



Review of the manuscript entitled “A new global gridded sea surface temperature data product based on multisource data” by Cao et al.

Prakki Satyamurty (Referee)

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-RC2>, 2021

The manuscript describes the methodology and procedure adapted to produce a new global dataset with 0.041° spatial resolution of monthly SST fields. The authors use MODIS SST data as benchmark and many other supplementary and complementary data sets including in situ observations and those retrieved from AVHRR infrared sensors, and AMSR and Windsat microwave sensors are utilized for obtaining a fusion. Essentially, the values in the blank or missing and low quality pixels are replaced by values of in situ observations and those derived or interpolated through the processes of Optimal Interpolation and Kalman Filter. The missing and low quality pixel problems arise due essentially to three reasons: (1) Cloudiness, fog, sea ice and proximity to shore influence the SST measurement. (2) Different sensors have different responses, that is, different sensors observe a pixel at different hours of the day and sense the temperature at different depths. (3) Latitudinal position of the pixel and the angle of sight from the sensor. The improvements in the new dataset, in comparison with earlier dataset, are statistically quantified.

A reader who is not highly specialized in the fields (of remote sensing and statistical manipulation of geophysical data) finds the manuscript difficult to read and assimilate. There is a certain amount of repetition in the description which did not contribute to clarity.

The methodology section should be improved to offer more clarity. In many places, a lot of empiricism is found about parameters that cannot be measured directly or easily or with sufficient accuracy. To make the understanding easier, they should provide units for the variables in the equations, tables and figures. Also the figure legends need to be more complete.

For Climatologists and Oceanographers who wish to use the SST, without bothering to go into miniscule details of the elaborate processing procedure, the present product provides a more accurate dataset. The quality control statistics presented shows substantial improvements in the new SST product.

One fundamental question over monthly time scales: Do we require a spatial resolution of 4,1 km, especially in the open oceans? This high resolution SST perhaps helps coastal studies like upwelling and estuary biology.

On the whole the authors did a good and useful job. Some specific points to be considered are:

- How can GOTM produce accurate values while utilizing several variables, such as 2 m temperature, 10 m wind, sensible heat flux, latent heat flux, which have poor accuracy? Does the ECMWF reanalysis of these variables present the accuracy needed?
- Fig. 7: Why the non-null pixel frequency is low off the Peru coast, exactly in the ENSO signal region? Because, the region is covered by low cloud almost all the time. You can see that the ITCZ region and other tropical oceanic regions west of the continents also present low non-null frequency due to clouds, high as well as low clouds. Inclusion of these comments may enrich your manuscript.
- Make the difference between skin temperature and surface temperature clear.
- I agree that the differences in time and depth of observation have to be compensated. What guarantees that Eqs. 1 and 2 can fix these problems? l is empirical, m is the frictional velocity in the water. These parameters may introduce uncertainties. What is the sanctity of the formula in Eq. 3? As you said in the Discussion section, the procedure relies on the performance of GOTM.
- Why can't you use the diurnal variability from in situ observations, at different places and in different months, instead of relying on a model?
- Lines 428-429: You say "I" is identity matrix and immediately after you say "H" is an identity matrix.
- The procedure described in lines 430 through 458 needs some more clearer explanation.
- Tables: You better provide in the text expressions for the statistical metrics shown the tables.
- Figs. 12 and 13 call for authors' comments. Fig. 16: what do blank circles and filled circles represent? Fig. 17: Indian_R and Indian_E. Tell what they represent in the legend.
- L 683: By the expression "different surface depths" you mean "different depths in the surface layer"?
- Many uncertainties you mentioned in your discussion will remain uncertain for a long time to come. Rewrite the last sentence, L 700. At many other places too the write-up needs improvement. Some repetitions can be suppressed while more explanation and comments are needed in some places. The conclusion section can be merged with discussion section and some repetitions can be avoided.

Please pay attention to comment 7 above.