Dear editor and referees,

Thank you for your valuable comments on our manuscript. First, we would like to express our sincere appreciation for your professional and insightful remarks on our paper. These comments are all valuable and have helped us to improve the quality of our paper. We have studied each comment and have made revisions that we hope will meet with approval. Please find our detailed responses below. For convenience, we also attach a version of the manuscript with changes incorporated. Thanks again.

With our best regards,
Mengmeng Cao and co-authors

##

Response to referees

Response to referee #1

 # Summary:

This is a review of "A New Global Gridded Sea Surface Temperature Data Product Based on Multisource Data" by Mengmeng Cao, Kebiao Mao, Yibo Yan, 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.

Response: We would like to thank you for reviewing our manuscript. Your comments and good suggestions are very important for us to improve the quality of manuscript and dataset. We have carefully addressed all the issues raised by you and the response is presented below.

At present, there are three main methods for obtaining ocean temperature: The first is the traditional method, which obtains sea surface temperature through sea observation sites. The main advantage of this method is that it has continuity in time and is hardly affected by weather. The disadvantage is that the number of observation sites is limited and the space lacks continuity, especially in remote sea areas. The second method is to obtain sea surface temperature through remote sensing retrieval. Remote sensing has advantages in space, but lacks continuity in time. Remote sensing is divided into two inversion methods: thermal infrared remote sensing and passive microwave remote sensing. Thermal infrared inversion of sea surface temperature has a high accuracy and resolution, but it has a great influence on clouds. There are more than 60% of the area covered by clouds every day, so there are more 60% of the area missing values. Although passive microwave is less affected by clouds, the resolution is relatively low. Passive microwave remote sensing is affected by the land near the coast, and the accuracy of sea surface temperature inversion is not high. The third is to output sea temperature products through the assimilation model. This method relies on the accuracy of the input parameters of the assimilation model.

Although different methods are used to obtain ocean surface temperature, they actually represent temperature information at different ocean depths, and the observation time is also inconsistent. The sea temperature observed by traditional sites is deeper than the temperature observed by remote sensing. Even if they are all the temperatures retrieved from remote sensing, the temperatures retrieved from thermal infrared and microwave are from different ocean depths. The sea temperature observed by thermal infrared is the skin temperature, and the sea temperature observed by microwave is a bit deeper than the depth observed by thermal infrared. The sea surface temperature obtained by the assimilation model should also be different.

Thermal infrared remote sensing is currently recognized as the most accurate method for obtaining sea surface temperature in a large area. Therefore, thermal infrared remote sensing is usually used to obtain sea surface temperature. For the ocean, the effective sea surface temperature value obtained through thermal infrared remote sensing every day is less than 40% of the total area (as shown in Figure 1), which means that more than 60% of the daily data have no value every day.

When calculating the monthly average data, some data sets use the average temperature value obtained by dividing the effective days of a month by the effective days. For example, if a certain pixel has only 25 effective temperature values in a certain month, then the average temperature of this month is calculated by using the average value of these 25 valid days. Although the monthly average temperature map calculated in this way has few missing values, the average temperature error of some pixels is relatively large (as shown in Figure 2).

The highlight of our work is to make full use of the research foundation of the predecessors to traverse the MODIS ocean temperature data set to find the pixels with low quality data, and then use high-quality daily data and other multi-source data (including sea surface temperature retrieved from passive microwave and observation site data, etc.) to improve the accuracy of the data, including pixels with low data quality and missing pixel data.

