Dear professor and Editor,

At this time when the news dominated by pandemic, I hope you and your family are healthy and safe. Thank you very much, and your comments and suggestions for modification are very good. We have tried to modify and emphasize your comments in our paper which are marked in red.

Global and regional surface temperature datasets are very important data for studying climate change, agricultural production, and urban heat island effects, and so on. At present, there are mainly two methods for obtaining global surface temperature data set. The first method is to directly use the weather station data for interpolation, and the other method is to use remote sensing data for interpolation. **Both methods have advantages and disadvantages. The first method has the problem of insufficient data volume and representativeness of weather station data. The second method is greatly affected by clouds and rainfall and the problem is still that the amount of available data is insufficient.** But there is no better method at present, which requires us further study. In order to improve the accuracy, on the basis of large-scale remote sensing data, we make full use of site data to obtain as more data information as possible, which improve the spatial and temporal continuity of the data. Thank you for give us guidance to our work.

Sincerely,

Kebiao Mao, et al.

A combined Terra and Aqua MODIS land surface temperature and meteorological station data product for China from 2003-2017 - revision 1

by Zhao, B., et al.

I reviewed the first version of this manuscript. Therefore I am referring to my first review regarding the summary of this manuscript.

First of all, I thank the authors for taking some of my concerns serious and for revising the manuscript with respect to some of the concerns brought up. I note that the authors did not answer all of my comments in the first review properly; some comments were not formulated clearly enough, it seems, so that these were misunderstood by the authors.

The topic of the paper is highly interesting and the data set behind has large potential. However, the description of data and methodology lack important and required details and needs to improved also with respect to the consistency of the wording and measures applied. The illustration of the re-construction method as a whole is very light and not satisfying. The application of the final bias-correction is neither well motivated nor does it appear to me that the final bias-corrected re-constructed LST data set is adequately evaluated because the 20% ground-observation data set aside are already used for the bias-correction. Finally, the results section lacks any example of the data and only focusses on trend analysis and interpretation of average conditions; hence it has to be restructured.

To my opinion this manuscript is not yet ready to go and requires major revisions.

Response: Thank you for your guidance. We have tried our best to improve the quality based on your opinion. We have used the data to make very detailed application cases from the monthly scale to the annual scale. The problem you mentioned is a common problem in the current large-scale research, which has not been solved well so far.

The traditional method to study surface temperature change at the regional scale in China is to interpolate using data from 2400 meteorological stations. This disadvantage is obvious, which is representative of the data from the ground observation station. Although everyone knows that the amount of data is not enough, there is no better way, and many people just use interpolation to obtain surface data.

Although remote sensing data sets can measure ground temperature in large areas, effective observations cannot be obtained where there are clouds and rainfall. In order to obtain continuous spatiotemporal data sets, people usually interpolate based on remote sensing data. This disadvantage is also obvious and the amount of data can still be insufficient.

In order to provide a set of high-precision spatial and temporal continuous Chinese regional temperature data sets, the only way is to obtain as much temperature data as possible. Although there is also the question of site representativeness that you mentioned, there is no other way to do better. We have made full use of the advantages of the two methods. Obviously, without comparison, this data set is theoretically more accurate than the previous method. Although there are still some shortcomings, more ground station data can only be obtained through technological progress.

At present, for global climate change research, there are mainly two methods for obtaining global surface temperature data set which are used for global climate change research. Please see references 1-17. The common problems raised by reviewers also exist in these studies, and there is no good solution now.

Traditional Methods

There are four major global indices of temperatures that incorporate station data. These efforts are led respectively by NOAA's National Climate Data Center (NOAA NCDC), NASA's Goddard Institute of Space Sciences (NASA GISS), a collaboration between the University of East Anglia's Climatic Research Unit and the UK Met Office's Hadley Centre (CRU), and the Berkeley Earth Surface Temperature group. They all primarily rely on the GHCN (Global Historical Climatology Network) for their input data. Their global temperature results all make use of the GHCN data collection in their analyses, which includes data from about 7000 stations [1-14]. The distribution of the observation stations is like Fig. 1, which is provided by NOAA's National Climatic Data Center. According to the introduction for data by the National Climatic Data Center (NCDC) in Asheville, NC, the data summaries provided are based on data exchanged under the World Meteorological Organization (WMO) World Weather Watch Program. For air temperature, the daily dataset (as available from each station) are only including mean temperature, maximum temperature, and minimum temperature.

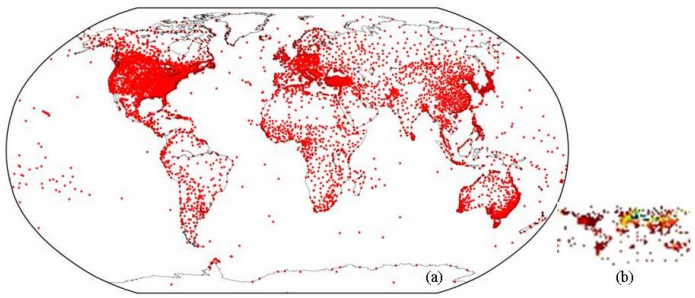

Fig. 1. (a). The distribution of the observation stations, (b) Global monthly mean temperature

Shown from Fig. 1(a), for land, although North America and Europe have large number of meteorological observation stations, the number of meteorological observation stations in other regions is not enough, especially for South America, Africa, Asia, ocean regions, and the Polar regions. Fig. 1(b) is the global monthly mean temperature map, and its data were obtained from GHCN (Global Historical Climatology Network) [15]. The total number of pixels is 2592 (36×72), the number of invalid pixels is 2112 which account for 81% of the global total pixels. Different research groups individually utilize different averaging techniques, quality control procedures, homogenization techniques, and datasets, thus these differences would cause uncertainty and obtain different conclusions[1-14].

References

1. Hansen, J., Ruedy, R. , Sato, M. & Lo K. Global surface temperature change. Rev. Geophys. 48, RG4004, doi:10.1029/2010RG000345 (2010).
2. Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B., & Jones P. D., Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *J. Geophys. Res.* 111(D12106).1-21, doi:10.1029/2005JD006548 (2006).
3. Smith, T. M., Reynolds, R. W., Peterson, T. C., & Lawrimore, J. Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880-2006). J. Climate 21, 2283–2296 (2008).
4. Ishihara, K. Calculation of global surface temperature anomalies with COBE-SST. Sokko-jiho [Weather Service Bulletin], 73 Special Issue, S19–S25, (in Japanese)(2006).
5. Rahmstorf, S. & Coumou, D. Increase of extreme events in a warming world. PNAS. 108, 17905–17909 (2011).
6. Hansen, J.E. & Lebedeff, S. Global trends of measured surface air temperature. *J. Geophys. Res.* 92, 13345-13372, doi:10.1029/JD092iD11p13345 (1987).
7. Hansen, J. E. et al. Global climate changes as forecast by Goddard Institute for space studies three-dimensional model. *J. Geophys. Res.* 93(D8), 9341-9364 (1988).
8. Hansen, J., Ruedy, R., Glascoe, J. & Sato, M. GISS analysis of surface temperature change. *J. Geophys. Res.* 104, 30,997–31,022, doi:10.1029/1999JD900835 (1999).
9. Hansen, J. E. et al. A closer look at United States and global surface temperature change. *J. Geophys. Res.* 106, 23947-23963, doi:10.1029/2001JD000354 (2001).
10. Hansen, J. et al. Global temperature change. PNAS, 103, 14288-14293, doi:10.1073/pnas.0606291103 (2006).
11. Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., & Kaplan A. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.* 108(D14), 4407, 1-22, doi:10.1029/2002JD002670 (2003).
12. Fan, Y., & Dool, H.V.D. A global monthly land surface air temperature analysis for 1948 – present. *J. Geophys. Res.* 113(D01103), 1-18, doi:10.1029/2007JD008470 (2008).
13. Jones, P.D., New, M., Parker, D.E., Martin, S., & Rigor, I. G. Surface air temperature and its changes over the past 150 years. Rev Geophys 37, 173–199 (1999).
14. Hansen, J., Sato, M., & Ruedy, R. Perception of climate change. PNAS. 6, 2415–2423, 10.1073/pnas.1205276109 (2012).
15. Lawrimore, J.H., Menne, M.J., Gleason, B.E., Williams, C.N., Wuertz, D.B., Vose, R.S., and Rennie J. An overview of the Global Historical Climatology Network Monthly Mean Temperature Dataset, Version 3. J. Geophys. Res., 116, D19121, doi:10.1029/2011JD016187(2011).

