Response to Reviewer #4

We appreciate a lot for your efforts in providing detailed comments and recommendation. They are very helpful to improve the quality of the manuscript. We have revised the manuscript according to your comments. The comments from the reviewers are kept in regular font, our responses use blue highlighting, and the revised sentences or words in the revised manuscript are highlighted with red color.

This paper developed a global historical twice-daily LST dataset (GT-LST) with a spatial resolution of 0.05° from 1981 to 2005. I believe this is an important study and it does make sense for earth science communities. The data and methods are clearly described, and the main results are well presented. However, there are some issues that need to be addressed or clarified before the paper can be published. Therefore, I recommend a major revision.

Some major comments:

1. This paper has inter-compared the GT-LST and MODIS LST over a variety of land cover types such as savannas and cropland/natural vegetation, permanent snow and ice, water bodies and etc., yet I wonder how much are the accuracies (such as RMSD and bias) over urban surfaces?

Response: Thanks a lot for your comments. To clearly quantify the RMSD and bias over various land cover types, we have redrawn Fig. 7, including urban and built-up lands (UBL). For your convenience, we listed it below. For UBL, the RMSD and bias are 3.4 K and 2.7 K, respectively.

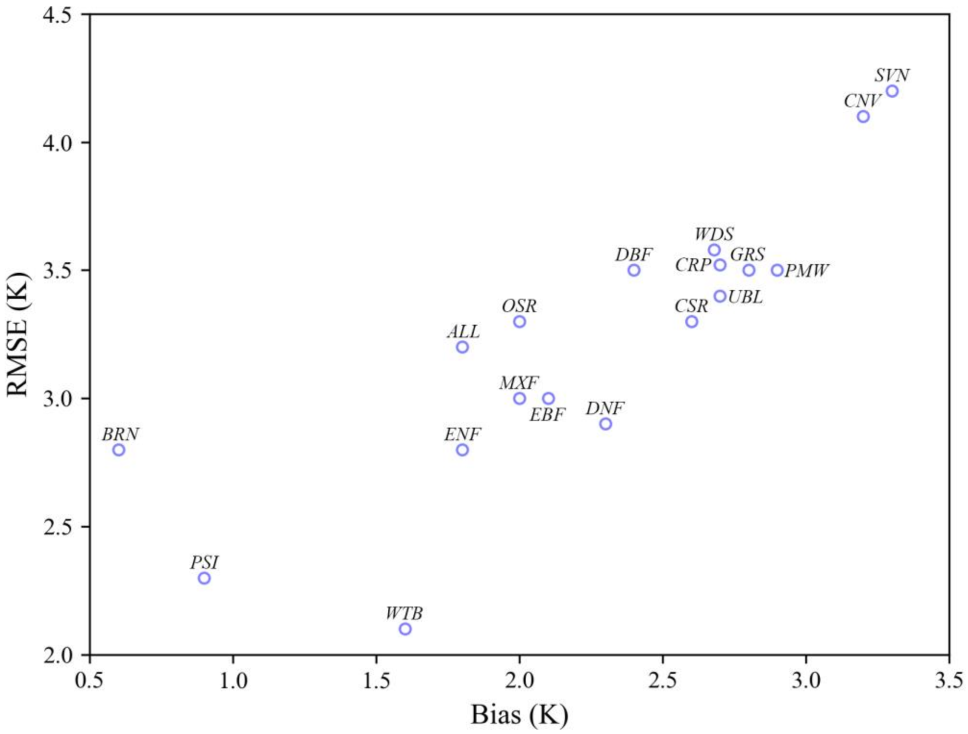


Figure 7: RMSD and bias between GT-LST and MYD11A1 LST in 2003 for various land cover types. ENF: evergreen needleleaf forests, EBF: evergreen broadleaf forests, DNF: deciduous needleleaf forests, DBF: deciduous broadleaf forests, MXF: mixed forests, CSR: closed shrublands, OSR: open shrublands, WDS: woody savannas, SVN: savannas, GRS: grasslands, PMW: permanent wetlands, CRP: croplands, UBL: urban and built-up lands, CNV: cropland/natural vegetation mosaics, PSI: permanent snow and ice, BRN: barren, WTB: water bodies, and ALL: all land cover types.