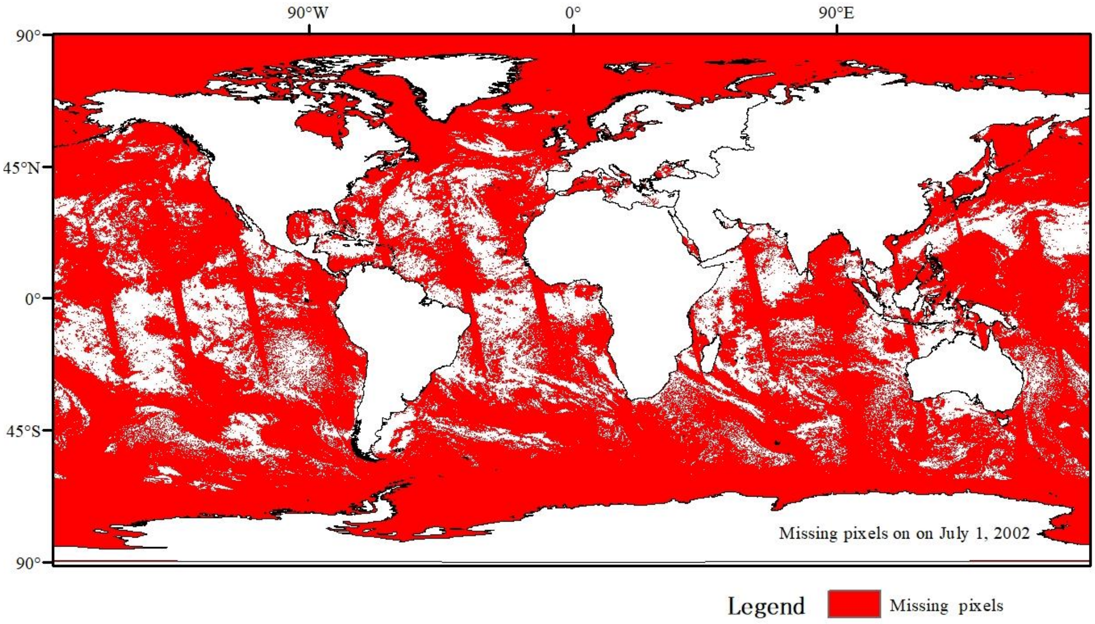

Figure 1. Distribution map of missing pixel values of thermal infrared remote sensing daily data (July 1, 2002).

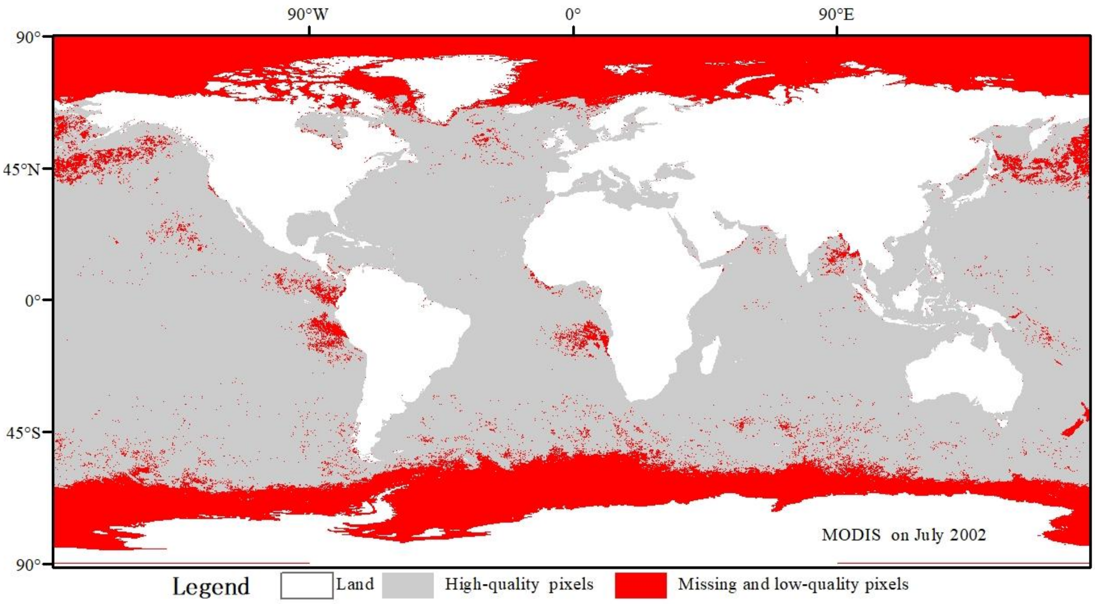

Figure 2. Distribution map of missing and low quality pixel of thermal infrared remote sensing monthly data (July, 2002).

General comments:

1. 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.

Response: Thank you for the valuable comment. It may be that we did not express it clearly in the manuscript.

We don't just use the in-situ observations (iQuam) data. In order to improve the low-quality and missing pixel values, we mainly use the sea surface temperature retrieved from microwave data, the sea surface temperature retrieved from AVHRR, and in-situ observations (iQuam) data, to fill in the relevant data by correcting consistency for time and space, and then part of in-situ observations (iQuam) data are used for verification. The purpose of the work is to build a long-term series of global major meteorological disaster remote sensing data sets with high spatio-temporal and consistency based on the current global multi-source remote sensing data and ground observation site data, and to provide key ocean temperature parameters (such as sea surface temperature) for marine meteorological disaster forecasting models, especially rapid forecasts of marine disasters such as typhoons, and provide early warning services for global fishing vessels and merchant ships.

It seems that there are not many data with low data quality in Figure 2 due to computer display. In fact, there are still many missing values for local areas. If we make up for vacancies or low-quality pixels through interpolation, the local accuracy needs to be improved. Although the data obtained by interpolation may have little impact on global research, it is not enough for local time-space analysis, local weather forecasting and disaster prediction with high accuracy.

2. 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?

Response: Thank you for your comments.

ERA5 provides hourly estimates of a large number of atmospheric, land and oceanic climate variables, which cover the Earth on a 30km grid. Although ERA5 claims that uncertainties for all variables will be included in the data information, the uncertainties for some variables cannot be downloaded yet, so it is difficult for us to make a direct evaluation. However, we have known from our research team's analysis for reanalysis product in some regions of China, such as the accuracy of air temperature, soil moisture and other products in the western region is not very well, and some products need to be further improved. We think the main reason is that the observation sites in these places are relatively sparse, resulting in limited data provided to the assimilation model. As a result, the accuracy of assimilation products in some regions with insufficient observation data is not high.