Remote sensing method

NASS has two polar-orbiting Earth Observing System (EOS) satellites (Terra and Aqua) in orbit at all times, with one satellite crossing the equator in the morning (10:30) and evening (22:30) and the other crossing the equator in the afternoon (13:30) and evening (1:30). The satellite sensors provide frequent global coverages which ensure the consistency of the retrieved surface temperature data and overcome disadvantages of Combined Land-Surface Air and Sea-Surface Water Temperature Index (Land-Ocean Temperature Index). The resolution of MODIS thermal bands is 1000 m. Wan et al have made a lot of good research work (especially for MODIS data), and the algorithms for retrieving surface temperature from MODIS are mature and NASA has provided the surface temperature product [16-22]. Since the observation time is symmetrical, the global mean surface temperature is more approximate to absolute mean surface temperature because of the high coverage of satellite data. However, more than 65% of the earth is affected by clouds and rainfall every day. It is difficult to obtain accurate data by remote sensing in these places. The problem of insufficient data can still exist.

16. Wan, Z. M. and Dozier, J.: A generalized split-window algorithm for retrieving land-surface temperature from space. IEEE T Geosci Remote., 34, 892–905, https://doi.org/ 10.1109/36.508406, 1996.
17. Wan, Z. M. and Li, Z. L.: A physics-based algorithm for retrieving landsurface emissivity and temperature from EOS/MODIS data. IEEE T Geosci Remote., 35, 980– 996, https://doi.org/10.1109/36.602541, 1997.
18. Wan, Z. M. and Li, Z. L.: MODIS land surface temperature and emissivity. Remote Sens. and Digital Image Proc., 11, 563-577, https://doi.org/ 10.1007/978-1-4419-6749-7_25, 2011.
19. Wan, Z. M., Zhang, Y., Zhang, Q., and Li, Z. l.: Validation of the land-surface temperature products retrieved from Terra Moderate Resolution Imaging Spectroradiometer data. Remote Sens. Environ., 83, 163–180, https://doi.org/10.1016/j.rse.2009.10.002, 2002.
20. Wan, Z. M., Zhang Y, Zhang Q., and Li Z L.: Quality assessment and validation of the MODIS global land surface temperature. Int J Remote Sen S., 25, 261–274 https://doi.org/ 10.1080/0143116031000116417, 2004.
21. Wan Z., New refinements and validation of the MODIS land-surface temperature/emissivity products, Remote Sens. Environ., 112, 59-74 (2008).
22. Otis B. Brown Peter J. Minnett With contributions from: R. Evans, E. Kearns, K. Kilpatrick, A. Kumar, R. Sikorski & A. Závody, MODIS Infrared Sea Surface Temperature Algorithm Algorithm Theoretical Basis Document (SST ATBD) Version 2, University of Miami, Miami, FL 33149-1098 (1999).

These two methods have their own advantages and disadvantages. The first method is guaranteed in time series and less affected by the cloud, but the number of sites is insufficient. The advantage of the second method is that it can be observed over a large area, but it is more affected by clouds. Through 20 years of temperature research, we found that if we can take advantage of both methods, which can improve the accuracy of global or regional temperature datasets. So we produced a combined Terra and Aqua MODIS land surface temperature and meteorological station data product for China from2003–2017.

General comments:

GC1: Still the paper is difficult to read and understand because the definition and naming of grid cells which are in one way or the other contaminated by clouds (or aerosols) is not consistent throughout the paper. It would be very helpful if the authors could put at one central place and for the entire paper a definition that they reconstruct the LST of grid cells they defined as missing. These grid cells might be defined as missing because a) for these simply no LST value exists [possibly due to thick clouds mostly but potentially also due to other factors which are - however -

not relevant for this paper] and b) the QA analysis provided an average LST error > 1 K or an average emissivity error > 0.01 - also potentially due to clouds [here perhaps only with a partial coverage of the grid cell but again this does not matter]. What matters is that there is still a suite (or zoo) of different names for the grid cells (or pixels) for which the LST is re-constructed. This needs to be harmonized!

Response: Thank you for your guidance. We have already made a definition for invalid pixel and low-quality pixel. Invalid pixel is defined when the value of pixel is missing, and low-quality pixel is defined as an average LST error > 1 K.

GC2: While the authors agreed that it would be important to illustrate and better describe steps of the LST reconstruction method in their reply to the reviewers' comments, they did not include any such information into the revised manuscript. Still the focus is on analysing the data set, showing trends and interpreting the data set. This is material for a geophysical paper. Earth System Science Data is a data set journal. It is neither like Journal of Geophysical Research nor like similar journals where geophysical interpretation is in the foreground of the interest.

Response: Thank you for your guidance. We have made some revisions, and flow chart in figure 2 is already very detailed. Most of analysis of the data set has moved the appendix as supplementary material. If the editor feels inappropriate, it can be deleted.

Therefore, I am still missing examples of how each of the three LST reconstruction steps 1) to 3) changes a LST time series for a specific month at a specific location. The reconstruction method is not illustrated adequately enough. It is not illustrated how a single LST value in-filled from a meteorological station might impact the GWR method in case that there are only 5 or 6 daily LST values in the 5 x 5 grid cell box used. It is not illustrated how the elevation-LST method fills in missing data. No investigation is presented which illustrates a potential different skill of the reconstruction methods during different times of the year. And, most severily, I don't find any high-resolution LST map which would illustrate how the three different steps of the re-construction process successively fill a LST map with a large count of missing data. In other words: The merit of the reconstruction method over the existing level 3 MOD11C3 and MYD11C3 products is demonstrated only to some extent. None of the trend analyses shown have been derived with the original product as well and compared with the results of the new product.

Response: Thank you for your guidance. We have made some revisions. The flowchart of figure 2 ((a)The summary flowchart for reconstructing MODIS monthly LST data, (b) The detailed flowchart for reconstruct missing daily pixels in (a).) has been drawn in detail, and the specific methods are also explained.

