Dear professor,

At this time when the news dominated by pandemic, I hope you and your family are healthy and safe. Let's get through the epidemic and welcome spring together. Our sincere apologies for the slow response point-to-point. The situation is not clear yet, we have to keep our social distance and self isolated at home. It causes difficulties to communicate, both person to person and data transfer. Our data is stored at the office, and the building is locked down most of the time. That's why we have to take longer time to response.

Thank you very much, and your comments and suggestions for modification are very good. We have tired to modify and emphasize your comments in our paper which are marked in red. Our work is very important and meaningful. The dataset has been downloaded 2203 times (https://zenodo.org/record/3378912#.XmwrCXK-s2w) and the method of data set paper has been cited by two papers. We have received many thanks from many users for our dataset. Thank you for attaching importance to our work,

Temperature is one of the most important geophysical parameters in studying ecosystems. 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. **Our method tries to combine the advantages of remote sensing and traditional methods to improve accuracy.** A detailed analysis is already made in "**AC9: 'Response to RC3', kebiao mao, 16 Mar 2020**". The point-to-point response is as follows.

Sincerely,

Kebiao Mao, et al.

Response

Summary: Data gaps, e.g. due to cloud cover or aerosol, limit the usefulness of land surface temperature products derived, for instance from MODIS aboard the EOS TERRA and AQUA satellites. Depending on region and season this applies to both daily and monthly products. This paper presents a potential improvement of this situation by proposing the development of an enhanced monthly LST product based on gap-filled daily LST data for China. Gap filling is carried out with the aid of LST observations from meteorological stations, a similarity analysis and linear

regression between observed LST and elevation. The paper provides an extensive interpretation of LST trend maps on annual, seasonal and monthly temporal resolution

This paper has been submitted to the journal "Earth System Science Data". My understanding of this journal is that it is a platform to present new data sets alongside a thorough description and evaluation of the used methodology and results, the data formats and the content of the data files. To my opinion, this paper does not fit into this journal. Neither is the method described adequately so that it is easily understood, nor is the method or steps of it evaluated or illustrated enough. Examples of the new data set are not shown, instead the authors present maps of the multi-year annual mean LST, trends and correlations. Too much focus is put on harvesting the data set. I don't see a reason why I should use this data set. It is neither evaluated properly enough nor does it come with a critical discussion of limitations, uncertainties or even improvements over the data sets existing so far.

Response: We would like to thank the anonymous referee for reviewing our manuscript. These constructive comments are very important for us to improve the present manuscript.

We agree that description should focus on the introduction of the data set production method and the evaluation of the data set. So, we added more information on the reconstruction method, especially some examples of graphs in the process, so that it can be more easily understood by users. In addition, we removed the analysis part of the LST trend in the main body of the manuscript according to your advice. After careful thinking and analysis of all the questions you raised, we have replied to all these questions and comments. In this process, we also have carefully modified and improved the corresponding parts of the manuscript. Changes in our manuscript are indicated by blue slanted font.

General comments:

GC1: The description of the LST restoration method requires rewriting, clarification and more illustration of what is done. Several issues are unclear. Please see my specific comments.

Response: Thank you for the valuable comment. We have made a point-to-point response to your question in the specific comments.

GC2: Section 3 lacks a final sub-section in which the reader learns what the output of all the measures explained so far is. Neither are examples of re-constructed daily LST time series at a

particular location nor gap-filled daily or monthly LST maps shown here. Missing at this place as well is a critical assessment of the validity of the obtained, gap-filled LST data. The evaluation shown in Section 5 it too global (see also GC5). The results section begins right away with the presentation and discussion of the annual national-mean LST.

Response: Thank you very much for your advice. For thermal infrared remote sensing data, there are large number of missing values (more than 60%) in daily MODIS LST image and the location of missing data is randomly distributed. In this study, we only reconstructed the missing pixels of the monthly image, resulting in many missing values still exist in the reconstructed daily MODIS LST image, though these missing values had no much effect on the accuracy of our final monthly image. To make the reconstruction method easier to understand, we added more explanations of the reconstruction steps and the corresponding output in the manuscript. Thanks again for your suggestions.

GC3: While the data section introduces data from both sensors MODIS TERRA and MODIS AQUA, it is not clear which are shown in the results (only for Figs. 3 and 4 it is mentioned that this is TERRA). This is particularly important since local overpass times of MODIS AQUA are closer to the daily minimum and maximum LST values than of MODIS TERRA. If for some reason the LST results shown in Fig. 5 are based on a combination of both, TERRA and AQUA, then there is a critical lack of information throughout the paper which needs to be mitigated.

Response: The Terra satellite acquires data daily at 10:30 and 22:30 local time and the Aqua satellite acquires data daily at 13:30 and 1:30, resulting in four LST observations per day. We agree that the Aqua satellite can better represent the minimum/maximum LST values. For the average LST, theoretically, the more data used in a day, the more accurate it is in representing the temperature of the day. At the same time, the combination of Aqua and Terra data can reduce some systematic errors more than using only Aqua data. Therefore, after obtaining the seamless LST data at the four local overpass times after data reconstruction, we averaged the Terra and Aqua data during the day to get the daytime LST. Similarly, the average of Terra and Aqua overpass at night was used as the LST at nighttime. Finally, the final result is obtained by averaging the day and night MODIS LST. We have added a brief description in the manuscript as follows.

"After LST data restoration data reconstruction, four overpass times of images are obtained each month. Calculated the average LST at four times to represent the LST image of the month."

GC4: A large fraction of the results section 4 comprises hypothetical considerations about potential causes of the observed LST changes, supplemented by hypotheses about the potential impact the observed LST changes may have. I suggest that all these hypotheses are taken out of the paper and used to motivate a suite of follow-on studies where an improved and evaluated LST data set is compared to and interpreted with the aid of additional data, such as CO2 concentration, atmospheric re-analyses, snow cover data, land-cover change data, etc. Results of those studies can then be published in highly ranked journal such as Nature Geoscience or similar.

Response: Thanks for your good suggestions, which has helped us to improve the manuscript quality. We agree that the statement about potential causes of the observed LST changes and impact should be deleted to better highlight the data set. We also plan to combine the reconstructed LST data set with other data for further research. Thanks again for your advice.

GC5: It is clear that the re-construction method presented in Section 3 cannot be ideal and is likely to have errors. However, the verification section is very global and does not touch upon an evaluation of the steps performed in the re-construction but instead, because the biases / RMSEs apparently are not "good" enough, comes up with a bias-correction method. In order to develop and apply such an additional step, first a thorough evaluation of the re-construction method should be carried out, taking into account into which direction the re-construction and hence gap-filling of the original LST product will change the LST values. This needs to be understood first and is not convincingly enough laid out in this paper. –> One example: Lets consider that the original monthly daytime LST of a month with high insulation but also a considerable number of data gaps due to invalid daily LST values due to clouds is +35degC. If we assume that this occurs in a region with high station density, then it is likely that the actual station measurements will have a high impact on the reconstructed LST. According to your method one can expect that a substantial fraction, if not all, of the missing daily LST values would be replaced by station observations, which since these are from cloud-covered days, provide a LOWER LST, e.g. +25degC. Now, reconstructing the monthly LST with this new daily LST time series which is composed of original clear sky and hence high LST values AND re-constructed cloudy-sky lower LST values would result in a lower monthly mean LST value. And this is actually your result for many of the grid cells. Such considerations and explanations are required by a reader of this paper and user of your data set to understand what is the new, enhanced, and credible part here.

Hence, what the evaluation section really needs is A) examples of time series of the original and the re-constructed DAILY LST for several selected stations for several years / seasons - taking into account all three flavours of the reconstruction method

B) examples of maps where the authors illustrate how for a region without valid data (like shown in Fig 3) the reconstruction method fills in LST values - again taking into account all three flavours.

C) examples which illustrate how the reconstruction method works for day-time and night-time (clouds have different impact) and in areas with high- or low open water areas and in areas with / without snow cover

D) examples of different topographic complexity which illustrate how the similarity concept and the elevation-determines-LST concept perform - ideally for two different seasons. Once this is done, I suggest a similar investigation with the monthly LST data. I as a reader and potential user of an enhanced LST data set want to see original and reconstructed monthly LST maps and not just multi-annual mean maps of trends. Last but not least a sub-section should be devoted to the true quality of the station observations in terms of the spatial representativity and other aspects I laid on the specific comments.

Response: After careful thinking and analysis of all the questions and suggestions you raised, we agree that the reconstruction method should be evaluated more comprehensively to show the credibility of the generated data set. The associated explanation has been added to the revised manuscript and can be seen in Subsection 3.3.2.

