Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2019-155-RC3, 2020

Open Access Earth System Science Data Discussions

Interactive comment on "A combined Terra and Aqua MODIS land surface temperature and meteorological station data product for China from 2003–2017" *by* Bing Zhao et al.

Anonymous Referee #2

Received and published: 6 March 2020

Review of

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

by

Zhao, Bing, et al.

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

Printer-friendly version

Discussion paper

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.

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.

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 rightaway with the presentation and discussion of the annual national-mean LST.

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.

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.

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 topograpic 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 re-constructed 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.

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. - 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.

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? ...

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

Line 89-92: I guess this sentence is one of the key motivations of this paper and should be expanded, e.g. by asking / answering the questions: How valid is it to use clear-sky LST retrievals in any sort of LST re-construction of grid-cells flagged unreliable, e.g. due to clouds? What is the effect of this assumption during day / night?

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

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.

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.

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.

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?

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.

Lines 193-198: 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. - Please mention why you don't use data before 2003. TERRA MODIS LST data begin in Feb. 2000 and AQUA MODIS LST data begin in July 2002.

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

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? - 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 heterogenuous surface properties in the 500 m grid cell. Here we talk about a grid cell an order of magnitude larger! - 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.

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.

Line 211: I note that for the verificaiton 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 underflights, other well-calibrated IR temperature measurements?

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?

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.

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

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

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.

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.

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. - 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.

Lines 247-260: I suggest to be more precise in your wording. - "high-precision LST dataset" should get "daily" in addition; otherwise this sentence is confusing. - "spatiotemporal integrity" –> what is this? What do you mean by "consistency"? Isn't consistency something you refer to when comparing time-series of long-term data sets derived from different products, figuring out that the time-series, for instance, consistently follow each other through positive and negative anomalies? - Line 251: "... composite data." –> Link to Figure 4 is missing. - "identify and reconstruct cloud-contaminated pixels" –> Sure. In which data sets? In the daily or monthly ones? - 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 VALUE (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 Line 258: "Quality information is almost indicative; thus, sufficient information ..." –> What does this mean? Again: "poor-quality pixels" are eliminated versus "low-quality pixels" are filtered ...? So to filter is not to eliminate?

Lines 261-265: - Here you use a new term "invalid". Does "invalid" mean poor-quality or completely cloud covered or low-quality or elimiated or rejected or ...? - Note in line

Printer-friendly version

Discussion paper

261 that this is the monthly LST data. - Please go back to Figure 3 and clarify whether the white pixels therein 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?

Lines 273-282: - Line 275: "precision" or accuracy? What are you referring to here? Are you sure you talk about precision? - 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. - 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. - 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.

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. - 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. - 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". - 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)?

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

Lines 308-315: - "spectral differences" –> so you compute spectra? From which parameter? - The "similarity threshold" mentioned in Line 310 is the one referred to in the previous paragraph? - Line 311: What is a "null 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? –> illustrative figure please with a clear notion which images a called how and which pixels a located on which images.

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.

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 clearsky (?) 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? Line 338: Is "cloudless contaminated pixel" = "clearsky"? 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?

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? - 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? - 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? - 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 day-time 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. - Is there a quality check with which you compare the LST values re-constructed by this method with the LST values re-constructed in the adjacent pixels using the GWR method? - Line 344: "null pixel"? "non-empty pixels"? - Lines 351/351: 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?

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.

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.

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.

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.

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

Lines 464-478 / Figure 8 - Please check Figure 8 a). It cannot be that the average daytime LST in southern China is around 0degC. - 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. - Line 470/471: The stated latitudinal dependence is only evident in the eastern half of China - possibly due to its comparably smoother relief. - 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.

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

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. Figure 9: - Please add annotations to the color bars. - Please

harmonize the font size of the annotations in the middle row. - I suggest to use grey (for the bars) and black (for the lines) in the middle row to avoid confusion with the colors used in the maps to the left and right. - Please check the values plotted for R in the middle row. They appear to be not correct. Examples: Autumn: region I versus region IV and region V versus region VI. - I don't agree with the caption for the middle column. These are not histograms but these panels appear to be graphs illustrating the regional average values of the slope and R.

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.

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.

Lines 546-558 / Figure 11: - Please state exactly what we see in this figure. The subset of station data used for the verification is 480 in total. How do these distribute of the six regions shown? Please provide the number (or count) of the stations actually used and the total count of the data pairs shown together with the statistical parameters in each panel. - 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? - 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. - Line 547: "better" can be removed. - Lines 557/558: This last sentence is not supported by Figure 11.

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

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?

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 absolut 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.

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

Typos / editoral 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. Line 65: "reconstruction of noise-contaminated ..." –> "reconstruction of the LST of noice-contaminated ..."

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

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 C006.

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.

Line 234: Markus needs to be Markus et al.

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

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

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". - 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. - 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.

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2019-155, 2019.

Printer-friendly version

Discussion paper