For large-scale remote sensing data on the land surface, almost every pixel is a mixed pixel, so verification is a very difficult problem, and so far there is no good solution. Because the surface condition of almost every mixed pixel is different, it is very difficult to verify. At present, for the verification of the surface temperature, two methods are usually adopted. The first method is simulated data verification. This method is mainly used to verify the retrieval algorithm, and the accuracy of the verification is the theoretical accuracy of the retrieval algorithm. This verification does not represent the accuracy of the actual product, because the simulated data assumes that the

pixels are endmember (pure pixels). Another method is actual data verification. Usually, a calm lake surface is selected as the ground measurement object, because only in this way the accuracy of surface data can be ensured (less affected by terrain and less affected by different types of surface). This approach is also to choose endmember (pure pixels) as much as possible. However, the land surface is almost mixed pixels, so it is difficult to obtain the true measurement value of the surface, and the verification results are relative. Because large-scale real data is difficult to obtain, it is difficult to draw a definite conclusion about the impact of other factors such as elevation, so we have not made too many descriptions.

On average, more than 65% of the world is affected by clouds and rainfall every day, and China's land is affected by clouds more than 45%. Our work is to make up for the missing values and the low-precision values affected by clouds and rainfall based on the original remote sensing data. We have already evaluated in section 4.

Another missing element of the methodology which radiates well into the entire manuscript is the brevity with which the about 2400 meteorological stations are described. These data are key to this paper - for re-construction, for the bias correction and for the evaluation. But the readers learn amazingly little detail about these stations, about what they regularly measure, how quality assessment is assured, and which surface conditions these stations represent. I had one major comment about this in my previous review and it seems that the authors accidentally ignored this missing piece of information. The description of the representativity of the station with respect to the MODIS grid cells (5600 meters x 5600 meters) is not included; neither did I find information about the biome and/or the topography for the more complex observation sites. I have seen too many papers showing wonderful results for the wrong reason. I cannot expect a comprehensive discussion of this issue in your manuscript. But I can expect that you spend more effort on this issue, because it is directly linked with the credibility of your data set, than you spend on interpretation and analysis of the data set - and I am well aware of that I am repeating myself here - which is material for a different publication.

Response: Thank you for your guidance. We have made some revisions. The problem you mentioned is a common problem in the current large-scale research, which has not been solved well so far.

The traditional method to study surface temperature change at the regional scale in China is to interpolate using data from 2400 meteorological stations, and even studying global temperature changes also has this problem. There is no other clever method, and the usual way is to interpolate whose disadvantage is obvious. Many people have made a lot of efforts and put forward various interpolation methods, but different interpolation methods have their own advantages and disadvantages, and the accuracy is different in different places or seasons. Although there are various comparative analyses, the significance is not very great. I think that Insufficient data measurement is the main reason.

Although remote sensing data sets can measure ground temperature in large areas, effective observations cannot be obtained where there are clouds and rainfall. In order to obtain continuous spatiotemporal data sets, people usually interpolate based on remote sensing data. This disadvantage is also obvious.

In order to provide a set of high-precision spatial and temporal continuous Chinese regional

temperature data sets, the only way is to obtain as much temperature data as possible. Although there is also the question of site representativeness that you mentioned, there is no other way to do better because we have made full use of the advantages of the two methods. Obviously, even without comparison, this data set is theoretically more accurate than the previous method. Although there are still some shortcomings, more ground station data can only be obtained through technological progress in the future.

GC3:

In Earth System Science Data data sets are introduced to potential users. I am missing a section which informs the reader about the data format and about the variables that are included in the product.

Response: Thank you for your guidance, and we have made some revisions. It has been provided at the place where the data was downloaded and the data file contains various information. Surface temperature is a very important parameter, which is needed for most geoscience studies especially for studying regional climate change and thermal environment changes, agricultural drought, crop yield estimation, ecosystems, etc.

GC4:

Again - despite my comment in the first review of the paper - you right-away jump into a discussion of annual-mean LST trends derived for China as a whole. At this stage the reader has still no clue how the reconstructed data set looks in comparison to the original data set. What, to my opinion, needs to be shown are maps of at least the monthly LST of the original product and of the new product. These should be supported by maps of the LST differences. Here you might want to zoom into critical areas (i.e. areas where QA of the original product indicates low-quality LST values).

All these trend analyses do not contribute to the evaluation of the product. For that they would have needed to be compared to other, independent products.

Response: Thank you for your guidance. The new data set is based on the original data, high-quality data is retained, and missing data and low-quality data are reconstructed. Therefore, in theory, the accuracy of the data is higher than the previous data. We have supplemented the comparative analysis, and actual verification shows that the accuracy has also been improved. The key is that the relevant research can directly use our data without having to re-do the data set.

In general, I don't understand why the flow of the results and discussion sections isn't like this:

1) Presentation of the product

2) Evaluation of the product

3) A few first examples (!) of the application of the product for trend analysis [for this journal possibly 2 pages maximum]

4) Discussion of potential caveats and limitations of use [this is for the user]

Response: Thank you for your guidance. We have made revisions.

GC5: You intend to use the independent 20% of the in-situ observations of the LST to evaluate the

product. However, instead of discussing observed differences (which are displayed and illustrated in a very global way without going into detail) based on physical principles you rather take the observed differences to perform a bias correction. One could provocatively say that your reconstruction methods hence does not work properly. In addition, now that you have used these independent LST observations for the bias correction - what are your independent observations to evaluate the final, bias-corrected LST product?
Response: Thank you for your guidance. There is no logical problem. After the data is filled and corrected, the data is evaluated and we get the accuracy. Through evaluation, some places will be overestimated or underestimated. We can correct it further, and the accuracy become higher, and we have not made any further evaluation.

Specific and editoral comments and remarks:

Abstract:
Lines 18/19: Please check sentence; it does not read properly.
Response: Thank you for your guidance. We have made revisions.

Lines 19/21: Neither "real" nor "true" appear to be the correct expressions here. About "actual" instead of "true" and simply delete "real"?
Response: Thank you for your guidance. We have made revisions.

Lines 23-25: Are these statistical measures valid / computed from the region mean LST values? Or are you referring to the statistical measured obtained for separate single station-to-LST-data-product-comparison within the regions? Please change the text accordingly so that this becomes clear. In addition, I recommend to explain what RMSE, MAE and Rsquared is.
Line 28: Please change "slope >0.10" such that it is clear that you write about an LST change of x Kelvin over time period y.
Response: Thank you for your guidance. We have made revisions.

Line 30: "exhibited" --> "exhibits"
Response: Thank you for your guidance. We have made revisions.

Introduction:
Line 36: "surface temperature" --> "LST"
Response: Thank you for your guidance. We have made revisions.

Line 38: You deal with LST in this paper while these "temperature changes" you are referring here to (with respect to glacier melt) are air temperatures, right? please connect these better to the LST (which should be 0degC over melting surfaces such as glaciers and snow).
Response: Thank you for your guidance. We have made revisions.

Line 41: "LST data ... measured" --> Either "LST is measured" or "LST data are obtained"
"by" --> "at"
Response: Thank you for your guidance. We have made revisions.

Line 42: "ground surveys" --> What is this? Are you speaking about specifically targeted campaigns or expeditions? Where can one find data of such "ground surveys" if these are not carried out in the context of regular measurements at the meteorological stations?
Response: Thank you for your guidance. We have made revisions.