GC - wording / editing:

- Please avoid usage of, e.g., "warming trends". A trend is either positive or negative, indicating an increase or decrease of the geophysical parameter assessed; in your case it is the LST, so we have an increasing LST which is a warming and we have a decreasing LST which is a cooling. The trends themselves are not warming.

Response: Thank you very much for your careful review. We have modified the words " warming trends " and " cooling trends " to the more appropriate " positive trends " " negative trends", respectively.

- You are introducing regions I through V in Figure 1 but hardly refer to those in the description and interpretation of your results. It would greatly aid the readability of your paper if you would more stringently follow your own notation of regions instead of the geographical names which may not

be so familiar to readers outside of China.

Response: Thank you very much for your careful review. We also agree that the numbering of each area is easier for readers to understand. We have checked the relevant parts of the manuscript and marked with the corresponding number of the geographical name.

Specific comments:

Lines 63/65/67/70: "low-quality" ... "noise-contaminated" ... "cloud contamination pixels" ... "poor-quality" –> Please, this is not a narrative but a scientific paper. I suggest to define clearly at this stage for which pixels you will re-construct the LST, write this down in a well-structured way, and introduce terms which hold for the entire paper. "low-quality" can be due to clouds, low observation angles, aerosols, etc."; "noise-contamination" as well, even though this sounds like sensor noise and cross-talk effects as well; "cloud contamination pixels" is clear. Which of the low quality pixels will you replace by a re-constructed value for which effects causing this low quality? - clouds? / - cloud shadows? / - aerosol? / - low observation angles? / sensor noise?

Response: We appreciate the referee for his/her suggestion, which has helped us to improve the manuscript quality. The pixels which need to be reconstructed include null pixels and low-quality pixels. Low-quality pixels are caused by undetected thin clouds, atmospheric disturbance, observation geometry and instrumental problems (Wan, 2014). We agree that we should define clearly at this stage for which pixels you will re-construct the LST. So, as the reviewer suggested, we have modified these words in the revised manuscript. The revised lines 63-70 are as follows:

"Moreover, other factors can also contaminate the observation signal and cause the data to be unavailable, such as atmospheric disturbance, observation geometry and instrumental problems (Wan, 2014). In general, the abnormally low values in LST maps caused by undetected thin clouds, together with other poor-quality values, need to be identified and filtered because these values greatly reduce the accuracy of the LST data.

Cloud cover and other factors, which causes extensive amounts of missing and low-quality pixels, significantly reduces the proportion of usable LST data and poses a problem to further applications. Thus, the reconstruction of these missing and low-quality pixels is necessary for satellite-derived LST applications."

Line 78: So far with the advantages of method I. What are its disadvantages?

Response: Thank you very much for your careful review and reminder, we have added the

corresponding statement in the manuscript. The first category of methods, which estimate missing MODIS LST data using only LST data, takes advantage of the similarity and interdependence of the available temporal/spatial attributes of neighboring pixels. To some extent, these methods have the advantage of simplicity and reliability. However, this category of methods is often not as reliable as expected especially in complex topographical regions and areas with many missing data, because it cannot obtain enough information for reconstruction.

"To some extent, these methods have the advantage of simplicity and reliability. However, this category of methods are often not as reliable as expected especially in complex topographical regions and areas with many missing data, because it cannot obtain enough information for reconstruction."

Line 93: Please provide a frequency here; what is meant by "high-frequency channels"?

Response: Thank you for your comments. High-frequency channels usually refer to channels with a frequency of 85 GHz and above. The radiation of the high-frequency channel mainly comes from the shallower surface emission layer, which is closer to the 0 cm surface temperature; while the low-frequency channel has better penetration, the emission layer maybe come from a certain depth of soil, and the obtained bright temperature is the volume bright temperature rather than surface temperature, so there is a large error. We also explain it in the manuscript.

Lines 111-122: I don't think this paragraph belongs into the "study area" section. I suggest to either delete it or put it into the Introduction section to underline the importance of a more accurate, medium-term LST data set.

Response: Thank you for your comments. We also agree that this paragraph shouldn't belongs into the "study area" section, so we removed this paragraph. Thank you for your guidance.

Lines 129-131: Also here I suggest to remove the hint to the crop production; this information appears not to be relevant for the paper - unless you want to refer to land surface type patterns ... which you, however, don't to for the other three regions in eastern China. So this info can be removed. Line 130: I have difficulties to believe that region I covers > 50% of China's land area as it is written currently.

Response: Thank you for your comments. The change of temperature has a significant impact on grain production. In order to better explain the impact of temperature in different regions on crop production, we have added more content of crop production. After careful consideration, we agree

that these information were redundant for this study, so we removed this sentence (Lines 129-131). We are sorry for carelessness on Line 130 and we have removed it. Thank you for your guidance.

Figure 1: - The coloring of the elevations is opposite to what is usually used. Was this done on purpose? If so why? - Did you try to use smaller black dots to mark the stations? These might come out clearer. - Did you try to use white or black for the delineation of the sub-regions; currently their boundaries are somewhat hard to follow. - I note that the red ellipses denoting the key regions used for evaluation are partly obscured by the dots marking the station. I suggest to a) use a different color for the ellipses (black or white) and b) plot them on top of the stations locations. - In the caption: "spatial patterns of the meteorological stations" –> perhaps better "spatial distribution of ... ", or simply "location of ... ". One could already add the total number of these stations in the text of the caption.

Response: Thank you for your comments. We also tried some methods to match the symbols and colors to better reflect the Terrain features, the stations locations, the area divisions, and the location of the stations for accuracy verification, but none of them achieved a good effect. Thanks for your good suggestions. We changed the color of the elevations, and used the light color combination to make the boundary of the sub-regions more clearly.

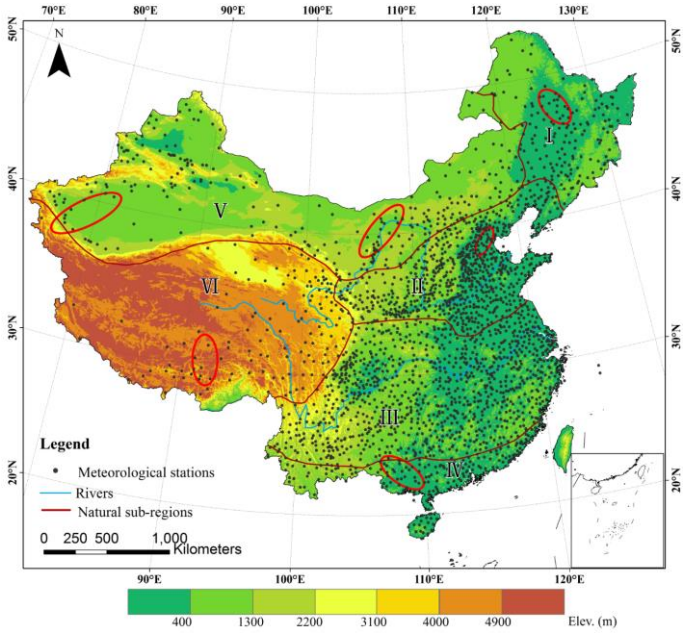


Table 1: It appears to be that the ellipses contain more stations than are actually listed in Table 1, e.g. regions c) and d). What is the reason for this?

Response: Thank you for your comments. The LST of the stations we selected (Table 1) has a

significant positive/negative change trend, and these stations are distributed in areas with different climatic conditions. We think these sites are representative when assess the effectiveness of the reconstruction method. In the process of data reconstruction, we used as much reliable site data as possible to improve the accuracy of the results. The eastern part of China（I, II, III and IV）has a flat Terrain and high site density. In order for readers to be able to clarify the distribution of sites in different regions, Figure 1 contains more sites than they actually are. We have also explained this in the manuscript.

Line 149: I suggest to put in some notion about the snow cover and its duration in the various regions because reflection of incoming solar radiation as well as thermal emission are considerably different from bare and vegetated surfaces and should play a role in the retrieval.

Response: Thank you for your comments.

Lines193-198: I don't understand the meaning of this paragraph. Is this something you did with the MOD/MYD11C1 and MOD/MYD11C3 data? Possibly not because none of these data sets include brightness temperatures. Hence I find this paragraph a little confusing and not well connected to the previous one. Clarification would be welcome.

Response: Thank you for your comments. This sentence is an introduction to MODIS LST data. It indicated that the monthly (MOD11C3/MYD11C3) data are calculated and averaged by daily data (MOD11C1/ MYD11C1). This is also the basis of our research that the daily data (MOD11C1/ MYD11C1) can be used to reconstruct the monthly (MOD11C3/MYD11C3) data.