The assimilation model is not a panacea. The first one requires a lot of input parameters. When some input parameters are missing or the accuracy is insufficient in some regions, the accuracy of some output parameters may be not well. In addition, the estimation of each parameter in the assimilation model uses some parameter estimation calculation models. These estimation models are not suitable for all situations. Different regions in the world are different in terrain and climate conditions. Especially when some abnormal situations occur, such as typhoons or heavy rains, the results estimated by the assimilation model may have a large deviation. Therefore, many different assimilation models have been developed. These models have their own advantages and disadvantages, and the accuracy is different in different places. In the case of missing input parameters or insufficient precision, such as thick clouds and heavy precipitation, the error will be relatively large. Therefore, many researchers have been constantly improving these models, including ERA5, which is also constantly evolving and improving (https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5).

US NOAA also produces ocean temperature products with a resolution of 30 km, which should be one of the most authoritative products, and they are updated and released every day. We use this product as a reference comparison. CAAS SST with a resolution of 4 km made by us are resampled based on the resolution of ERA5. Figure 3 is the distribution map of OIS SST obtained from NOAA in July, 2002. Figure 4 is distribution map of CAAS SST (July, 2002). Figure 5 is distribution map of ERA5 SST with a resolution of 30 km obtained from ECMWF (July, 2002). Figure 6 is a distribution map of the difference between CAAS SST and NOAA SST. Figure 7 is the difference between ERA5 and NOAA.

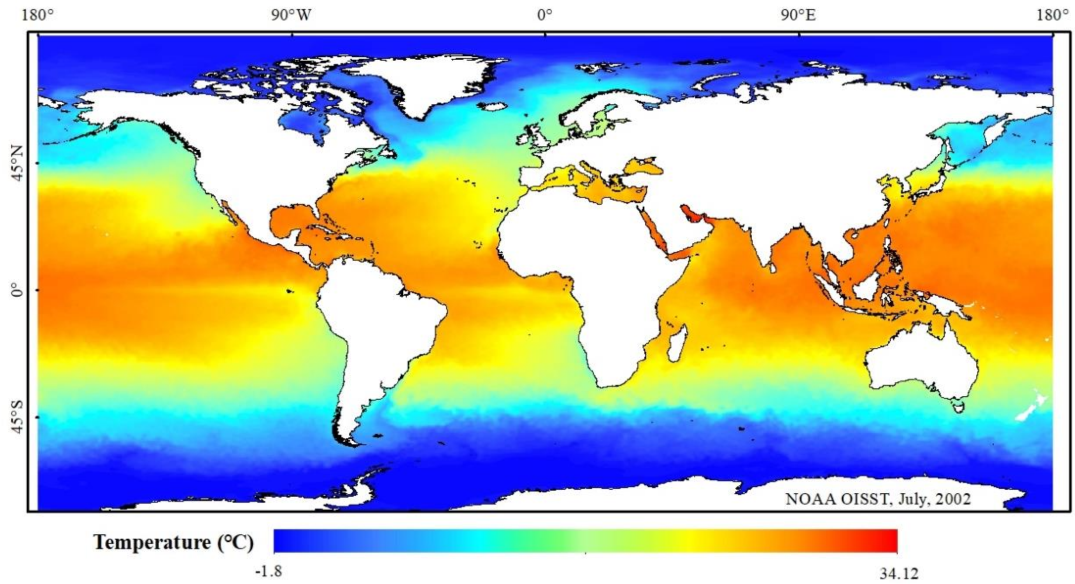

Figure 3. Distribution map of OIS SST obtained from NOAA (July, 2002).

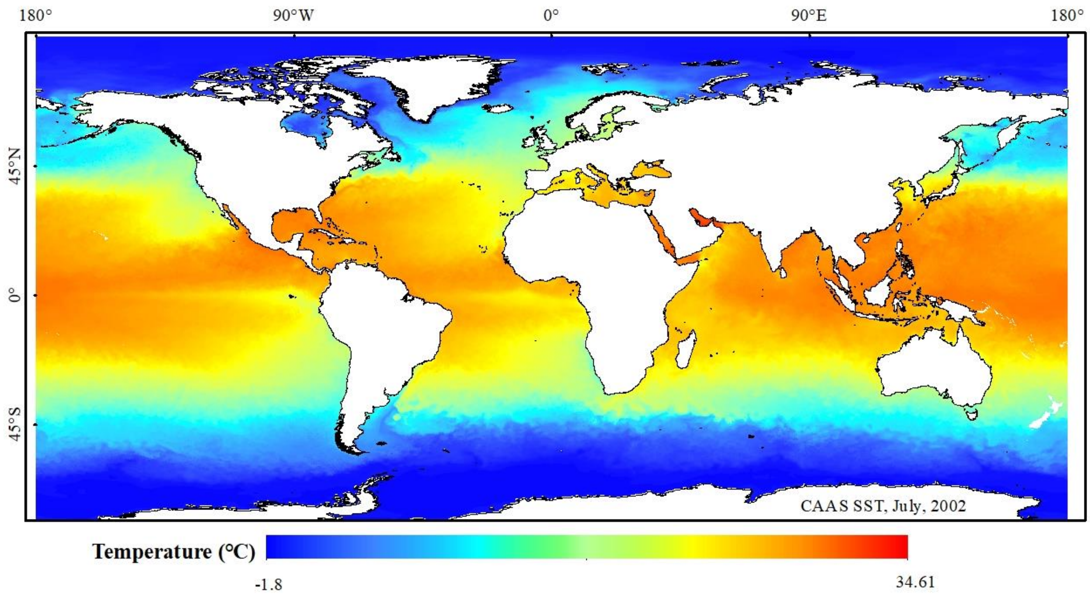

Figure 4. Distribution map of CAAS SST made by us (July, 2002).