Line 54: "throughout the world" --> "and its global coverage"
I note that you could and should have mentioned NOAA-AVHRR data as well here - particularly since with AVHRR data one can build a substantially longer time series than with MODIS and since data of that sensor are successfully used for sea surface temperature retrieval back to 1981.
Response: Thank you for your guidance. We have made revisions.

Line 58: "low quality values from undetected cloud-low-quality pixels" --> consider re-writing this part please; it is confusing.
By the way, if you intend to use this expression "cloud-low-quality pixels" you might want to define once what you mean by it.
Response: Thank you for your guidance. We have made revisions.

Lines 60-62: "Although the integrity of the data has been greatly improved" --> What do you mean with this? Further: "synthetic data" ? Further: "thus contain an insufficient quantity of daily LST pixels" Again, what do you mean with this? Perhaps you wanted to write something along these lines: "The number of missing or low-quality LST retrievals particularly high in the daily LST products. However, even in the 8-daily and the monthly LST products are subject to a considerable number of data gaps - particularly in regions and at time periods with persistent cloud coverage." ???
Response: Thank you for your guidance. We have made revisions.

Lines 62/63: "Cloud cover and other factors" appear to be several; therefore "cause", "reduce" and "pose".
Response: Thank you for your guidance. We have made revisions.

Line 64: One "low-quality" has to be deleted.
Response: Thank you for your guidance. We have made revisions.

Line 75: "only LST data" --> perhaps better: "only adjacent high-quality MODIS LST data"?
Response: Thank you for your guidance. We have made revisions.

Lines 78/79: "it cannot obtain enough information for reconstruction" --> perhaps better: "data coverage is too sparse for a reliable reconstruction."?
Response: Thank you for your guidance. We have made revisions.

Line 90 "improve" --> "improved"; "add" --> "added"
Response: Thank you for your guidance. We have made revisions.

Lines 92/93: "... clear-sky conditions instead of cloudy conditions, which cannot fulfill the need to obtain the real situation at the land surface" --> perhaps better: "... clear-sky conditions. However, clouds reduce night-time surface cooling and day-time surface warming due to solar irradiance. These effects are not taken into account using this assumption and therefore the derived LST values are likely biased towards clear-sky conditions."
Response: Thank you for your guidance. We have made revisions.

Line 94: "the real" can be deleted.
Response: Thank you for your guidance. We have made revisions.

Line 95: "is capable of penetrating clouds" --> This is definitely not true for the higher frequencies (near-90 GHz) and also not completely true for even the lower frequencies. Please reformulate this statement therefore.
Response: Thank you for your guidance. We have made revisions.

Study Area
You could, this is just a suggestion, provide some additional information about the average cloud cover in these six regions. With that you could underline right from the beginning where your approach potentially will have the largest added value.
Lines 114/115: "the key areas" --> "key areas"; "significantly" --> really? Or just substantially. How about stations outside these red circles? Do none of these exhibit significant / substantial changes in LST?
Response: Thank you for your guidance. We have made revisions. We have done this data set several times. After doing some analysis, the regional temperature change is relatively large and the representative area is analyzed and discussed again.

Data and Methods
Line 148: "near-polar orbit" --> "polar orbiting"
Response: Thank you for your guidance. We have made revisions.

Line 149: "with a flying height of" --> "flying at an altitude of"
I suggest to delete "with a temporal resolution".
Response: Thank you for your guidance. We have made revisions.
Lines 150-152: This is not entirely correct because the satellite switches orbit direction between the two overpass times given, i.e. descending and then ascending for Terra and ascending and then descending for Aqua. Hence you need to correct your notion of the direction. Please also include the names "ascending" and "descending". If you are in doubt I recommend to take a look here: https://www.ssec.wisc.edu/datacenter/aqua/ or https://www.ssec.wisc.edu/datacenter/terra/
Response: Thank you for your guidance. We have made revisions.

Line 153: I don't get the notion of the repeat orbits every 1-2 days. What is the exact repeat cycle? Which areas of China receive complete daily coverage using one satellite (I guess you can specify a latitudinal boundary)? Which area require data of one satellite from two days for complete coverage?

Response: Thank you for your guidance. We have made revisions.

Line 158/159: "another important algorithm:" can be deleted.
Response: Thank you for your guidance. We have made revisions.

Line 162/168: You write of the latest LST V006 product and refer to "MOD07_L2". This puzzles me because when I check respective LP.DAAC web pages I only see that these LST products are based on MOD11_L2 / MYD11_L2 products; MOD07_L2 is a product containing air temperature and water vapor profiles without any notion of an accurate land surface temperature. It appears to me that mentioning of MOD07_L2 (you should add MYD07_L2?) is ok for the atmospheric data included in the improved retrieval but that the LST retrieval itself is based on different MODIS products which you should give here as well.
Response: Thank you for your guidance. We have made revisions.

Line 172: Now I am even more puzzled: When I look at the LP.DAAC web page about the MOD11C1/MYD11C1 data set I find:
"The MOD11C1 Version 6 product provides daily Land Surface Temperature and Emissivity (LST&E) values in a 0.05 degree (5,600 meters at the equator) latitude/longitude Climate Modeling Grid (CMG). The MOD11C1 product is directly derived from the MOD11B1 product. "
Response: Thank you for your guidance. We have made revisions.
and further
"The MOD11B1 Version 6 product provides daily per pixel Land Surface Temperature and Emissivity (LST&E) in a 1,200 by 1,200 kilometer (km) tile with a pixel size of 5,600 meters (m). Each MOD11B1 granule consists of the following layers for daytime and nighttime observations: LSTs, quality control assessments, observation times, view zenith angles, number of clear-sky observations, and emissivities from bands 20, 22, 23, 29, 31, and 32 (bands 31 and 32 are daytime only) along with the percentage of land in the tile. Unique to the MOD11B products are additional day and night LST layers generated from band 31 of the corresponding 1 km MOD11_L2 swath product aggregated to the 6 km grid."
Response: Thank you for your guidance. We have made revisions.
Together with the information you gave in the paragraph ending in Line 165 I obtain the impression that these three paragraphs should be condensed and put into a more logical flow. Would it perhaps make sense to state upfront, i.e. in Line 158 that you use MOD11C1/MYD11C1 and MOD11C3/MYD11C3 and that this is the last generation of this V006 products which utilizes the day/night algorithm explained?
Response: Thank you for your guidance. We have made revisions.
Lines 172-182: How relevant is information about that collection 6 is an improved version of collection 5 and that there is an 8-daily product as well? I recommend to substantially condense this paragraph and merge it with the two previous ones. Please check in this context also the last sentence "resampling from 5600 m spatial resolution to a resolution of 7200 columns ..." which appears to be strange given the fact that the C1 and C3 products have this number of columns and rows exactly at that grid resolution of 0.05 degrees (=5600 m).
Response: Thank you for your guidance. We have made revisions.
Lines 183-188: What is the purpose of this paragraph? I don't think that you need to describe the

steps required to generate the 11C1 and 11C3 data sets (which seems to be your aim here). I also have a problem with the last sentence about the conversion of "brightness temperatures". As said above the C1 product is based on the B1 product which already includes surface temperatures. Therefore, this whole paragraph could potentially deleted.
Response: Thank you for your guidance. We have made revisions.

Line 190: "LST records at the hourly intervals" --> "Hourly LST observations"
Question 1: Do all these stations provide LST measurements over the entire period 2003-2017 at exactly the same location?
Response: The number of sites is variable, and some sites' data is not available for a certain period of time. A small number of sites have no surface temperature observation data.