- Please mention why you don't u se data before 2003. TERRA MODIS LST data begin in Feb. 2000 and AQUA MODIS LST data begin in July 2002.

Response: Thank you for your comments. As you said, AQUA satellite did not acquire data for the whole year before 2003. It is not very reliable to take the average temperature obtained only from two overpass times as the average LST values. Moreover, the average LST values from TERRA may deviate from the value from TERRA and AQUA after 2003.

Line 203: Please provide a reference to the Jackknife method.

Response: Thank you very much for your careful review. Jackknife method refers to the random resampling of the entire observations set several times (Benali et al., 2012). We also added this reference to the manuscript.

Lines 200-209: - Did all these stations record without data gaps over the time period given?

- Is the way the land surface temperature is observed homogeneous across all stations for the entire period? It is automated of manual? Were there changes in instruments? Were there changes in location (and hence land surface properties) for any of the stations during the period? Were there changes in the location of the stations with respect to the heat island of a growing city / faculty complexes / nearby reservoirs / reforestation? In other words: How confident are you that these data provide high quality means to supplement your product and to use it for (a first) evaluation?

Response: Thank you for your comments. Thank you for your comments. Stations record were provided and subjected to strict quality control and evaluation by the China Meteorological Administration. Only a small part of the observation records are excluded due to disqualification. We have uniformly processed the surface temperature data of the observation station before it is used in this study. For the integrity of the site data, only recorded site data was used in this study, and null values were ignored. For the change of site location， due to urban expansion and other reasons, from 2003 to 2017,the detection environment of some stations was destroyed and relocated. For the relocated stations, we only retained the observation data and coordinate information of the long observation time series before / after the relocation.

- How representative are the locations of the stations in terms of the surface conditions (over which surface exactly is the LST measured in relation to the matching (co-located) CMG 0.05 degree pixel of the MODIS data?)? This is critical for your step I (see Line 225): filling the daily pixel by the in-situ observation. It has been shown that already at the scale of 500 m the evaluation of a geophysical quantity derived from a satellite sensor with an in situ observation can yield misleading results because of heterogeneous surface properties in the 500 m grid cell. Here we talk about a grid cell an order of magnitude larger!

Response: Thank you for your comments. Large-scale data includes LST and soil moisture, and there are indeed representative problems in the data verification part of ground stations. Because thermal infrared is greatly affected by clouds, if you do not use ground station observation data information, direct interpolation will result in higher interpolation results than actual results. In order to further improve the interpolation accuracy, we use as much ground station information as possible to make the ground products more consistent with the actual situation. This is currently the best selection method for such large-scale data. In order to further increase the representativeness of the

site data, the number of sites will need to be increased in the future to improve accuracy where the surface type changes greatly and the Terrain is undulating. For this part of the content, we also made supplementary explanations in the paper.

- I don't understand the mentioning of / focus to the 4 equator local overpass times. China is a big country and local overpass times at the equator differ from those further north. I suggest to provide maps (4, two for AQUA, two for TERRA) of average (typical) local overpass times of your region of interest for illustration that this approach is correct.

Response: Thanks for pointing this out. Yes, we agree that four maps for illustration is more appropriate for this study. We have changed it for the whole context in the revision.

Lines 210-211: I don't understand the connection between the SRTM data and the LST; please be more specific. Could it be that your DEM had gaps which you filled with the SRTM data? Please clarify in the text.

Response: Thank you for your comments. SRTM (Shuttle Radar Topography Mission) DEM(Digital Elevation Model) data were used in this study. Temperature dem is closely related. There have been many studies on the relationship between temperature and altitude in geography. For example, Barry studied the relationship between mountain temperature and altitude (Barry, 1992). Körner redetermined the geographical significance of altitude and discussed the influence of altitude on temperature (Körner, 2007). The vertical decline law of temperature decreasing by 0.55℃ when the altitude increases by 100 meters is applicable to most areas.

Line 211: I note that for the verification you are using a subset of the same data set you use for the improvement of the data set. Isn't there perhaps another, more independent means to verify your product, e.g. airborne satellite under flights, other well-calibrated IR temperature measurements?

Response: Thank you for your comments. Due to its optimal temporal and spatial resolution throughout the world, the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor has become an excellent data source for satellite-derived LST data and is widely used in regional and global climate change and environmental monitoring models. We have also tried other data, but the stability and accuracy of the data are not high enough, and the data obtained by different sensors are quite different.

Line 283-290: Also here I have several concerns with the formulation as well as the content: - Whether the approach described here is viable for a 0.05 degree grid cell needs to be proven and

needs to be supported with the information you provide in the revised version of this paper with respect to how representative the station measurement is with respect to the grid cell in terms of land cover, elevation, etc.

Response: Thank you for your comments. This method is to take advantage of the similarity of geographically adjacent geographical units, and the formulation has been widely used in geographic information system large-scale interpolation. It is widely used at the macro scale, please refer to the following references (Hansen et.al.,2010; Smith et.al.,2018; Hansen et.al.,1987; Hansen et.al.,1988; Hansen et.al.,1999; Rayner et.al.,2003). We have added references to the manuscript.

- It is not correct that the LST in adjacent cloud covered grid cells is always lower than the clear-sky ones. This is only valid for daytime LST values and might also be not fully valid for snow-covered grid cells.

Response: Thank you for your comments. We revised this sentence in the revised version.

- I suggest to not write "prediction" in the context of the satellite LST observations. Likewise, I suggest to replace "predictor factors" by, e.g., "influencing factors" and add "elevation".

Response: Thank you for your comments. We revised this term in the revised version.

- How is the co-location between the MODIS grid cell and the station done? Is there a threshold distance to the grid-cell center required to use the station observation or is it sufficient if the station falls just within the approximately square-shaped grid cell? How did you measure this distance (in kilometers / meters or in degrees)?

Response: Thank you for your comments. We deal with this problem based on this principle: (1) when there is no cloud, we use MODIS temperature product data; (2) in the case of cloud, if the pixel has ground observation site data, we will calculate the difference when there have no cloud based on nearby dates between the pixel and the site data of the MODIS temperature product as a threshold to modify the site observation data as the temperature data of the pixel; (3) When the pixel has no site data, it is processed by classic interpolation methods. See references (Hansen et.al.,2010; Smith et.al.,2018; Hansen and Lebedeff, 2018).

Line 214: "... it is difficult to reconstruct the operational LST dataset under clear-sky conditions on a daily scale ..." –> I don't understand. Why do we want this in case of a clear-sky pixel?

Line 215: "and it is even more difficult to retrieve the LSTs to identify the real performance

of the LST reconstruction ..." –> retrieve LSTs from what? What do you want to say with this sentence?

Response: Thank you for your comments. We are sorry for our unclear expression. We have revised sentence, and the revised Figure is as following.

"*It is difficult to reconstruct the cloudless LST dataset on daily scale, and it is even more difficult to reconstruct the dataset that can reflect the LST of the ground under the cloud cover.*"

Line 217: This "high-precision data set", does this have daily or monthly temporal resolution or both? It is not clear what the main output is.

Response: Thank you for your comments. The new data set is monthly temporal resolution, and we are sorry for our unclear expression. In addition, we also modified this sentence in the original text as follows. We create a reconstruction model that combines meteorological station data and daily and monthly MODIS LST data to reconstruct a high-precision monthly dataset that can reflect the true LST under cloud coverage. The associated explanation has been added to the revised manuscript and can be seen in Subsection 4.3

Lines 219/220: You mention "poor-" and "low-quality" pixels. Are these different?

Response: Thank you for your comments. In the original manuscript they have the same meaning, and we have modified it to "low-quality".

Line 223: What is "traversed"? What are the "corresponding daily pixels"?