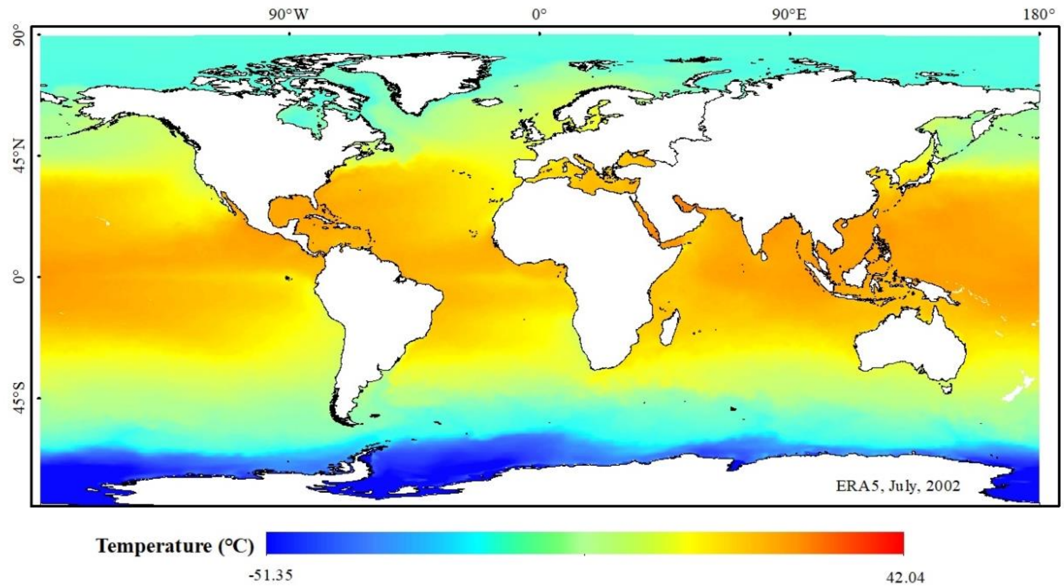

Figure 5. Distribution map of ERA5 SST obtained from ECMWF(July, 2002).