Furthermore: Do all these stations provide LST measurements or are some of these only measuring atmospheric parameters and as such measure the 2m air temperature but not necessarily the LST? Please provide an adequate sentence / statement into your manuscript.
Response: The new observing stations generally measure the surface temperature and air temperature and other related parameters. A small number of old stations only measure the air temperature, and most of the old stations simultaneously measure the surface temperature after improvement.

Question 2: What is the surface type of the stations listed in Table 1? By surface type I mean the "biome" represented by the station. Is it urban, grassland, forest, barren, whatsoever? Please include this information into Table 1.
Response: Thank you very much for your careful review and reminder, we have added the corresponding statement in the manuscript and can be seen in Table 1. The surface types of most sites are bare land, grassland and agricultural land. Since the surface types of some sites change with the seasons, and many sites do not provide specific introductions, we have not listed them in detail here.

Question 3: You write hourly. Is that true? Do really all these stations provide these measurements with hourly temporal resolution? I am asking because typically such stations report either every 6 hours (0,6,12,18UTC) or, if it is climate stations, at three different times per day. I just want to be sure that you stress to the reader the apparently amazing consistency and temporal resolution of this data set of in-situ observations.
Response: yes, it is true. A few stations have half-hour observations. Relevant units usually only provide maximum and minimum values and average values.

Line 196: "overpass times" --> "local overpass times";
Furthermore: How is the "extraction of meteorological station data" done? Did you interpolate between two hourly temperature readings? I note that the MOD11C1 files contain the local overpass time so that you have a very exact measure of it; did you use this information for the co-location in time with the in-situ observations?
Response: Thank you for your guidance. We have made revisions. Most sites have observation data that are close in time to satellite transit. If not, we interpolated the two observations.

Line 197: "warming/negative trends" --> "positive/negative trends" or "warming/cooling"
Response: Thank you for your guidance. We have made revisions.

Line 198: Which LST data are meant here?
Response: Thank you for your guidance. We have made revisions.

Lines 205/206: It appears some information is missing here?
Response: Thank you for your guidance. We have made revisions.

Lines 208/209: Please check this sentence. It is not complete.
Response: Thank you for your guidance. We have made revisions.

Line 209: Why do you want to reconstruct a cloudless [please write "clear-sky"] LST data set? I suggest to delete the first part of this sentence and write seomething along the lines: "It is difficult to fill data gaps caused by clouds in LST data products based on satellite infrared imagery with data of the same quality as the clear-sky LST observations."
Response: Thank you for your guidance. We have made revisions.

Line 212: "that can reflect the true LST under cloud coverage." --> perhaps better: "that takes into account the actual LST under both clear-sky and cloudy conditions."
Response: Thank you for your guidance. We have made revisions.

Line 213: Aren't the highly accurate pixels preserved in both, the daily AND the monthly data?
Response: Thank you for your guidance. We have made revisions.

Line 217: "filtering(see ... details) . " --> "filtering (see ... details). "
Response: Thank you for your guidance. We have made revisions.
Line 218: "low qualityFor" ???
Response: Thank you for your guidance. We have made revisions.

Line 219: "at the corresponding time" --> perhaps better "for all days of the respective month" ?

Response: Thank you for your guidance. We have made revisions.
Lines 221-224: wrongly placed spaces
Also: "substeps" --> "steps"
Furthermore : "1) in situ ... " --> "1) Where possible we filled invalid grid cells with co-located in situ observations of the LST."
And: "2) In case in situ observations are lacking, we employed the ... method to interpolate ..."
And: "3) The remaining invalid grid cells we filled with LST values reconstructed based on regression ..."
Response: Thank you for your guidance. We have made revisions.

Lines 224/225: "Finally, the averages ..." --> I don't understand. I would have thought that after steps 1) to 3) all grid cells with invalid LST values in the daily data were filled so that one only needs to average over the full monthly record of daily LST values. If this is not the case then please say so in the manuscript, e.g.: "Even after these three steps, some grid cells still contain invalid values. These we filled by "... doing what? However, if after steps 1 to 3 all daily grid cells contain valid data, then : "Finally, we averaged over all daily data of the respective month and replace the invalid data in the original monthly LST product with the new monthly LST value based on the reconstructed LST time series of that month."
Response: Thank you for your guidance. We have made revisions.

Line 231/234: "other atmospheric disturbances" --> a very vague expression. Would it make sense to simply only write "aerosols" and leave it with that?
Response: Thank you for your guidance. We have made revisions.

Lines 232/233: "temperature values in the MODIS LST data" --> "LST"
Response: Thank you for your guidance. We have made revisions.

Liens 235-237: Is this general statement required? I mean, the MODIS LST data contain a suite of quality flags which give information about the cloud cover as well as an uncertainty value for the LST between < 1 K and > 3K. So, I'd say the producers are well aware of that and put adequate measures into the existing data sets.
Response: Thank you for your guidance. We have made revisions.

Line 238: "Statistical calculations were performed ..." --> What exactly did you do? Did you use the quality information (which and how) included in the MODIS LST product? Please be more specific and write explicitly in the manuscript which data you used for which step. Yes, I note that you write further down how you defined invalid grid cells in the monthly LST data but your description how you "created" Figure 3 is light.
Response: Thank you for your comments. Figure 3 is an example to illustrate the lack of daily MODIS LST data caused by cloud cover, so the original daily data without pixel filtering is used here. We directly use the ARCGIS software to statistically analyze the missing value of daily products. For the monthly MODIS LST data, the original data has fewer missing values directly caused by cloud coverage, so we did not show it in the manuscript and put more attention on the filtered data.

Lines 243/244: "to reconstruct a high-precision daily LST images that represents real values under clouds using ... " --> "to reconstruct high-precision LST under clouds using ..."
Response: Thank you for your guidance. We have made revisions.
Lines 250/251: "It is necessary ..." --> We know this already from the introduction and hence this sentence can be removed.
Response: Thank you for your guidance. We have made revisions.
Line 255: "Finally, pixels ..." --> "Grid cells ..."
Response: Thank you for your guidance. We have made revisions.
Line 257: "Finally, we reconstructed all the invalid pixels in ..." --> "Our aim is to reconstruct the

LST for all these grid cells with invalid data."
Response: Thank you for your guidance. We have made revisions.

Line 257/258: "Quality ... pixels (Benali et al., 2012)." --> What do you want to state here? I don't understand.
Response: Thank you for your guidance. We have made revisions.

Lines 272-276: This paragraph can be deleted.
Response: Thank you for your guidance. We have made revisions.

Lines 277-282: What is the intention of this paragraph? Are you making the statement that even if there are, say, 20 accurate daily LST data in a given month the data quality of the respective monthly LST value (if retrieved at all) is reduced? Do you have a reference for this statement? In other words: An interesting information you could pass on to the reader is: What is the minimum number of days with valid daily LST values (MOD11C1) which is required for a monthly LST value (MOD11C3)?
Response: Thank you for your guidance. The actual operation is to use quality control files, so this seems a bit redundant, which has been deleted.

Line 283: "filter the monthly image" --> "filter each monthly image"
Response: Thank you for your guidance. We have made revisions.