Response: Thank you for your comments. The associated explanation has been added to the revised manuscript,

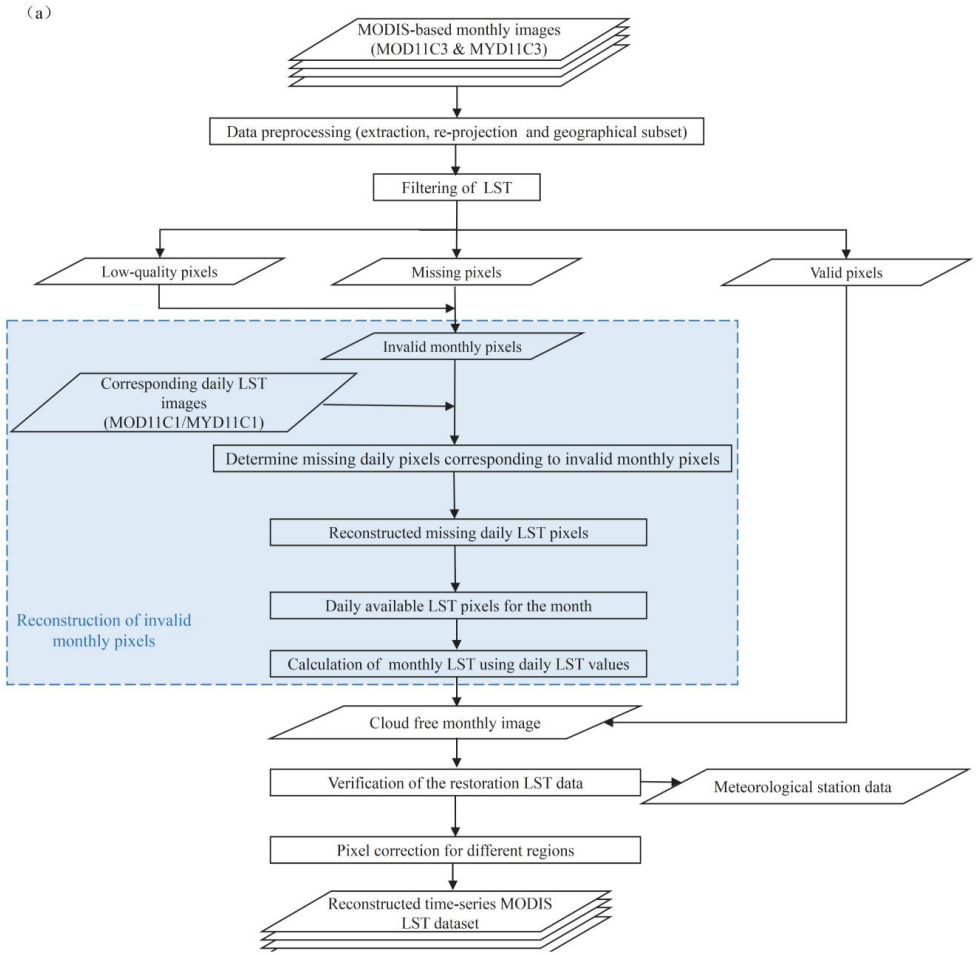
Line 229: I recommend that in the paragraph ending in this line you add references to the subsections (3.3.1, 3.3.2 and so forth) and note that the aspects mentioned here are explained there in more detail.

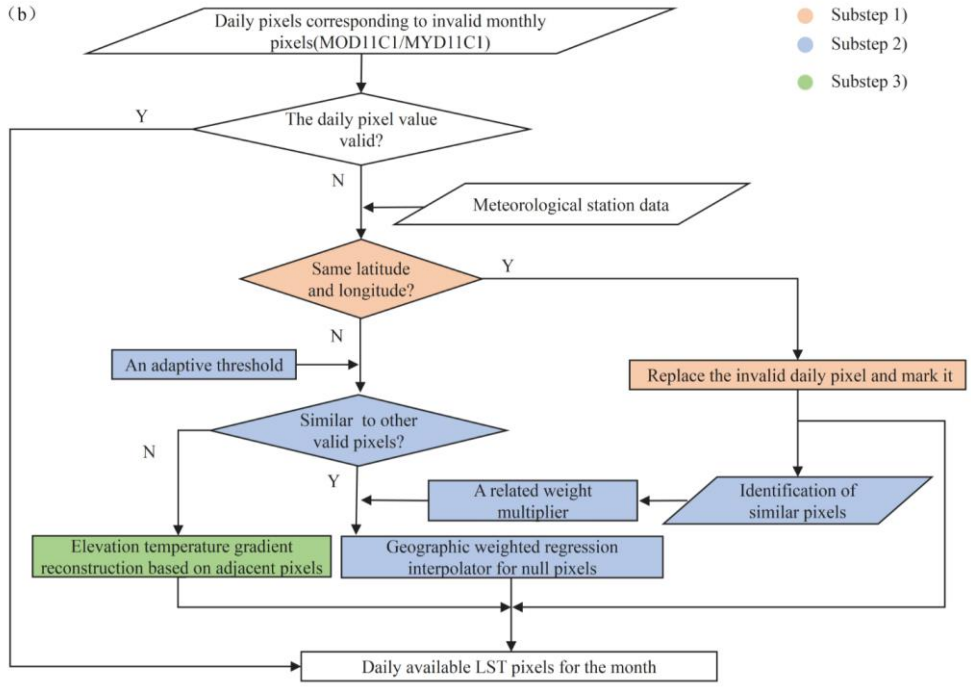
Response: Thank you for your comments. As the reviewer suggested, and we revised these sentences in the revised version.

Figure 2: - What do you mean by "traversal of daily pixels ... separately"? Please be more specific. - The MOD/MYD11C1 daily LST products come in at the side. Are these auxiliary products? - Are the steps below the diamond shaped box with "The daily pixel value valid?" repeated for every day, i.e. is this a loop? - If possible I would try to color those parts of the flowchart which belong to steps I to III described in the text with different colors. - If possible

I would also try to add one more illustrative figure which better explains what you do with the missing and the poor-quality pixel (of the monthly (?) data) in connection with the daily data (possibly of the same pixel but the entire time series). I guess what would greatly help in the understanding of your method if you would further illustrate the steps carried out.

Response: Thanks for your comment and good suggestions, which has helped us to improve the manuscript quality. "Traversal of daily pixels ... separately" means that we determine the invalid pixels in daily LST images at the same location at the corresponding time, and we revised these sentences in the revised version. In addition, as the reviewer suggested, Figure 2 has been revised to explain more clearly. It is also presented below for the ease of reviewing.





"Figure 2: (a)The summary flowchart for reconstructing MODIS monthly LST data, (b) The detailed flowchart for reconstruct missing daily pixels in (a)."

Figures 3 and 4, caption, and interpretation - I suggest to step back from the notion of an exact time for which these maps are valid and instead simply state that this is "daytime". While the local overpass time at the equator is 10:30am, China is a) further north and b) extend over a substantial latitude area. In addition, these maps are composited from MODIS data of several adjacent overpasses.

Response: Thank you for your comments. The Terra satellite passes over the China daily at 10:30 and 22:30 local time, and we agree that "daytime" is more appropriate than "10:30", and we have revised it in the manuscript.

- I suggest to harmonize the color table for both maps. Currently the same temperature range is color coded differently in a) and b). Either you use red over white to blue or red over yellow/green to blue to display the gradient from high to low temperatures. In addition it is very confusing (counter-intuitive) to have high temperatures given by bluish colors (in a). - The white areas in Fig. 3, are these the result of the filtering described in lines 256- 258? If not clarify what is shown please.

Response: Thank you for your comments. We agree that multiple colors usually indicate better. We have also tried using a color table, similar to red to white to blue, to represent the distribution

of temperature values. However, due to the large difference in temperature range between the cold season (January in the manuscript) and the warm season (July as the example), July is concentrated in the high temperature range and January is in the low temperature range. Therefore, the same color coding will concentrate the colors, and the distribution of high and low LST in an image cannot be visually displayed. So, we chose two legends and used green and red to represent low-temperature January and high-temperature July, respectively. In addition, Figure 3 presents the distribution of missing values of unfiltered waves to illustrate that the daily LST data is heavily affected by the cloud.

Lines 247-260: I suggest to be more precise in your wording. - "high-precision LST dataset" should get "daily"

Response: Thank you for your comments. We are sorry for Inaccurate wording and we have modified it .

Line 251: "... composite data." –> Link to Figure 4 is missing.

Response: Thanks a lot for pointing this out. We revised it.

- "identify and reconstruct cloud-contaminated pixels" –> Sure. In which data sets? In the daily or monthly ones?

Response: Monthly LST data set is reconstructed in the text. In addition, we have modified the sentence in the original text as follows.

"It is necessary to identify and reconstruct cloud-contaminated pixels, which seriously affect the use and analysis of monthly LST data."

- Line 255 / 258: Further up you wrote "identify and reconstruct" ... here you write "eliminate to ensure the quality of the LST data" or ... "rejected". Please use one common set of expressions for filters, qualities, and actions undertaken. My assumption is that you identify low-quality pixels (or grid cells) by means of the QA filters. For further analysis you (possibly) set values in all these grid cells as missing values (preferably the same as is used for those pixels which don't have a valid LST value anyways because of cloud coverage). Later, you are replacing the missing values by LST values derived with one of the three methods presented in Section 3.3.2

Response: Thanks a lot for pointing this out. We agree that inconsistent representations in the manuscript may confuse readers, and we have modified the sentence in the original text as follows.