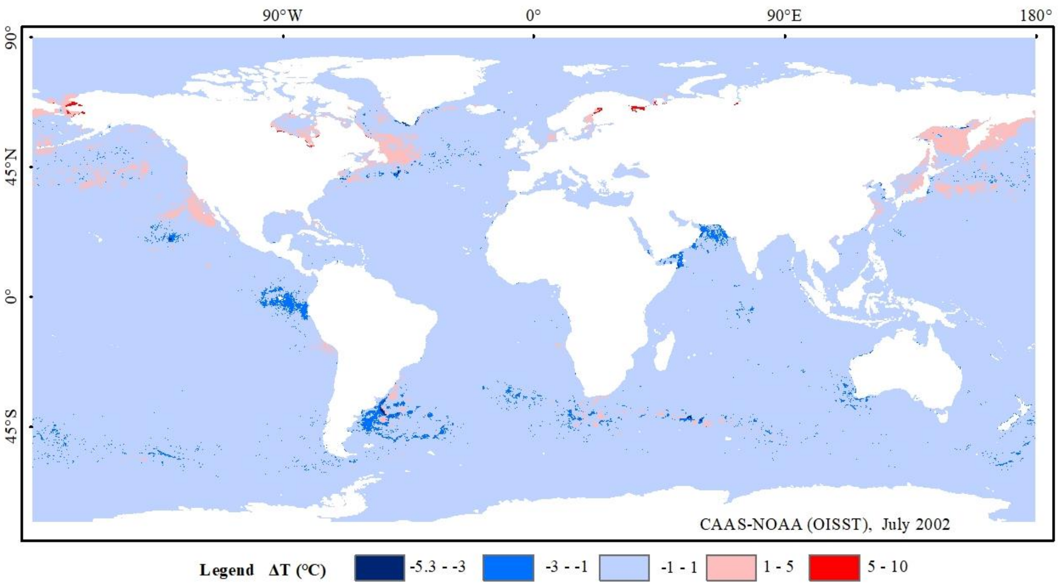

Figure 6. Distribution map of the difference between CAAS SST and NOAA SST

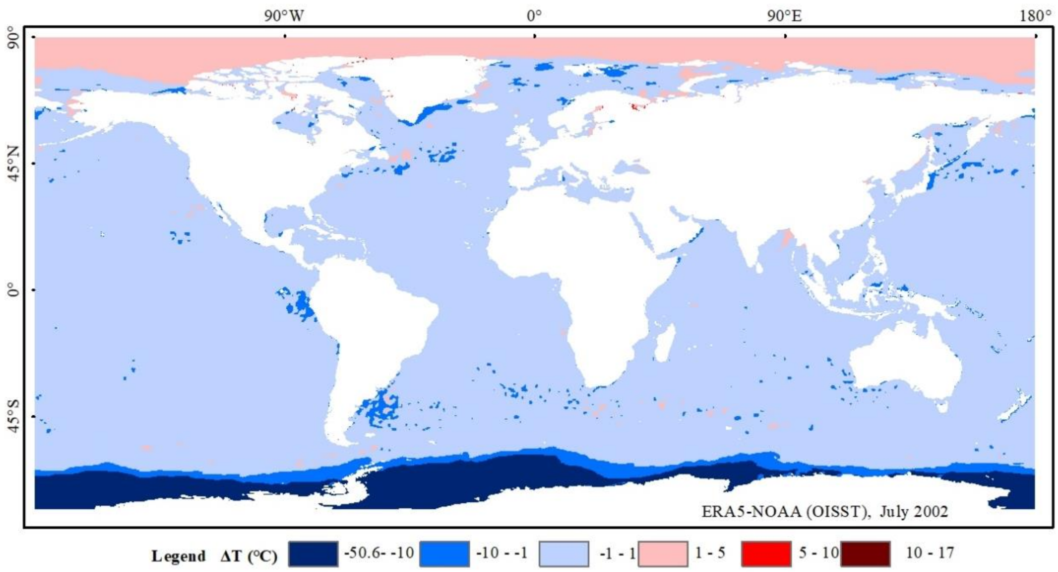

Figure 7. Distribution map of the difference between ERA5 SST and NOAA SST

Shown from figure 3-5, on the whole, the distribution trend of CAAS SST and NOAA SST ocean temperature products is not much different, and the distribution range of ERA5 temperature products is larger than that of CAAS SST and NOAA SST. We took NOAA SST ocean temperature products as a reference and made a distribution map of the difference between ERA5 and CASS SST and NOAA SST. Shown from figure 6 and 7. The regions with the largest differences are mainly distributed in the Antarctic and Arctic regions, as well as the regions with low MODIS product quality and no pixel values, which can be seen in Figure 2.

The difference between ERA5 sea temperature products and NOAA products in the Antarctic and Arctic regions is relatively large. The main reason is that there are less observation data in the Polar Regions, especially the low resolution of ERA5, and the polar region itself is not large. Due to the influence of mixed pixels, in places where the sea and land meet, especially where ice and water are mixed, the temperature of the polar coast or the ice surface is used as the pixel temperature as the pixel temperature value. We know that if the temperature of the ocean is too low, it will freeze. The resolution of remote sensing data is relatively high, and the ice along the coast of the polar regions can be distinguished. Therefore, at this point, high-resolution remote sensing observation data still has advantages. For other areas with large differences, we used microwave and AVHRR and site data to improve low-quality and invalid pixels. So theoretically, our accuracy should be slightly higher in these regions except that our data resolution is higher.

Re-analyzed data such as ERA5 is a data set (about 271 global products) produced by the assimilation model considering the global multi-factors, which has an overall advantage. However, when a single parameter or special parameter has a high spatiotemporal observation, the accuracy of the ERA5 product is not necessarily higher than that of the inversion data product. Especially in the local area when there is not enough input parameter information or high-precision parameter to input assimilation model, the accuracy of the observation data inversion should be higher than that of the assimilation model.

In addition, the advantage of our data set can be used for local research and analysis. The 30-kilometer resolution of assimilated data products cannot meet the requirements of high spatial resolution in local areas. Especially in offshore weather forecasting and analysis of near-shore agricultural disasters, we need to further improve the spatial resolution. This is also one of the main reasons why many countries have been continuously increasing ocean ground observation sites and improving the spatial and temporal resolution of satellite observations. Different data sets are not contradictory, and we can absorb each other's advantages, and further improve the accuracy of monitoring and forecasting.

 # Specific comments:

1. 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?

Response: Thank you for the valuable comment. The observation time of different sensors is different, and the calculation must be corrected to the same time to be comparable. For example, when the satellite transits at 10 o'clock, the observed sea temperature data is the temperature at 10 o'clock. The data of the sea surface observation station is 10:30, so we must calibrate the observed sea temperature at 10:30 to 10 o'clock and then calculate or compare. The calculated average must also be consistent in time to be comparable.

2. Section 2.4: Suggest separating the section of data and method.

Response: Thank you for your good suggestions. We have tried to make revisions.

3. Line 224: Figure caption in wrong place.

Response: Thank you very much for your advice. We have modified it in the manuscript.

4. 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?

Response: Thank you very much for your careful review. When we made this figure, we did not remove the outliers in the original MODIS data, so there were some large differences between MODIS SST and in situ observations. Therefore, we re-extracted the high quality SSTs and redrew the figure, which can be seen in line 346 of the revised manuscript. It (Figure 6 of the revised manuscript) is also presented below for the ease of reviewing.