Line 284: "Then, we filter all the daily pixels from the month in which the cloud-low ... " --> Then, for each month, we filter all daily images of the respective month by determining all missing and low-quality grid cells." You use the term "cloud-low-quality pixels" in the sense that this is the group of pixels which are either missing completely or which are defined low-quality using the above mentioned QA information. If this is the case then also Line 285: "low-quality daily data" and Line 286: "low-quality pixels" should have this term used; in both case you aim to replace both kind of missing pixels: those which are missing anyways because they have been completely flagged as useless because of clouds and those which are of low quality.
Response: Thank you for your guidance. We have made revisions.

Line 286: "with the corresponding averaged pixel values from the daily data" --> "with the average LST derived from the gap-filled daily LST time series of the corresponding month."
Response: Thank you for your guidance. We have made revisions.

Line 289: "real ground LST" --> "actual LST"
Response: Thank you for your guidance. We have made revisions.

Line 290: "Many influencing ... reconstruction" could be deleted. Also, I would not group in-situ observations of the LST into "influencing factors". Therefore, please rewrite L291/292 as well.
Response: Thank you for your guidance. We have made revisions.

Line 292/293: "Therefore, the latitude ... same location" --> I would simplify this sentence

towards: "Grid cells with invalid LST values were co-located with meteorological stations". Please provide more information (this is one of the things which were not properly answered in your reply to the review comments): How is the co-location done? Were you co-locating distances in degrees or in kilometers? Were you computing distances to the grid cell centers? Was there more than one station in the grid cells? If there were more than one station in the grid cell, did you average over the values?

Response: Thank you for your guidance. We have made revisions. We make average because the resolution is very low, we cannot consider such a fine.

Line 293: "Invalidt" --> "Invalid"

Response: Thank you for your guidance. We have made revisions.

Line 294: "ground-based" --> "in situ"

Response: Thank you for your guidance. We have made revisions.

Lines 298-309: Is the degree of detail with which you describe this method really needed for your purposes? I note that for instance seasonal changes are possibly not important because you are using daily LST images here. I find the description overly complex and not to the point with respect to your specific problem.

Response: Thank you for your guidance. We have made revisions. Here we mainly introduce the general principles of this method, and it is flexible in practice according to the actual situation.

Line 306: "temporally closest at the same overpass time" --> What does this mean? In the ideal case this is a pair of partly overlapping consecutive overpasses (~100 minutes time difference and the closest descending overpass of the same day (within 12 hours).

Also, you refer to "reference images" but then in Line 307 to "a adaptive threshold is then used in the reference image ..." --> So, it is first several reference images, then it is just one.

Response: Thank you for your guidance. We have made revisions.

Line 311: seaons? See above.

Response: Thank you for your guidance. We have made revisions.

Line 312: Check for missing " "

Response: Thank you for your guidance. We have made revisions.

Line 313/314: "calculated from the standard deviation" --> of what?

Response: Thank you for your guidance. We have made revisions.

Line 315: "closer" --> Are you referring to a distance here?

Response: yes

Line 325: "greater than 4" --> So of the 5 x 5 = 25 pixels minus the central one = 24 pixels only 5 valid pixels are required for application of this method. Since, as you said, this is not done in a sliding window approach this reduces the grid resolution of the LST product effectively to 5 x 5600 m = about 25 km ... because to take into account the information of a 25 km x 25 km area to replace a value at the center grid cell.

Response: Thank you for your guidance. When the effective data is insufficient, the interpolation

method is to reduce the amount of information.

Line 327: "... the LST values of the pixels low-quality by clouds are determined through GWR." --> I suggest to use "reconstructed" or "filled" instead of "determined". In addition, I thought so far that you aim to replace any missing pixels and not just those which are potentially low-quality by clouds. This is confusing. I am still missing a clear, consistent and stringent definition and usage of which pixels you replace. It is tedious work to try to understand what you did.
Response: Thank you for your guidance. We have made revisions.
Line 330: "The" -->"the"
Response: Thank you for your guidance. We have made revisions.
Line 348: After all this discussion I still don't get whether grid cells for which the LST has been "reconstructed" using in situ observations (actually in this case a missing value is simply replaced by the respective value, so we cannot speak of a reconstruction) are getting a higher weight in the GWR than grid cells which contain valid high-quality clear-sky LST values. Please provide this information in the manuscript in a direct, non-hidden way. In case my assumption is wrong, then I don't understand the reasoning given in Lines 327-330.
Response: Thank you for your guidance. We have made revisions. If there is ground observation site data, similar pixels are obtained directly from ground stations which are the most representative and can better reflect the LST under clouds than under clear-sky conditions.

Line 350: "the elevation factor" --> "elevation"
Response: Thank you for your guidance. We have made revisions.
Line 351: "variation" --> "variation of the LST"
Response: Thank you for your guidance. We have made revisions.
Equation 7: An interesting question would be whether alpha is in the range between moist and dry adiabatic lapse rate of the air temperature? Is this the case? If not, are there enough physically reasonable explanations about why this is not the case?
Response: We did not consider it so complicated, in fact, it can be simply expressed as the effect of corrected elevation on temperature change. In this study, due to the low resolution, the actual effect is limited.

Line 371: This is the end of the methodology section. I have one major concern. See GC2.
Response: Thank you for your guidance. We have made revisions.

##########
Results
I have one general concern with the product. See GC3.
I have one general concern with respect to the structure of the results section. See GC4.
Response: Thank you for your guidance. We have made revisions and response above.

Lines 380/381: I don't think these first two sentences are an adequate introduction for a subsection about trend analysis of average Chine annual-mean LST data.
Response: Thank you for your guidance. We have made revisions. We have made a lot analysis in 2017 (Mao et al, 2017), and found that the mean surface temperature of MODIS surface

temperature in 4 time periods is close to the annual mean surface temperature because the observation time of MODS sensor is symmetrical, so it is feasible to use monthly mean instead of annual mean. Detailed derivation and comparative analysis of how to calculate the average temperature can be referred to reference (Mao et al. 2017).

Mao, K. B., Ma, Y., Tan, X. L., Shen, X. Y., Liu, G., Li, Z. L., Chen, J. M., Xia, L.: Global surface temperature change analysis based on MODIS data in recent twelve years, Advance Space Research, 59, 503-512, http://dx.doi.org/10.1016/j.asr.2016.11.007, 2017.

Lines 381: "To obtain ..." --> How did you do this? Did you first compute the monthly all-China average LST and then the mean over every year? Or did you first compute for each grid cell an annual mean LST and averaged the values over China as a whole? What kind of a trend analysis did you perform? A linear regression analysis? Any weights? Any significance tests? Please refer back to the respective subsections in Section 3.

Response: We have made derivation in paper (Mao et al., 2017). Detailed derivation and comparative analysis of how to calculate the average temperature can be referred to reference (Mao et al. 2017).

Mao, K. B., Ma, Y., Tan, X. L., Shen, X. Y., Liu, G., Li, Z. L., Chen, J. M., Xia, L.: Global surface temperature change analysis based on MODIS data in recent twelve years, Advance Space Research, 59, 503-512, http://dx.doi.org/10.1016/j.asr.2016.11.007, 2017.

Lines 388/389: What does abnormal rainfall have to do with LST? Please add the physical reasoning for this observation.

Response: Thank you for your guidance. We have made revisions.

Lines 380-392: When you write "significantly" in this paragraph, is this statement based on a significance test result?