"It is necessary to identify and reconstruct low-quality pixels, which seriously affect the use and analysis of monthly LST data. A reliable method for removing low-quality pixels is implemented using the data quality control information for MODIS LST data. The data quality control information is statistically calculated and stored in the corresponding QA layer and is represented by an 8-bit unsigned integer and can be found in the original MODIS LST HDF files. Therefore, we use the quality control labels for daily and monthly files as mask layers to identify low-quality pixels to ensure the quality of the LST data. Finally, pixels 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. Finally, we reconstructed all the invalid pixels in monthly LST data."

Line 258: "Quality information is almost indicative; thus, sufficient information ..." –> What does this mean?

Response: Thank you for your comments. Quality control information is a reliable method to judge the quality of pixel, it is merely indicative, no other quality information were used to eliminate lower quality LST pixels (Benali et al., 2012).

Again: "poor-quality pixels" are eliminated versus "low-quality pixels" are filtered ...? So to filter is not to eliminate?

Response: Thank you for your comments. Pixel filtering is to identify low-quality pixels and set them as missing values. We have modified the sentence in the manuscript for a better understanding in Page 8 Lines 249-252.

"Therefore, we use the quality control labels for daily and monthly files as mask layers to identify low-quality pixels to ensure the quality of the LST data. Finally, pixels 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. Finally, we reconstructed all the invalid pixels in monthly LST data."

Lines 261-265: - Here you use a new term "invalid". Does "invalid" mean poor-quality or completely cloud covered or low-quality or eliminated or rejected or ...? - Note in line 261 that this is the monthly LST data.

Response: Thank you for your comments. Invalid pixels include missing pixels and low-quality pixels. We also have added corresponding explanations in the text as follows.

"Both missing pixels and low-quality pixels are considered invalid pixels that need to be reconstructed."

- Please go back to Figure 3 and clarify whether the white pixels there in results from your filtering or are an inherent feature of the daily LST product, i.e. in those white areas no daily LST values could be derived?

Response: Thank you for your comments. There is no daily LST values could be derived in the white area. We have revised the related contents as following.

"Figure 3: Spatial distribution of valid data for daily MODIS LST data from Terra during the daytime on (a) January 1, 2017, and (b) July 1, 2017. Areas of missing data are blank.

Figure 4: Spatial distribution of valid data after pixel filtering for monthly MODIS LST data from Terra during the daytime on (a) January and (b) July. Areas of invalid data are blank."

- In Line 275 you have "poor-quality values in monthly pixels"; in lines 280/281 you have "poor-quality daily data ... low-quality pixels in the monthly data" –> inconsistent. It is not clear what you do and why you differentiate between poor- and low-quality.

Response: Thanks a lot for pointing this out. We revised it. In the manuscript, they have the same meaning, which were determined by the pixel filtering, and we have revised them all to "low-quality" in Page 9 Lines 274 and in lines 280/281.

- Lines 275/276: "The contributions of multiple valid daily pixels, despite their good precision, are rejected along with the final poor-quality values in monthly pixels." –> I don't understand what you mean. What I assume is that you refer to a pixel in the monthly LST product, where the QA suggests, for instance, an accuracy of the LST > 2 K. Your filter identifies this pixel. Good. The monthly LST value of the grid cell is based on daily LST values, presumably those you investigate as well. Now it possibly depends on what the criteria (set by the MODIS LST production team) are to use a daily LST value to compute a monthly LST value. It seems to me that this information is not known - otherwise you would have given it. I am sure, however, that a documentation exists where it is written up to which QA flag daily LST values are used in the monthly LST product. Given the fact that this piece of information is not given it is not entirely clear which direction your approach has.

Response: Thank you for your comments. You're right, the monthly LST value of the grid cell is based on daily LST values. Because we get no missing value by reconstructing the day data, and then calculate the monthly value according to the average value of the day, which is not affected by the original method, so there is no introduction to the original data discrimination method.

- Line 281: "from the daily data" –> If I understood you correctly then this would be the "gap-filled" or reconstructed daily data, am I correct? In that case I would mention it here. If not, then I did not understand what you did.

Response: Thank a lot for the comments. You're right.

Lines 297-307: - Lines 301/302: "abrupt transformation" "wheat harvesting" ... "expansion of a city" I can agree that wheat harvesting is an abrupt transition as it happens within a day; time scales for the expansion of a city might not be days, though. Consider rewriting please.

Response: Thanks a lot for pointing this out. We revised it.

- Line 303: "nearest phase" –> What is meant here by "phase"? -

Response: Thank a lot for the comments. We are sorry for the wrong expression. We have revised “nearest phase” as “images from the other three overpass times”.

After equation (1) where the target pixels is the one without a subscript and the "similar pixels" carry a subscript I became confused with the text beginning in Line 303: "for the target pixel i ...". I have difficulties to imagine where pixels i and j come from. I again recommend to add a figure illustrating the process.

Response: Thank a lot for the comments. We added a subscript for the target pixels and also added letter marks.

“$T_t = \sum_{i=1}^{m} W_i \cdot T_i + \sum_{j=m+1}^{n} W_j \cdot T_j$, *(6)*

where T_t *is the reconstructed LST value of a target pixel,* T_i *and* T_j *represent LST values for the similar pixel i and j, the sum of* W_i *and* W_j *values is 1.“*

- Why three images that are temporally closest at the same overpass time? These three images form the "reference images" ... meaning that pixel i is not on them but pixel j? No entirely clear.

- "valid pixel j" –> What is a "valid" pixel and how is it defined? - The concept of the threshold is not clear to me. How is a threshold determined?

Response: Thank a lot for the comments. Both missing pixels and low-quality pixels are considered invalid pixels that need to be reconstructed. valid pixels are reliable pixels, and they can be used in the data reconstruction process. We are sorry for our inadequate interpretation of the threshold making it difficult to understand. We added an explanation of the threshold, which can also be seen below.

"The adaptive threshold φ^{τ}, calculated from the standard deviation, indicates the local area smoothness. Local area is a certain size area centered on similar pixel, which is located in the three reference images. The closer the pixel is, the more similar the environment is, so the smoother the local area will be."

Lines 308-315: - "spectral differences" –> so you compute spectra? From which parameter?

Response: We are sorry for carelessness and we have removed " spectral".

- The "similarity threshold" mentioned in Line 310 is the one referred to in the previous paragraph?

Response: Thank a lot for the comments. Yes, it is the same as mentioned above.

- Line 311: What is a "null pixel"?

Response: Thanks a lot for pointing this out. We revised it as invalid pixel.

- Line 312: "target image" –> So, in addition to the 3 (?) references images we have a target image. I assume that is the image which contains the target pixel and the i-th pixels mentioned in Equation 1? How does this fit with what is written here?

Response: Thanks a lot for pointing this out. You are right that the target image is an image which contains the target pixel.

"Missing daily pixel is defined as the target pixels T_t, the image contains the target pixel T_t is target image."

Lines 318-322: - In addition to target and valid pixels we now also get a "local pixel"–> what is this? In addition we get the "local area" mentioned in line 311 already –> which is located where? On the target or one (or each) of the reference images? The description of the steps given in these last paragraphs also leave open the question: Which parameters are actually used to define "similarity"? LST? Elevation? NDVI? All together? Is "smoothness" a measure of the spatial variability of the respective parameter in (?) the local area? - Line 321: "greater than 4" –> where? in the local area? in one image? Not clear.

Response: Thank a lot for the comments. Local pixel is the mean value of all pixels in local

area. We also revised it in Page 6 Lines 26-28. The adaptive thresholdφ^{τ}, calculated from the standard deviation, indicates the local area smoothness. Local area is a certain size area centered on similar pixel, which is located in the three reference images. We set the range of the local area to 5 pixels by 5 pixels centered on the target pixel. So, the LST values used to define "similarity". It means in one image. The associated explanation has been added to the revised manuscript and can be seen in Subsection 3.3.2.