2. Why did you choose January 15 and July 15, 1997 for the GT-LST and RT-LST comparison over continental Africa? Please clarify the selection criteria.

Response: Thanks a lot for your comments. The RT-LST is a twice-daily LST product at 8-km resolution over continental Africa, which spans of 6 years from 1995 to 2000. Two days, January 15 and July 15, 1997, were chosen because they represent the median time of different seasons (winter and summer, respectively). In addition, the number of matchups is enough to guarantee the reliability of the intercomparison. We have added some descriptions in Line 334-335 as follows:

"Two days, January 15 and July 15, 1997, were selected to implement the comparison over continental Africa because they represent the median time of different seasons (winter and summer, respectively)"

3. I just suggest combining Figs. 3 to 7 into a single figure for clarity.

Response: Thank you for the suggestion. We have redrawn the schematic of the workflow according to your suggestion. For your convenience, we listed it below.

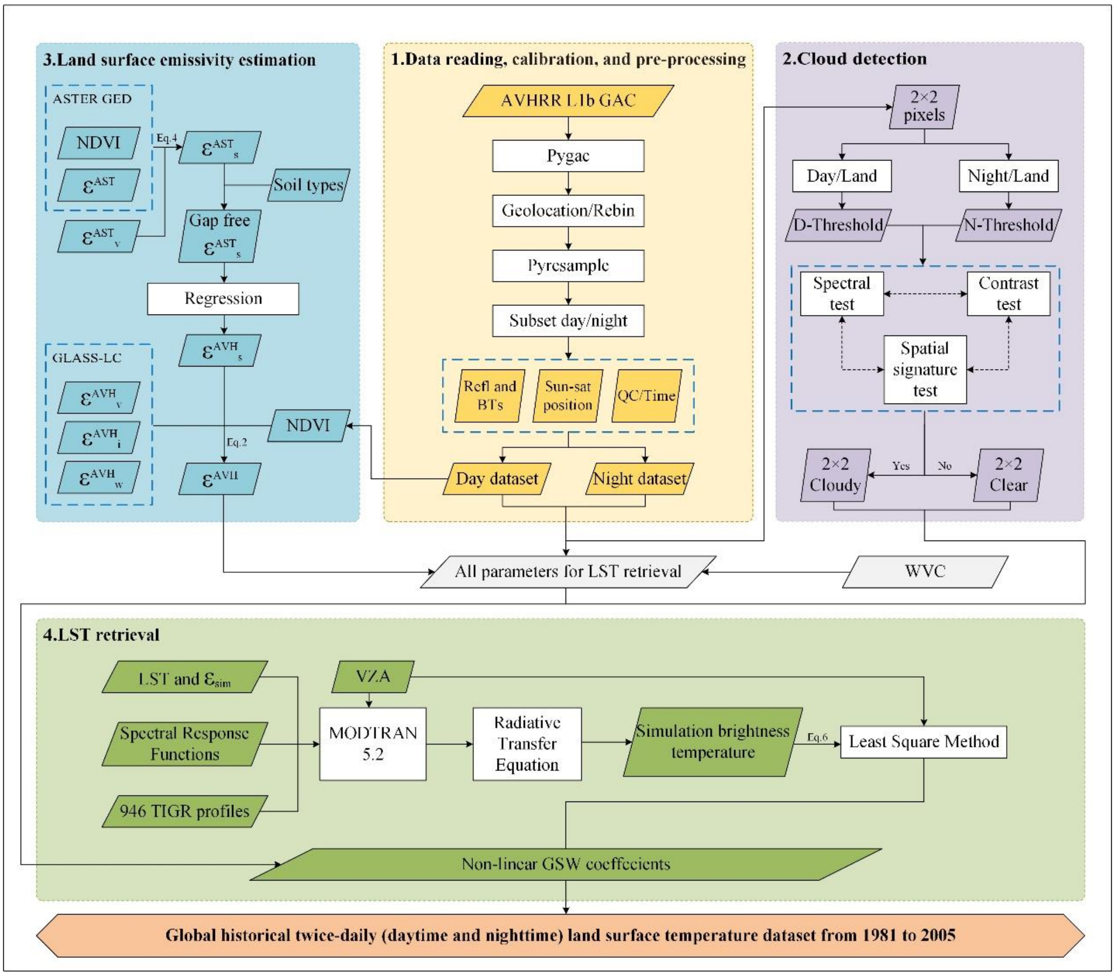


Figure 3: Schematic of the workflow used to generate the GT-LST product

4. To what extent the differences in the emissivity between MODIS LSTs and GT-LST will influence their inter-comparison results, can you provide some quantitative results?

