrupt changes to wider smoother areas. I think it should be acknowledged as a weakness of the product and the causes should be further explored. In fact, in L471, the opposite is suggested: “At the hourly scale, the GHA-LST map (Figure 9c) can also produce the reasonable spatial variation in LST without artificial textures”
- I would also recommend releasing the dataset in the NetCDF4 format, as it would reach a wider audience (in particular the climate modelling community).
- I couldn’t find the way to convert from int32 to physical values (i.e., where is the scaling_factor and add_offset values we must use to produce actual LSTs from the information within the geoTiffs?).
- I would also like to be able to isolate clear and cloudy situations in the analysis, would the authors be able to provide a cloud mask together with the product itself as an ancillary variable?
- There is no uncertainty information, as required by the journal guidelines.

L572 – I strongly disagree with this speculation; as the authors say across the manuscript, most of the lower accuracies over mountainous regions can be explained by lower site representativity; therefore MW sensors with larger footprints will not help correcting those issues. This was discussed in Martins et al (2019).

L577 -and maybe soil moisture?

L750 – Please check this reference, there is a typo.

João P. A. Martins, PhD.

30/09/2022