"Here, we use an adaptive threshold φ^{τ} to determine similar pixels for each invalid pixel (Eq. 3). The adaptive threshold φ^{τ}, calculated from the standard deviation, indicates the local area smoothness. Local area is a certain size area centered on similar pixel, which is located in the three reference images. The closer the pixel is, the more similar the environment is, so the smoother the local area will be. For example, the jth valid pixel in the target image is determined to be a similar pixel of the target pixel i only when the relationship described in Eq. (2) is satisfied in the reference image τ. Simultaneously, similar pixels were determined based on all valid pixels in the image rather than a sliding window because missing values are often arbitrarily clustered in a large area rather than scattered.

$$|P_s^{\tau} - P_t^{\tau}| \leq \varphi^{\tau}, \quad (2)$$

$$\varphi^{\tau} = \sqrt{\sum_{i=1}^{n}(P_s^{\tau} - \varepsilon)^2}, \quad (3)$$

where P_s^{τ} and P_t^{τ} are the values of pixels corresponding to the position of the similar pixel and the target pixel in the reference image, respectively. φ^{τ} is the threshold used to determine similar pixels. ε is the mean value of all pixels in local area. τ is the reference image (value=1, 2, 3). Here, we set the range of the local area to 5 pixels by 5 pixels centered on the target pixel (Zeng et al., 2013). In this paper, the number of similar pixels of the target pixel in the target image should be greater than 4 to apply the GWR method to reduce the error due to an insufficient number of similar pixels."

Line 323-329: - "related weight multiplier" –> related to what? - "relative multiweight values of the ground stations were set to 3" –> Not clear ... relative is often something per 100 ... so 3 percent? I don't understant what the "relative" stands for and I am lost about the "multiweight values" –> are these determined by Eqs. 4 to 6? - In order to understand correctly what you do here: You have an area of missing LST data. You find, e.g., two stations collocated with two of the pixels of this area

of missing LST data. These pixels you assign the LST values from the stations. You keep these pixels in mind [according to what you wrote these pixels would have an LST value for every time step from the station data, correct?]. For re-constructing LST values in all other pixels of the area with missing LST data you proceed with the GWR method. And in this method, you first identify similar pixels and subsequently incorporate (if present) LST (?) data from similar pixels filled with station data and clear sky (?) similar pixels with different weight coefficients M (3 and 1, respectively) to derive weights W. If there is not station around then it is a W_i; if there is a station around and included then it is a W_j ...? Question: What is the threshold or measure which determines whether a pixel filled with a station value is included such that the weights become a W_j?

Response: Thank a lot for the comments. The regression weight coefficient of a similar pixel is determined by its Euclidean distance from the target pixel. In addition, we assign a related weight multiplier to the marked ground station data based on the GWR. "Relative multiweight values of the ground stations were set to 3 " means that for similar pixels that have been assigned to the site, the weight coefficient is obtained by the GWR method, and then multiplied by three to increase the weight of the ground site data. The weight coefficient is determined by Equation (4)-(6). We agree with what the reviewers said about the reconstruction process. The threshold is determined by Equation (2)-(3), calculated from the standard deviation.

Line 338: Is "cloudless contaminated pixel" = "clear sky"? Otherwise this is very confusing. - I note that D in eqs. 5 and 6 carries subscripts i and j while equation 4 makes no difference between i and j; I suggest to change this accordingly. How is D measured? In kilometers or meters? - How connect the W_i and W_j of Eqs. 5 and 6 to Eq. 1? Is this the same W_i? If yes, then how is it ensured that the sum of all W_i equals 1?

Response: Thanks a lot for pointing these out. You are right and we revised it. Equation 4 applies to pixels i and j at the same time. In addition, we modified it in the original text. D represents the Euclidean distance from the similar pixel (i, j) to the target pixel t. We revised the formula and adjusted the structure of this part. It can be seen in Page 6 Lines 26-28.

Lines 340-352: - Is it correct that you apply the method described here only based on the elevation and that this is the general method the be applied in case the 5x5 pixel local area (in whatever image) contains only 4 or less similar pixels? - Does this means at the same time that elevation is not among the similarity criteria used in the GWR method?

Response: Thank a lot for the comments. The elevation temperature gradient regression method was used to reconstruct the remaining low-quality pixels that did not have enough similar pixels. So, the two methods are used in different situations.

- Line 343: "spatial trend" –> It is perhaps a matter of taste but "trend" is something I connect to time series. Here you refer to the spatial variation of LST as function of the elevation, am I correct?

Response: Thank a lot for the comments. Generally, "temporal trend" is related to time series. As you said, "spatial trend" means that the temperature drops vertically when the altitude increases. In addition, similar expressions have also been found (Sun, 2016).

- While I can imagine that the pure elevation has in most cases the dominant influence I am wondering about the impact of the slope and the orientation with respect to solar illumination. Did you take this into account as well? If not, could this be the reason why you require a sliding window of 19 x 19 pixels -which is about 1 degree x 1 degree or 106 km x 106 km - to minimize your noise?

Response: Thank a lot for the comments. We agree that LST is not only closely related to altitude, but also affected by slope and aspect, especially in mountainous areas. Reconstruction of mountainous areas where the topography changes in small areas and slopes has a greater advantage. We mainly reconstructed the missing and low-quality sky pixels by GWR method. For pixels that do nsot have enough similar pixels, a reliable elevation temperature gradient regression method was used. Especially for the plain areas of eastern China, the slope and aspect have less significant influence on the surface temperature than mountain areas. In future research, we will further develop a study on the use of slope and aspect for remote sensing image reconstruction in mountainous areas. In addition, using a sliding window of 19 x 19 pixels can effectively reduce noise and increase calculation speed.

– Did you compute alpha and beta for every pixel and every day for which you need to use this method instead of the GWR method? Or do you compute global values for every grid cell of China for these parameters? Otherwise I have problems to understand the "sliding window" technique. - Did you compute alpha and beta separately for the daytime and nighttime LST data? If not then how did you take into account that nocturnal cooling often results in near surface "lakes" of cold air in areas with sufficient topography, offsetting the classical temperature-to-elevation relationship? The same applies to cases with pronounced inversions.

Response: Thank a lot for the comments. We estimate the temperature value in the area based

on the linear relationship between the elevation and temperature in a certain area. The reconstructed pixel values are obtained based on neighboring pixels and elevation information. SO, we cautiously think do not need to compute alpha and beta for every pixel and every day. It is a problem for the cooling of the surface caused by the nocturnal cooling you mentioned.

- Line 344: "null pixel"? "non-empty pixels"? - Lines 350/351: What do you mean by cropping here? Does that mean that you carry out the computation for an areas which is actually expanded by 9 pixels in each direction beyond the Chinese border?

Response: Thank a lot for the comments. We revised these words. In order to avoid insufficient data in the sliding window at the edge of the study area, we crop the geographic subset of the LST images according to latitude and longitude position of the study area during preprocessing in Section 3.1.

Line 363: "changes" ??? So you mean that, e.g. a negative LST trend during the first half of the period changing towards a positive LST trend during the second half of the period (i.e. a change) causes a large correlation value? Please clarify. Perhaps I misunderstood what Eq. 9 does ... but apparently you do not compare two different data sets but you compute the correlation of the time series with itself? ... You write that "the LST is positively correlated with the time series" ... I guess I don't get what you here ... correlating one annual mean value with the 15-year time series ... ? Perhaps you enter spatial information but I don't understand where.

Response: Thank a lot for the comments. A negative LST trend during the first half of the period changing towards a positive LST trend during the second half of the period (i.e. a change) causes a small correlation value. To some extent, the correlation coefficient can be used to reflect the reliability of the surface temperature trend over time. In order to more clearly express the significance of the trend, student's t-test was performed for different time scales. The significance maps has been added to the revised manuscript and can be seen in supplementary material.

Lines 396-408: This paragraph does not fit to an ESSD paper about a new data set. If at all it would belong either to the discussion and/or into a conclusion/outlook section; see GC4 also.

Response: Thanks for your good suggestion. We agree that the discussion /or into a conclusion/outlook section should be deleted. We believe that through the study of the changing trends of new data sets, readers can have more understanding of our data sets, so we hope this

section can be kept.

Lines 409-417: - This paragraph repeats partly the paragraph where Figure 6 has been described. Please consider to delete it. I have the same view of it as I voiced for the previous paragraph; see GC4. - Without a map where those Taihang Mountains are located it is not possible to take any added value out of what is written about it here.

Response: Thanks for your suggestions , which has helped us to improve the manuscript quality. According to your suggestions, we have deleted these contents.

Lines 418-448: Same as stated for the previous two paragraphs. This is certainly all nice and well collected information but to my opinion this journal is the wrong place to lay out these issues in that degree of detail. Consider deleting it and - if you feel it needs to be included - summarize the key messages in 3-4 sentences in the conclusions; this would be my recommendation for the while part of lines 396-448; see GC4.

Response: According to your suggestions, we have deleted these contents.

Lines 450-462 / Figure 7: - I note that you show the correlation in Fig. 7 b) but do not comment on it in your text. You might consider to not show this parameter or put it into supplementary material. However, isn't it interesting to note that while Fig. 6 a) appears to be dominated by the daytime LST changes Fig. 6 b) resembles a lot of the pattern shown by the nighttime R (Fig. 7 b). Do you / we understand why this is the case?

