

Comment on essd-2021-6

Anonymous Referee #1

Referee comment on "A new global gridded sea surface temperature data product based on multisource data" by Mengmeng Cao et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-6-RC1>, 2021

Summary

This is a review of "A New Global Gridded Sea Surface Temperature Data Product Based on Multisource Data" by Mengmeng Cao¹, Kebiao Mao, Yibo Yan¹, Jiancheng Shi, Han Wang, Tongren Xu, Shu Fang, and Zijin Yuan. The authors have merged the sea surface temperature (SST) data from multiple sources to create a new high-resolution global dataset of monthly SST. The new data product does not contain any missing values and has been shown to be more accurate than the unmerged datasets. However, I have some concerns about the usefulness of the new dataset and the assessment of accuracy (see general comments). Moreover, the authors may need to emphasize the novelty and the uniqueness of the methods used to create the new data set.

General comments

- The conclusion of the better performance of the new dataset is drawn based on the smaller differences between the new data set and the in-situ observations (iQuam) than between the original datasets and iQuam. Should it simply because the iQuam data is used to create the new dataset? If iQuam is supposed to be the closest to the truth (as it is used as reference to validate other datasets), then why researcher do not use iQuam but use the new dataset created by the authors? I think author needs to better justify it. Missing values generally do not have a significant impact on the statistical analysis of climate. Moreover, the missing values can be obtained by interpolating other data using simple interpolation methods without losing much accuracy.
- I wonder how accurate the new dataset is compared with global reanalysis such as ERA5. The global reanalysis may not have such a high-resolution as the new dataset. But is it possible to compare the new dataset with some regional high-resolution reanalysis? The reanalysis is created using both information from observations and model simulations. Therefore, should one expect a reanalysis product to be more

accurate than the new dataset generated using only observations?

Specific comments

- Line 109-110: Why need to correct the observation time difference if the new dataset is monthly data. Isn't the difference measured in days?
- Section 2.4: Suggest separating the section of data and method.
- Line 224: Figure caption in wrong place.
- Figure 6: The scatter points for B looks strange (different behavior than the pdf and the box plot). Are the box plots already enough to demonstrate what you want show?
- Eq. (8): Suggest using bold font for vector and matrix.
- Eq. (9): $X_{\{t\}}$ looks a bit uncommon, suggest $X_{\{t\}}$ or $X(t)$. But it is the author's choice.
- Eq. (10): I don't really understand this equation, if Eq. (10) holds, then Eq. (11) shouldn't be correct.
- Eq. (12): Should it be $(P+R)$. Additionally, they are matrices and should be write in such a way $P(P+R)^{-1}$.
- Fig. 15: the original SST data seems quite interesting, as it has a lower bound of around 9 Celsius.

I apology that I haven't checked the quality of data set. I don't have necessary tools to access the data with the given format on my laptop.