Response: Thanks a lot for your comments. From Fig.R1, one can conclude that a negative relationship between the GT-LST and MYD11A1 LST difference and their corresponding emissivity difference. The mean biases (GT-LST – MYD11A1) for LSTs calculated with emissivity differences less than -0.05, between -0.05 and -0.03, between -0.03 and -0.01, between -0.01 and 0.01 and more than 0.01 are 7.0, 4.3, 2.3, 0.8 and 0.7 K, respectively. We have added some descriptions in Line 406-408 as follows:

"*As a result, the dynamic emissivity of GT-LST is typically lower than that of MYD11A1, which leads to overestimation of the LST (Hulley et al., 2016; Guillevic et al., 2014; Reiners et al., 2021; Ren et al., 2011). Fig. A3 shows that the mean biases (GT-LST – MYD11A1) for LSTs calculated with emissivity differences less than -0.05, between -0.05 and -0.03, between -0.03 and -0.01, between -0.01 and 0.01 and more than 0.01 are 7.0, 4.3, 2.3, 0.8 and 0.7 K, respectively.*"

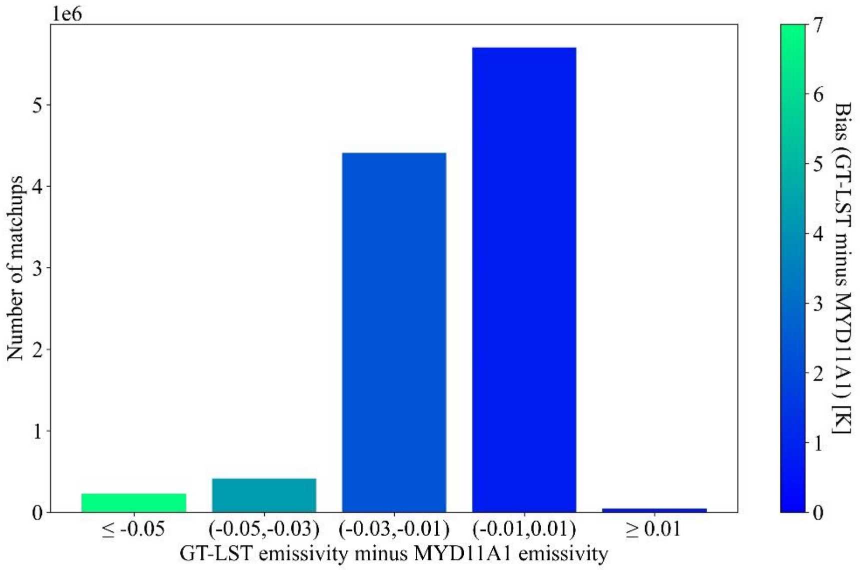
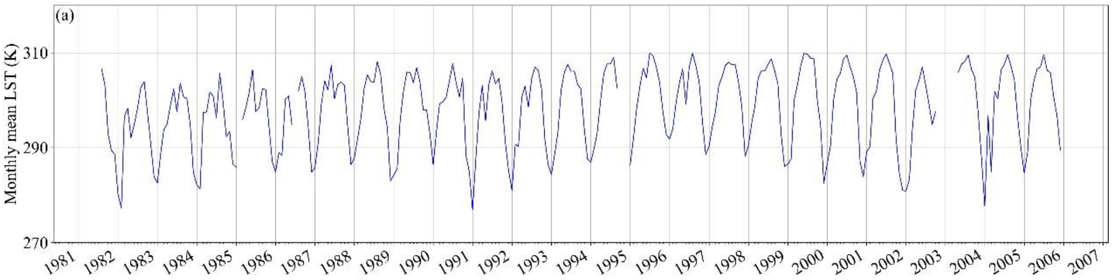


Figure A3: Difference between GT-LST and MYD11A1 LST stratified by the difference between GT-LST and MYD11A1 emissivity (water vapor content < 5 g/cm^2; satellite zenith angle < 50°)

5. As you stated, the LSTs for a long period such as > 40 years are important for monitoring and evaluating global long-term climate change. Thus, the validation of tendency consistency for the generated GT-LST products is also of vital importance in addition to its spatiotemporal pattern. Could you test the accuracy of time series GT-

LSTs over several typical regions, as I guess the orbit drift of AVHRR could also introduce uncertainty for the tendency estimation.