Response: Thank a lot for the comments. To some extent, the correlation coefficient can be used to reflect the reliability of the surface temperature trend over time. We hope that the value of r indicates the reliability of the change. In addition, according to Dr. He's opinion, we have added saliency maps in the supplementary material, hoping to provide some reference.

Yes, it is also interesting that the changes and R in the daytime LST are similar to the interannual LST, which is mainly due to the small changes of LST at night. The annual daytime positive /negative trends of LST in almost all regions from 2003 to 2017 are significantly higher than those in the evening (-0.03<slope <0.03); thus, the average LST positive/negative trends can be attributed to changes during the daytime.

- I note that the color bar below the panels is twice as large as for, e.g., Fig. 6 a, b but that an annotation of what is shown by the color bar is missing

Response: Thank you very much for your careful review. There is no missing comment on the

content shown in the color bar we see. If there is any deviation in understanding or displaying version, please contact us and we will continue to change it.

- I suggest to tie the text closer to the Figure by explicitely referring to it as Fig. 7 a) day and Fig. 7 a) night; it might be easier actually to use panels a) through d) instead of a) and b).

Response: Thanks for your good suggestions, we revised it.

"Figure 7: Spatial dynamics of Fig. 7 a) day and Fig. 7 a) night LST change trends based on slope (a) and correlation coefficient (b)."

- I find it interesting that the negative daytime LST trend in the northern part of region IV is offset by a positive nighttime LST trend.

Response: I'm sorry we didn't find the opposite phenomenon you point out.

- Line 451: "average annual diurnal surface temperature" , also in Line 450 you use "diurnal" –> I am not convinced that usage of "diurnal" is adequate here. I suggest to separately refer to day- and nighttime LST because diurnal implies that you know more about the diurnal development of the LST - but since you only use data of one MODIS sensor here you only have two anchor points

Response: Thank you for your comments. Adopted.

- Line 454: "evening" –> "nighttime"; in addition: there appears to be a substantial fraction (25% or even more) where the absolute slope is larger than 0.3 in Fig. 7 a) night; the range given in parenthesis appears not be correct therefore.

Response: Thank you for your comments. Adopted; we are sorry for the careless expression we have revised the related contents as following.

"The annual daytime positive/negative trends of LST in most regions from 2003 to 2017 are significantly higher than those in the nighttime."

- Line 456: "daytime human activities" –> what does this have to do with LST observations?

Response: Thank you for your comments. We have changed "daytime human activities" into "daytime human production " .

Lines 464-478 / Figure 8 - Please check Figure 8 a). It cannot be that the average daytime LST in southern China is around 0degC.

Response: Thank you for your comments. I am puzzled by what you said about "0degC ". We have carefully checked Figure 8 and the average daytime LST in southern China is around 25 degC.

- I note that quite a large fraction of the LST shown in Fig. 8 b) is saturated at the lower end of the color scale shown. I suggest to expand that accordingly. I don't think it would hurt to cut off the legend at +20degC and add the respective two 5 Kelvin bins at the left hand side. - Why did you use such 5 Kelvin wide bins? Is the LST distribution too noisy otherwise? - All legends lack the annotation of what is shown in which unit.

Response: Thank you for your comments. In order to more directly reflect the difference in temperature difference between daytime, nighttime and day and night temperature difference, we use the same legend for the three figures. We cautiously believe that the current distribution of temperature in the same image and between different images can be well reflected. In addition, we will add units to all the graphs.

- Line 470/471: The stated latitudinal dependence is only evident in the eastern half of China - possibly due to its comparably smoother relief.

Response: Thanks a lot for pointing this out. We revised it.

"This result may suggests that the spatial temperature change is related to the latitude range, and also to the smooth Terrain in the east."

- Line 475: Here you use "temperature difference between day and night" –> This appears to be a better wording than diurnal. - Finally, I suggest to stress that you look at a multiple-years average of the annual mean LST and that hence individual day-to-night changes can be much larger (or smaller) depending on season - to have an adequate link to the next section。

Response: Thanks for your good suggestions. As the reviewer suggested, we have modified these words in the revised manuscript.

Lines 479-489: Same comment as for Lines 396-408. Hence, remove the Hu line in Fig. 8 c); see GC4.

Response: Thank you for your comments. we have deleted these contents.

Section 4.2: - Please check for GC - wording/editing; there are many increasing and warming trends here. - I find this section overly long and suggest to condense the material to the key elements, omitting any links to changes in crop yield or the like and omitting attempts to explain observations with changes in atmospheric circulation / precipitation or the like. If at all then this can be included in a condensed paragraph in the conclusions.

Response: Thank you for your comments. As the reviewer suggested, we have revised the related

contents as following.

"Specifically, in spring, the warming area is mainly concentrated in the northern areas (I, II, and V), while a weak negative trend is observed in the southern areas. The largest positive trend over the northern areas appears in the Inner Mongolia Plateau (slope >0.18, P < 0.01). In addition, rapid warming also occurred in the North China Plain in the eastern part of the North China Region (II) (especially near Beijing and some areas of Hebei Province, slope >0.12, R > 0.6, P < 0.01)f

As shown in Fig. 9, compared with the other two seasons, both summer and autumn showed weak positive trends throughout the country. In summer (Fig. 9b1, b2 and b3), there were slight increasing trends in most areas of China, while there were still negative trends in the Northeast Region (I) (details in Fig. 9). Significant increasing trends were mainly observed in the Qinghai-Tibet Plateau, North China Plain, Inner Mongolia Plateau, Tarim Basin and some areas in the north, with the largest positive trend in the Qinghai-Tibet Plateau. In autumn, the negative trends were mainly present in the Northeast Region (I) and the Northern Chinese Tianshan Mountains in the Qinghai-Tibet Plateau Region (VI). In contrast, the Qinghai-Tibet Plateau was still controlled by strong positive trends (near Lhasa city, slope=0.09, R=0.60, P < 0.05), especially in the southern part of the Tanggula Mountains.

In winter, 69.4 % of the areas experienced warming, which is significantly higher than in other seasons; thus, winter is the most important source of interannual increases in the average LST. The most remarkable positive trends in winter were observed in the Northwest Region (V) and the Qinghai-Tibet Plateau Region (VI)."

Lines 561/562: "dramatic" and "rapid" –> Please reconsider these formulations because "dramatic" appears not adequate and "rapid" implies that something happens particularly fast but it appears that it is rather the magnitude of the trend which strikes here. Ines.

Response: Thank you for your guidance. They're all part of the analysis of temperature effects, and they've been removed

Section 4.3: - While I see the merit to also take a look at the months it appears to me that the paper should contain an analysis of either seasonal or monthly distributions / changes in the multi-annual development of the LST. I tend to favor the seasons and to delete Figure 10 and this section for the sake of keeping more space for the illustration of the method and the evaluation.

Response: Thank you for your comments. According to your suggestion, we have deleted these

contents.

Response: Thank you for your comments.

What is the temporal resolution of the MODIS LST data shown in Fig. 11? Daily or monthly? Please indicate in the figure caption. - What is the time period used?

Response: Thanks a lot for pointing this out. It is monthly MODIS LST data shown in Fig. 11. We have added descriptions of temporal resolution in the figure caption and original text.

- Single data pairs obscure each other. I suggest to plot two sets of panels, one with the re-constructed values (currently in blue) and one with the linar model corrected values (currently in grey), in which you show the count per data value bin, i.e. my suggestion is to plot a 2-dimensional histogram, using a bin-size of the LST values of 1 Kelvin and displaying the count of data pairs falling into these bins with a color code. This way one would have a better impression where data pairs concentrate. Currently, I find it un-natural that we see for all regions - excpet region I - an elongated cluster of points with an almost constant width across; only region I shows some variation here and is the most credible of the panels shown.

Response: Thank you for your comments. Although single data pairs obscure each other, it can be intuitively judged by the change of the slope of the site that the quality of the point after the linear model is stretched has been significantly improved. At the same time, we only showed a rough distribution here, and the specific difference can be obtained according to the numerical indicators(R^2, RMSE, MAE).

- Line 547: "better" can be removed.

Response: Thanks. We have revised.

- Lines 557/558: This last sentence is not supported by Figure 11.

Response: Thank you for your comments, and we have deleted these contents.

Lines 559-571: - Lines 566-571, beginning with "Simultaneously ..." can be deleted; see GC4.

Response: Thanks a lot for pointing this out. We revised it.