Response: Thank you for your guidance. We have made revisions. These indirectly verify the correctness of the reconstructed data through meteorological events, indicating that the reconstructed data can be used to analyze the long-term spatiotemporal changes in surface temperature.

Line 408: "approximately 20.80%" --> "20.8 %"

Line 414: This is the second time you refer to the reported hiatus in global warming. Would you think it makes sense to leave these statements to the discussion / conclusions section? There it would be much more visible than somewhere inbetween the description of the results.

Response: Thank you for your guidance. We have made revisions.

Lines 433-445: Please check this paragraph for superfluous blanks " ".

Response: Thank you for your guidance. We have made revisions.

Line 479: ")f" --> ")"

Response: Thank you for your guidance. We have made revisions.

Lines 502-511: See GC5

Response: Thank you for your guidance. We have made revisions.

Line 511: "as been resolved" --> perhaps better "reduced"?

Response: Thank you for your guidance. We have made revisions.

Line 520/521: I am sorry but Figure 3 and 4 are only snap-shots, showing TWO days and TWO months of the respective year. This does not support the global statement made in this sentence.

Response: Thank you for your guidance. We have made revisions. The east and south of China are connected to the Pacific Ocean, so the amount of water vapor (clouds) in the sky is higher than in the west.

Line 537+: Please note whether you used the re-constructed bias-corrected LST data in Table 2. It is not clear which of your two LST products you compare here.

Response: The evaluation here is the result of filling in invalid values and low-quality pixel values. According to the results of the regional evaluation (higher or lower), we made further corrections. The accuracy is theoretically higher than the accuracy after the first filling. We take the accuracy evaluation of the first filling as the accuracy of the final product. Because this correction can be cyclically iterated, in theory, the accuracy can be continuously improved by dividing into smaller blocks, but each time requires independent data to verify.

Line 543/544: Here you write that you use the original MODIS LST product directly without filter using the QA flags. I have two comments here:

1) Your product merges over data from MOD11C3 and MYD11C3. Hence you need to merge these two products for your comparison as well and/or perform the comparison with both these products.

2) Doing the intercomparison without taking the quality flags into account is ok - as long as you also present in addition an intercomparison where you applied the same QA filtering as you used for the re-construction to the original LST data. For this comparison, the improvement of your approach should be - if at all measurable close to zero. Why? Because with your re-construction you are supposed to only correct the LST for the missing and low QA grid cells. You could also split your evaluation and in addition to what I just asked for show a comparison of the original low QA data and the reconstructed high-precision LST. If I understood your approach correctly then this latter LST difference should be pointing to the achieved improved even more clearly. I am looking forward to this result of your quality check in the revised version.

Response: After careful thinking and analysis of all the questions and suggestions you raised, we agree that we should present in addition an intercomparison to show the credibility of the generated data set. The revised section is as follows.

(1) Figure 5 shows scatter diagrams relating ground station data and original MODIS LST monthly data (MOD11C3 / MYD11C3) without QA filtering. It can be seen from Figure 5 that for each region, the deviation of some points causes the distribution of points to be more discrete. The error of the deviated points is mostly concentrated between 0-3 ℃, and the deviation of a few stations reaches 4-6 ℃.

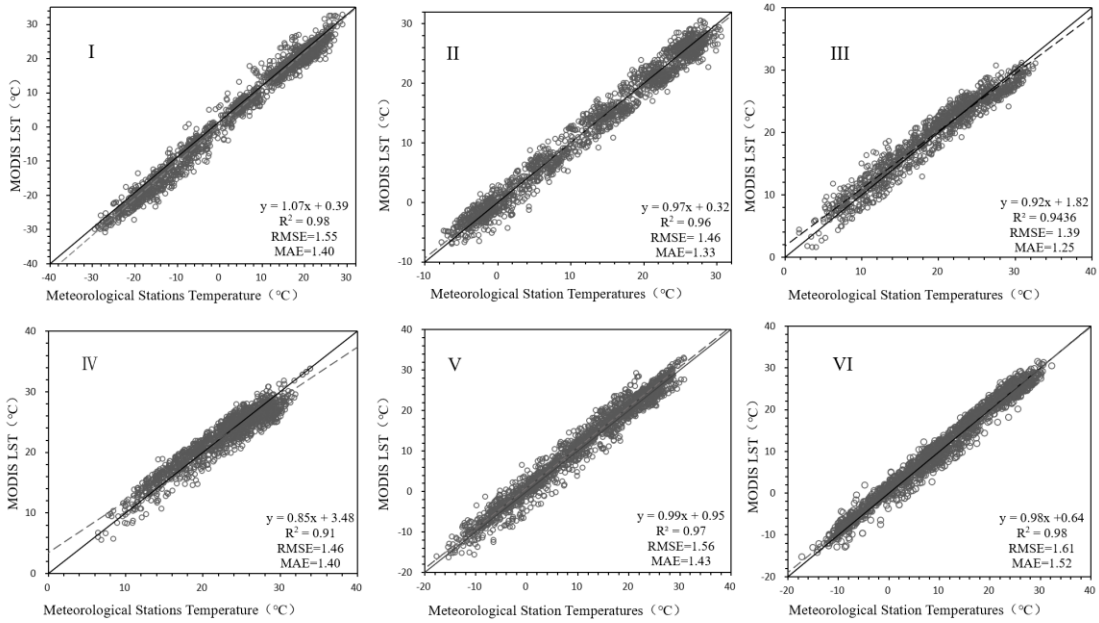

Figure 5. Scatter diagrams of original MODIS LST monthly data (MOD11C3 / MYD11C3) against ground station data, the statistical accuracy measures (R^2, RMSE, and MAE) are also indicated.

(2) Scatter plots of the low-quality MODIS LST data and reconstructed results versus their corresponding ground station data were also plotted (Fig.7) in order to show the accuracy comparison of low-quality pixels before and after reconstruction more clearly. It can be seen from the figure that the overall result between the reconstructed MODIS LST data and the ground station data presents a better linear relationship, with more clustered distribution on both sides of the 1: 1 line, indicating that the accuracy of the reconstructed data has been significantly improved.