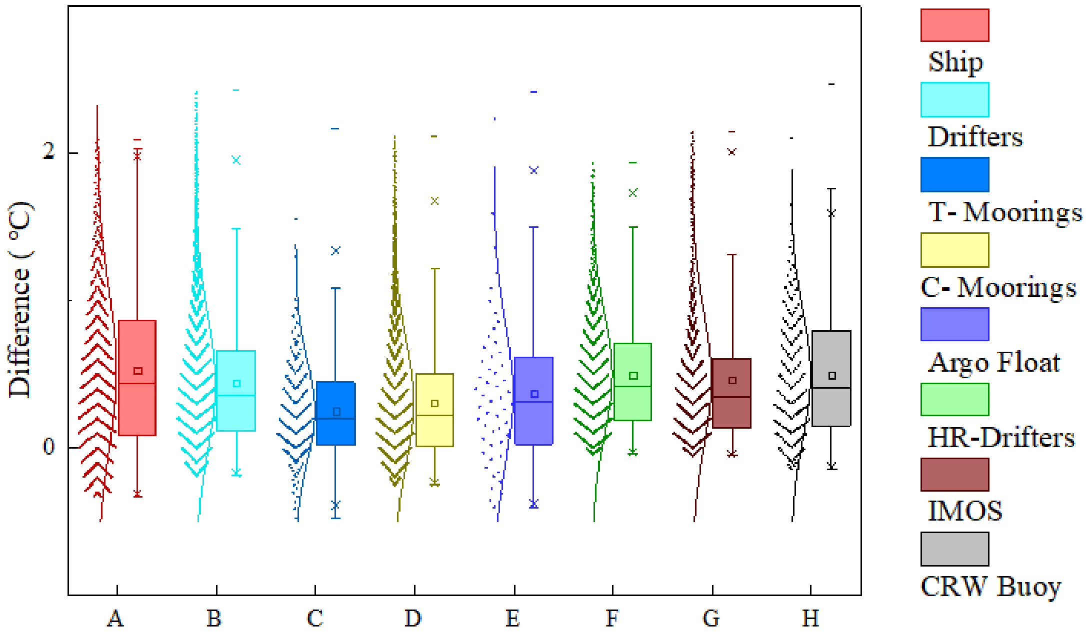

Figure 6. Box chart with scatters representing the differences between the original MODIS data and eight types of in situ SST observations.

5. Eq. (8): Suggest using bold font for vector and matrix.

Response: Thanks for your good suggestion. We have modified it, which can be seen in line 427 of the revised manuscript.

6. Eq. (9): X_{(t)} looks a bit uncommon, suggest X_{t} or X(t). But it is the author's choice.

Response: Thanks for your good suggestion. We have modified it, which can be seen in line 434 of the revised manuscript.

7. Eq. (10): I don’t really understand this equation, if Eq. (10) holds, then Eq. (11) shouldn’t be correct.

Response: Thanks for your good suggestion. We have modified it, which can be seen in line 443 of the revised manuscript.

8. Eq. (12): Should it be (P+R). Additionally, they are matrices and should be write in such a way P(P+R)^{-1}.

Response: Thanks a lot for pointing these out. We are sorry for our unclear expression and carelessness. We have modified it in the manuscript. It can be seen in line 446.

9. Fig. 15: the original SST data seems quite interesting, as it has a lower bound of around 9 Celsius.

Response: Thank you for the valuable comment. Most of the global ocean observing station data are distributed in low and middle latitudes (shown as figure 9), and relatively few in high latitudes. Data is very valuable. When doing verification, we mainly choose the site data in the low and middle latitudes, so there are not many low values. The biggest advantage of our data set is the high spatial resolution.

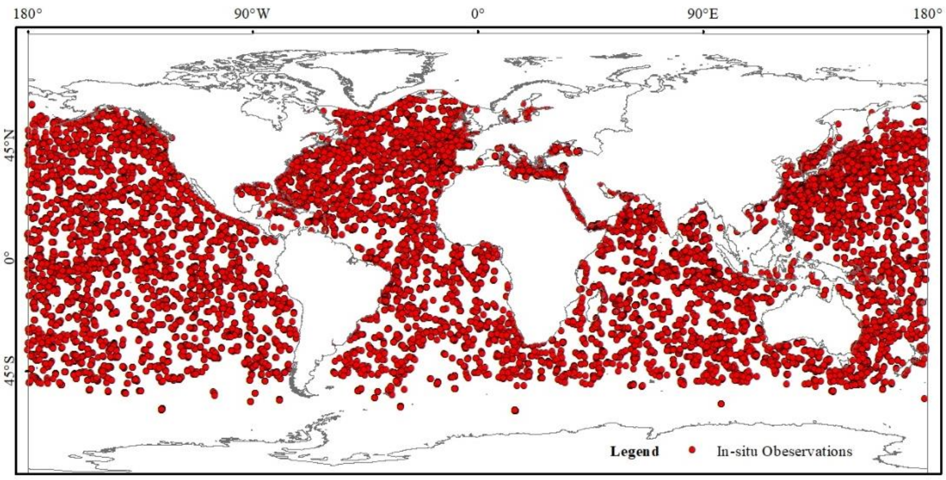

Figure 9. Distribution map of in-situ observation (iQuam) stations.