Lines 572-581: - Please describe how this comparison is carried out. It is not known how you computed the seasonal averages and then finally carried out the comparison. The number of seasonal LST values per station entering this comparison is not clear; one can guess that per station it is 3 months (per season) times the number of years, i.e. 15. It is furthermore not clear how you dealt with data gaps in the original LST data in this comparison. Also, when using the original LST data,

did you take into account the quality flags as users should do when using the data?

Response: Thanks a lot for pointing this out. We are sorry that we have ignored the description of the division of seasons in the text.

"The original MODIS monthly LST data were used directly without filtering quality flags. For the original MODIS LST images, we averaged the LST data of the month corresponding to the season, and obtained the seasonal LST images. The pixels with missing LST values in original MODIS LST images for the corresponding months of the season were not used in the verification process. Therefore, if there is no missing value for the LST pixel corresponding to the site, each station can have a maximum of 15 values in each season."

Lines 594/595: I don't agree to this statement. Why should the reconstruction of the LST be particularly vulnerable in areas (the red ellipses) where the absolute LST changes over the period chosen are at maximum? These areas (red ellipses) in fact contain regions of large topographic complexity, yes, but this is possibly not the only error source which needs to be discussed.

Response: Thank you very much for your careful review. We are sorry for the statement on Line 594 and we have revised it as follows. Thank you for your guidance.

"We also note that the selected ground stations shown in Table 2 located in six key zones are examples of where the local LST warming/cooling rate changed by more than the average rate, and these areas actually include areas with greater Terrain complexity."

Lines 597-607: This steps is not transparently enough explained and appears not justified; see GC5.

Response: Thank you for this comment. We have revised and deleted some contents in the manuscript. The revised statement is as follows.

"Moreover, the examples indicate that the reconstruction model proposed here is effective even in the areas of complex topography."

Typos / editorial comments:

Lines 56/57: These lines read as if the sensor is used in the models but I assume it is the LST derived from the MODIS observations which is used in the models; please reformulate accordingly.

Response: Thank you very much for your careful review. As the reviewer suggested, we have modified the sentence in the revised manuscript.

"Due to its optimal temporal and spatial resolution throughout the world, the Moderate Resolution

Imaging Spectroradiometer (MODIS) sensor has become an excellent data source for satellite-derived LST data, and the MODIS LST values are widely used in regional and global climate change and environmental monitoring models (Tatem et al., 2004; Wan et al., 2014).”

Line 65: "reconstruction of noise contaminated ..." –> "reconstruction of the LST of noise-contaminated ..."

Response: Thank you for your comments. we have modified the sentence in the revised version for better understanding.

“Thus, reconstruction of these missing and low-quality LST pixels low-quality is necessary for satellite-derived LST applications.”

Line 154: Typo: "mm" –> needs to be micrometer

Response: Thank you for your comments. we have modified this word in the revised version.

Line 231: "a thermal infrared band" –> this contradicts the information given further up, where you explained the day/night algorithm used for MODIS LST retrieval since V006.

Response: Thank you very much for your careful review. For the LST V006 version products used in this article, seven thermal infrared bands (bands 20, 22, 23, 29, 31, 32, 33) were used for retrieving LST values. We apologize for the carelessness, and we revised this term in the revision version.

“MODIS LST data are retrieved from thermal infrared bands in clear-sky conditions and contain many missing values and low-quality values caused by clouds and other atmospheric disturbances.”

Lines 232/236: You need to clarify what you mean by "atmospheric disturbance"; for some people this would be a low pressure system ... I doubt that this is what you mean as you give this in addition to clouds.

Response: Thank you for your comments. Atmospheric disturbance means the influence of atmospheric molecules and aerosols.

Line 234: Markus needs to be Markus et al.

Response: Thank you for your comments. We have revised the wording.

Line 236: Why is "illumination" a problem if we talk about IR data?

Response: Thanks a lot for pointing this out. Thermal infrared remote sensing data is not affected by illumination conditions. We have corrected this error in the revised version of the manuscript.

Line 359: "LST image time series" –> "LST time series"

Response: Thank you for your comments. We have revised the wording.

Line 360: suggestion: "A positive slope indicates an increase in LST (warming); a negative slope indicate a decrease in LST (cooling)." This avoids to write something about trends which become warmer.

Response: Thank you for your comments. As the reviewer suggested, we have modified this sentence in the revised manuscript.

Lines 384-395: - Please stick with "areas" and do not mix "areas" and "districts". - See GC - wording - Consider to use again "area" or "region" instead of "pattern".

Response: Thank you for your comments. As the reviewer suggested, we have modified these words in the revised manuscript.

- Line 391: I doubt that the area with a slope > 0.05 degC/year (panel a) coincides with the area with R > 0.6 (panel b).- Line 394: Mentioning of R < -0.6 appears to be not that informative here? - The slopes lack a unit.

Response: Thank you for the valuable comment. To some extent, the correlation coefficient can be used to reflect the reliability of the surface temperature trend over time. In order to more clearly express the significance of the trend, we further calculated the significance according to the correlation, and student's t-test was performed for different time scales. The significance maps have been added to the revised manuscript and can be seen in supplementary material. In addition, we added the unit in the revised manuscript.

- Figure 6: I particularly like panel (c). Cool! However, also here the unit of the warming / cooling regimes should be degC / year, right? Annotation "slope" and "R" in panels (a) and (b) is too small. The slope lacks a unit.

Response: Thank you for your appreciation of our Figure. We revised the units and annotation in Figure 6.

Thank you again for your efforts in reviewing our manuscript. We hope our modification will be to your satisfaction.

References:

Kilibarda, M., T. Hengl, G. B. M. Heuvelink, B. Gräler, E. Pebesma, M. Percec Tadi ˇ c, and B. Bajat (2014), ´ Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution, J. Geophys. Res. Atmos., 119, 2294–2313, doi:10.1002/2013JD020803.

Benali, A., Carvalho, A. C., Nunes, J. P., Carvalhais, N., and Santos, A.: Estimating air surface

temperature in Portugal using MODIS LST data. Remote Sens. Environ., 124, 108-121, https://doi.org/10.1016/j.rse.2012.04.024, 2012.
Sun, Z., Wang, Q., Batkhishig, O., and Ouyang, Z.: Relationship between evapotranspiration and land surface temperature under energy-and water-limited conditions in dry and cold climates. Adv. Meteorol., 1-9, 1835487, https://doi.org/10.1155/2016/1835487, 2016.
Benali, A., Carvalho, A. C., Nunes, J. P., Carvalhais, N., and Santos, A.: Estimating air surface temperature in Portugal using MODIS LST data. Remote Sens. Environ., 124, 108-121, https://doi.org/10.1016/j.rse.2012.04.024, 2012.
Barry B R G. Mountain Weather and Climate[M]// Mountain weather and climate. Routledge, 1992.
Körner C. The use of 'altitude' in ecological research[J]. Trends in Ecology & Evolution, 2007, 22(11):0-574.
Hansen, J., Ruedy, R., Sato, M. & Lo K. Global surface temperature change. Rev. Geophys. 48, RG4004, doi:10.1029/2010RG000345 (2010).
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).
Hansen, J.E. and Lebedeff, S. Global trends of measured surface air temperature. J. Geophys. Res. 92, 13345-13372, doi:10.1029/JD092iD11p13345 (1987).
Hansen, J., Ruedy, R., Sato, M. and Lo, K.: Global surface temperature change. Rev. Geophys., 48.4, 1-29, https://doi.org/ 10.1029/2010RG000345, 2010.
Smith, T. M., Reynolds, R. W., Peterson, T. C., and Lawrimore, J. Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880-2006). J. Climate 21, 2283–2296, 2008.
Hansen, J.E. and Lebedeff, S. Global trends of measured surface air temperature. J. Geophys. Res. 92, 13345-13372, doi:10.1029/JD092iD11p13345 (1987).
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).
Hansen, J., Ruedy, R., Glascoe, J. and Sato, M. GISS analysis of surface temperature change. J. Geophys. Res. 104, 30,997–31,022, doi:10.1029/1999JD900835 (1999).
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).
Hansen, J. et al. Global temperature change. PNAS, 103, 14288-14293, doi:10.1073/pnas.0606291103 (2006).
Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and 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).
Fan, Y., and 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).
Jones, P.D., New, M., Parker, D.E., Martin, S., and Rigor, I. G. Surface air temperature and its changes over the past 150 years. Rev Geophys 37, 173–199 (1999).
Ghafarian Malamiri, H.R.; Rousta, I.; Olafsson, H.; Zare, H.; Zhang, H. Gap-Filling of MODIS Time Series Land Surface Temperature (LST) Products Using Singular Spectrum Analysis (SSA). *Atmosphere* 2018, *9*, 334.