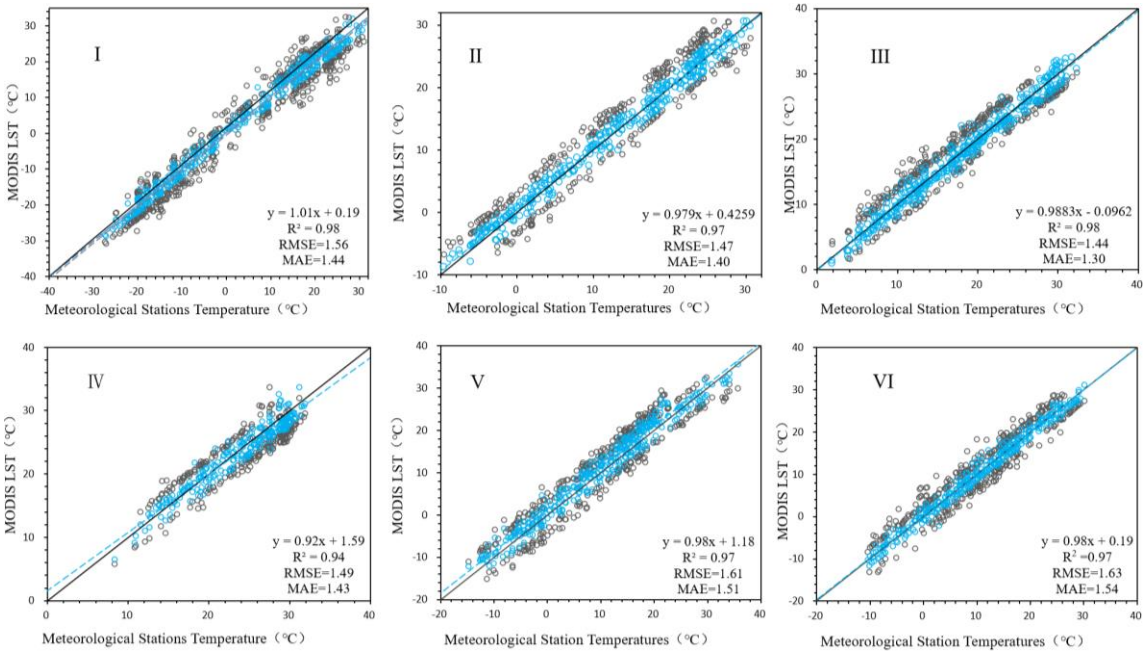

Figure 6. The scatter diagrams of the low-quality MODIS LST data and reconstructed results versus their

corresponding ground station data in six natural subregions (I, II, III, IV, V, and VI). The gray points indicate low-quality LST pixel values in original MODIS LST data. The blue points represent the values in new LST dataset, and the statistical accuracy measures (R^2, RMSE, and MAE) are also indicated.

The associated explanation has also been added to the revised manuscript and can be seen in Subsection 4.1 and 4.2. We hope our revision will give you satisfaction.

Line 553/554/562/567: "The main reason for this difference may be the complex and diverse terrain" --> Don't you think that this is exactly one of these cases where a better illustration of your approach would allow you to be more to the point here? You have all the data at hand and instead of speculating what the reasons might be you could come up with a clear statement about... "as we see in figure xy in the complex terrain of region xyz the 19x19 sliding window used for the LST reconstruction based on the relationship between LST and elevation results in unrealistic LST values" ...?

Response: Thank you for your guidance. We have made revisions. After consideration, I still think that the main reason is the insufficient number of surface meteorological observation stations in these areas.

Line 577:

I agree that a better spatial coverage is achieved with this re-constructed LST data set. While the overall accuracy has improved compared to the in situ LST observations it remains unspecific whether this improvement applies indeed only for those grid cells which have not been defined a lower quality because a separate uncertainty investigation has not been carried out, i.e. an answer to i) What is the difference between the original and the (bias-corrected) re-constructed LST data for QA of the original LST of < 1K? ii) What is the respective difference between the LST data sets for original LST QA > 1K?

Response: Thank you for your confirmation. Your requirements are very high, the data volume and accuracy of the observation site are limited because of the technical limitations at that time, which is a bottleneck. We just use different information sources as much as possible to improve accuracy. This can only be better than before, but we have not solved all the problems, and it will still be difficult to solve in the future. We can add more ground-level meteorological observation stations in the future and add more data sources to further improve accuracy.

Figure 2:

- Why does the first step include a "re-projection"? The MOD11/MYD11 data are on a regular 0.05 degree x 0.05 degree grid. Hence for which step of your processing do you require the re-projection?

Response: Thank you for your comments. MOD11/MYD11 data provided in an equal-angle geographic. They were converted to Albers Conical Equal Area projection to keep the area undistorted.

Figure 3 / Figure 4:

- I am sorry, but didn't get your argument why in panel a) you kept blue for warmer LST values instead of reversing the color table and use blue for cold and yellowish / red for warm. This would be far more intuitive. I know, these figures is shown mainly to illustrate the extension of the cloud

influence / number of data gaps but this does not prevent a reader from being puzzled by a color table that is reversed to what is usually used. Therefore please reverse the color table in panel a).
Response: Thank you for your guidance. We have made revisions for Figures 3 and 4. We made a mask with the quality control file, and then obtained the required results and made a drawing.

- Figure 3: How is "valid" defined? Did you use QA information as well? If not, why not?
Response: Thank you very much for your careful review. In this study, for daily data, all pixels with LST error> 3 K in daily data are rejected. This is mainly due to the fact that there are too many missing values (about 50%) in the original daily data. This is a compromise between maintaining the quality of the data and having a certain proportion of available data. In addition, the use of adjacent ground station data also helps to improve the accuracy of monthly data reconstruction results. Thanks a lot for pointing this out and the associated explanation has been added to the revised manuscript and can be seen in Subsection 3.3.1.

- Figure 4: Reading the bit-wise encoded QA information is not straightfoward. Did you check how the number of invalid data changed when using a larger average LST error of, e.g. > 3 K?
Response: Thank you very much for your careful review. The data quality control information is statistically calculated and stored in the corresponding QA layer and is represented by an 8-bit unsigned integer. According to the bit flags defined for QA information (Table S1), it can be seen that the pixels with DN values of 0 and 8 can satisfy the condition of "00 = average LST error <= 1K" after being converted into binary. According to our statistics, pixels with values of 0 and 8 in the QA information account for most of the total number of pixels.
Taking 2017 as an example, the cells with DN values of 0 and 8 accounted for 76.55% of the total in QA information of Terra day (MOD11C3) in January 2017, of which 38.69% had a DN value of 0 and 37.85% had a DN value of 8. In the images (MOD11C3/ MYD11C3) of all months in the study area in 2017, the total number of pixels with DN values of 0 and 8 averaged about 78%. Therefore, at least about 78% of the pixels can meet the requirement of average LST error <= 1K.

We also counted the number of valid data in different situations. We also counted the number of valid data in different situations. When LST error> 3 K, that is, LST error ≤ 3K is considered to be a good quality pixel, the average number of invalid data in 2017 accounted for 0.9% of the total data (MOD11C3 / MYD11C3). At the same time, when LST error> 2K is considered to be poor quality data, the average invalid data percentage in 2017 was 1.2%.

Data with large errors seriously affects users' use of data and further analysis. Therefore, in this study, in order to ensure the accuracy of the data, we only kept high-quality data. Grid cells with QA layer labels of "the average LST error <=1 K", "LST produced, good quality" and "the average emissivity error <=0.01" are considered to be high-quality data, and the remaining pixels are low-quality pixels and are set to missing values.

Table S1. Bit flags defined for QA information in MOD11C3/MYD11C3

bits	Long Name	Key
1 & 0	Mandatory QA flags	00=LST produced, good quality, not necessary to examine more detailed QA

		01=LST produced, other quality, recommend examination of more detailed QA 10=LST not produced due to cloud effects 11=LST not produced primarily due to reasons other than cloud
2	Data quality flag	0=good data quality 1=other quality data
3	Terra/Aqua Combined-use flag	0=no 1=yes
5 & 4	Emis Error flag	00=average emissivity error <= 0.01 01=average emissivity error <= 0.02 10=average emissivity error <= 0.04 11=average emissivity error > 0.04
7 & 6	LST Error flag	00=average LST error <= 1K 01=average LST error <= 2K 10=average LST error <= 3K 11=average LST error > 3K

Table 1:

Columns latitude and longitude miss the direction, i.e. degrees North and degrees East.

Response: Thank you for your guidance. We have made revisions.