##

Response to referee #2

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.

Response: We would like to thank the referee for reviewing our manuscript. These comments are very important for us to improve the present manuscript. We have carefully addressed all the issues raised by the referee. Please find our detailed reply below.

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.

Response: Thank you for the valuable guidance. We have deleted these repetitive statements and revised some sentences that are difficult to understand.

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.

Response: Thank you for your guidance, and we have tried to make revisions.

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.

Response: Thank you for your positive evaluation and guidance. Your evaluation is very correct. The high-resolution ocean surface temperature data set is far more important for coastal studies than for open oceans studies. But the high-resolution ocean temperature data set is also helpful for us to capture temperature anomalies in open oceans, helping us to understand the ocean more accurately, such as the migration of central location of El Niño or La Niña.

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

1. 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?

Response: Thank you for your comment and guidance. You are right. When the accuracy of the input variables cannot be guaranteed, the GOTM simulation value will have deviations, such as 2 m temperature, 10 m wind, sensible heat flux, latent heat. In general, the use of ECMWF reanalysis of these variables can meet the requirements. But in special circumstances, when the input parameter deviation is too large, it will also cause a relatively large error. When there is a large deviation, we will make adjustments. For example, the difference between the sea temperature depth obtained by microwave inversion and the temperature depth obtained by thermal infrared inversion is less than 1mm, and the temperature difference is within 0.6 K, and we have tried to control this error through statistical methods. The associated explanation has been supplemented to the revised manuscript and can be seen in lines 309-317 of the revised manuscript. The relevant content is also presented below for the ease of reviewing.

"*In order to ensure that the corrections of depth and time are effective for each pixel, we calculated the difference range of high quality pixels for different SST data. Then, we manually checked the correction results of each invalid pixel, and determined the outliers according to the statistical difference range and other satellite SST data. Finally, these outliers were adjusted based on mathematical statistics. For example, to determine the temperature difference (Δt) between the skin surface temperature and sub-skin surface temperature of the pixel i of the MODIS data, we first calculated the high quality value of pixel of MODIS data and the microwave data at the corresponding time during the study period. Then we extracted the data of wind speed, cloud cover, humidity and other environmental factor corresponding to these values. Further, based on these environmental factors, we determined the SSTs corresponding to the environmental conditions at the moment when the outlier of pixel i appeared. Lastly, the average value of the differences between these high quality SSTs was Δt.*"

2. 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.

Response: Thank you very much for your guidance and good suggestion. We have added a brief description in the manuscript as follows. It can be seen in lines 354-358.

"*There are many missing pixels distributed off the Peru coast, exactly in the ENSO signal region, due to the widespread low clouds over the eastern South Pacific off the coasts of Chile and Peru (Satyamurty and Rosa, 2020). In addition, there is a lower frequency of non-null pixels in the Inter-Tropical convergence zone (ITCZ) region and other tropical oceanic areas west of the continents due to the cloud cover in these areas (Ackerman et al., 2008; McCoy et al., 2017)..*"

3. Make the difference between skin temperature and surface temperature clear.

Response: Thank you very much for your careful review. We have modified it in the manuscript as follows. It can be seen in lines 225-227.

"The thermal infrared remote sensor measures the sea surface skin temperature at a depth of 10-20 µm, while the microwave remote sensor can retrieve the sea subcutaneous temperature at a depth of 1-1.5 mm (Minnett, 2003; Minnett et al., 2011)."

4. 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.

Response: Thank you for the valuable comment. You are very right. Some parameters cannot be obtained in real time, especially in some remote areas. Some parameters in the formula are empirical in Eqs. 1 and 2, which will cause certain uncertainty. The reanalysis data such as 2 m temperature, 10 m wind, sensible heat flux, latent heat flux, etc. are used as input parameters for control, so the uncertainty of calculating the sea surface temperature will be controlled within a certain range, which can basically meet the current requirements. In addition, there is also microwave inversion temperature as a control condition. We try to improve the accuracy of sea surface temperature products as much as possible.

In our research, Eq. 3 is mainly used to determine the representative depth of the temperature observed by different observing instruments for stratification. Some parameters, such as reanalyzed data as input parameters, can still bring certain errors when the accuracy deviation is relatively large. Therefore, we need more and more high-precision input parameters, and this is also one of the main reasons why many countries have been continuously increasing the number of ocean ground observation sites and improving the spatial and temporal resolution of satellite observations.

5. 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?

Response: Thank you for the valuable comment. You are right. We originally planned to do this as you have said at the beginning, but we found that the number of data from the observed observation sites was very limited. Another important reason is that the data depth information of the observation sites is also inconsistent with the MODIS thermal infrared observation depth. We can still need to use the model for calibration. Therefore, we directly chose the model for calibration.

6. Lines 428-429: You say "I" is identity matrix and immediately after you say "H" is an identity matrix.

Response: Thanks a lot for pointing these out. We have modified it in the manuscript.

7. The procedure described in lines 430 through 458 needs some clearer explanation.

Response: Thank you very much for your careful review. We have modified it in the manuscript, which can be seen in lines 439-462.

8. Tables: You better provide in the text expressions for the statistical metrics shown the tables.

Response: Thank you for your guidance. We have made revisions, which can be seen in lines 539-546.

9. 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.

Response: Thank you for your guidance. We have modified it in the manuscript as follows.

1) Figs. 12 and 13 call for authors' comments.

"--- Figure 12 (a) shows the validation results of the AMSRE SST data, which shown that the overall result between the corrected data and the MODIS data presents a good linear relationship. The RMSE is reduced from 1.137 ℃ to 0.508 ℃ and the absolute bias is reduced from 0.718 ℃ to 0.302 ℃, which indicates that the model can simulate the SST at different depths well. Furthermore, we compared and analyzed the nighttime products of the same sensor with the corresponding daytime products after a time correction to verify the time correction performed by the model. Taking AMSRE daytime SST products as an example, we corrected these SST values to the corresponding nighttime SST values (Figure 12 (b)). Shown from figure 12 (b), there was an obvious diurnal warming before the correction, and the data after the correction had lower absolute bias and RMSE values. Thus, the model also can simulate the diurnal variation in the SST well and can be used to normalize the SSTs observed at different times.

--- It can be shown from figure 13 that the range of temperature difference between uncorrected in situ data and MODIS data is about -2 ℃ - 2.5 ℃, while the difference range between corrected in situ data and MODIS data is reduced to the range of -0.5 ℃-0.75 ℃."

2) Fig. 16: what do blank circles and filled circles represent?

Blank circles and filled circles represent original product and the reconstructed product, respectively. We have modified the legend as follows. Thank you

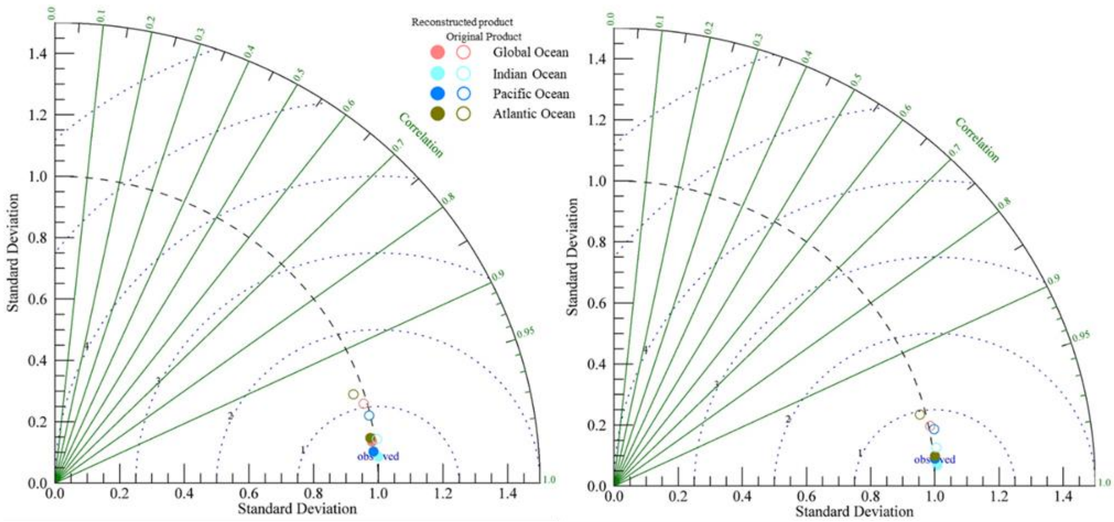

Figure 16. Normalized Taylor diagrams showing differences between matched SST from in situ data before (a) and after (b) correction and the corresponding SST products.

3) Fig. 17: Indian_R and Indian_E. Tell what they represent in the legend.

_E and _R are monthly mean SST changes of ERSST and new SSTs in different ocean from 2002 to 2019. We have made revisions, which can be seen in line 670.

10. L 683: By the expression "different surface depths" you mean "different depths in the surface layer"?

Response: Thank you for your guidance. We are sorry for our unclear expression. It refers to "different depths in the surface layer". When we modified the article, we found that this sentence was redundant, so we removed it. Thank you for your comment.

11. 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.

Response: Thank you for your guidance. We have removed some redundant sentences and combined the conclusion and discussion sections. We have also rewritten the sentence in line 700. The associated content can be seen in section 6 of the revised manuscript.

Please pay attention to comment 7 above.

Response: Thank you for the valuable comment and guidance. We have made revisions.

##

Special thanks are extended to you for your valuable comments.

We have tried our best to improve the manuscript and made substantial changes to the manuscript to correct certain shortcomings.

We greatly appreciate your help and hope that the corrections will meet with approval.

Once again, we would like to extend our sincere gratitude and appreciation for the valuable comments and suggestions.