Response: Thanks a lot for your comments. Indeed, global long-term climate change requires daily, monthly or annual mean LST (i.e., DMLST, MMLST, and AMLST) more than instantaneous LST as these mean LSTs are key indicators when monitoring global LSTs over a long time series (Li et al., 2023; Liu et al., 2023; Xing et al., 2021). Impacting of the NOAA satellite orbital drift, daytime and nighttime observations of NOAA afternoon satellites cannot represent maximum and minimum temperatures well. Therefore, calculating the daily and monthly mean LST by averaging daytime and nighttime LSTs derived from GT-LST has a significantly lower accuracy than other studies (Fig. A4). Inspired by the work of Xing et al. (2021), we use simple linear combinations of daytime and nighttime LST values that were observed at observation times for NOAA to estimate DMLST and MMLST. In order to validate the accuracy of DMLST and MMLST according to the simple linear regression method, we compared DMLST and MMLST derived from GT-LST with that of in situ LST observations from SURFRAD sites, and reported RMSE values of approximately 2.4 K and 2.7 K, respectively. These results are similar to that of Xing et al. (2021) and Chen et al. (2017). In this way, we still obtain accurate DMLST and MMLST without satellite orbit drift correction. In order to demonstrate the tendency consistency of GT-LST products, Fig. R1 shows time series of MMLST using the simple linear regression method from 1981 to 2005 for two small area (5°×5°) in the Sahara Desert and the Tibetan plateau: no significant inconsistencies can be seen.



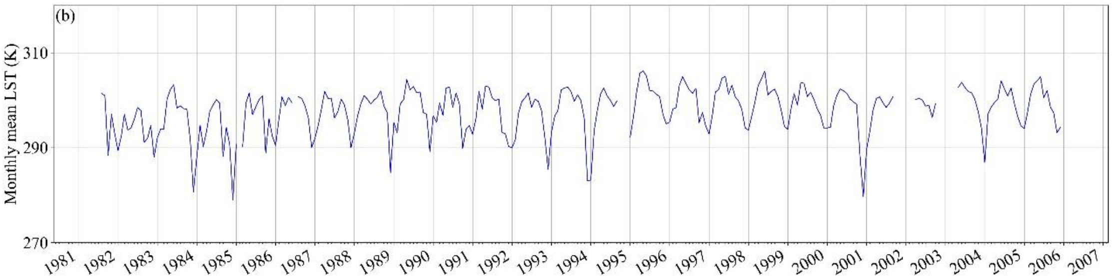


Figure R1: Monthly mean LST time series for 1981 to 2005 over two small area (5°×5°) in the Sahara Desert (a) and the Tibetan plateau (b).

Then, we rephrase the paragraph in Line 429-436 as follows:

"…*To estimate MMLST, first obtain the mean instantaneous clear-sky LST at daytime and nighttime, and then use these mean values to estimate MMLST according to the simple linear regression method (see Appendix B). In order to validate the accuracy of MMLST results, we compared MMLST based on GT-LST with that of in situ LST observations from SURFRAD sites for 1994–2005. All in situ LST measurements are all-sky and complete on a certain month, which means that the in situ MMLST is true MMLST. Fig. 15 showed that MMLST derived from GT-LST are related to the true MMLST, with an* R^2 *value of 0.94 and an RMSE value of 2.7* K. *This result is similar to that of Chen et al. (2017), who compared MMLST from MODIS day and night instantaneous clear-sky LST with actual MMLST from 156 flux tower stations, and reported RMSE bias values of approximately 2.7 K.*"

We have redrawn Fig. 13 according to the simple linear regression method. For your convenience, we listed it below.

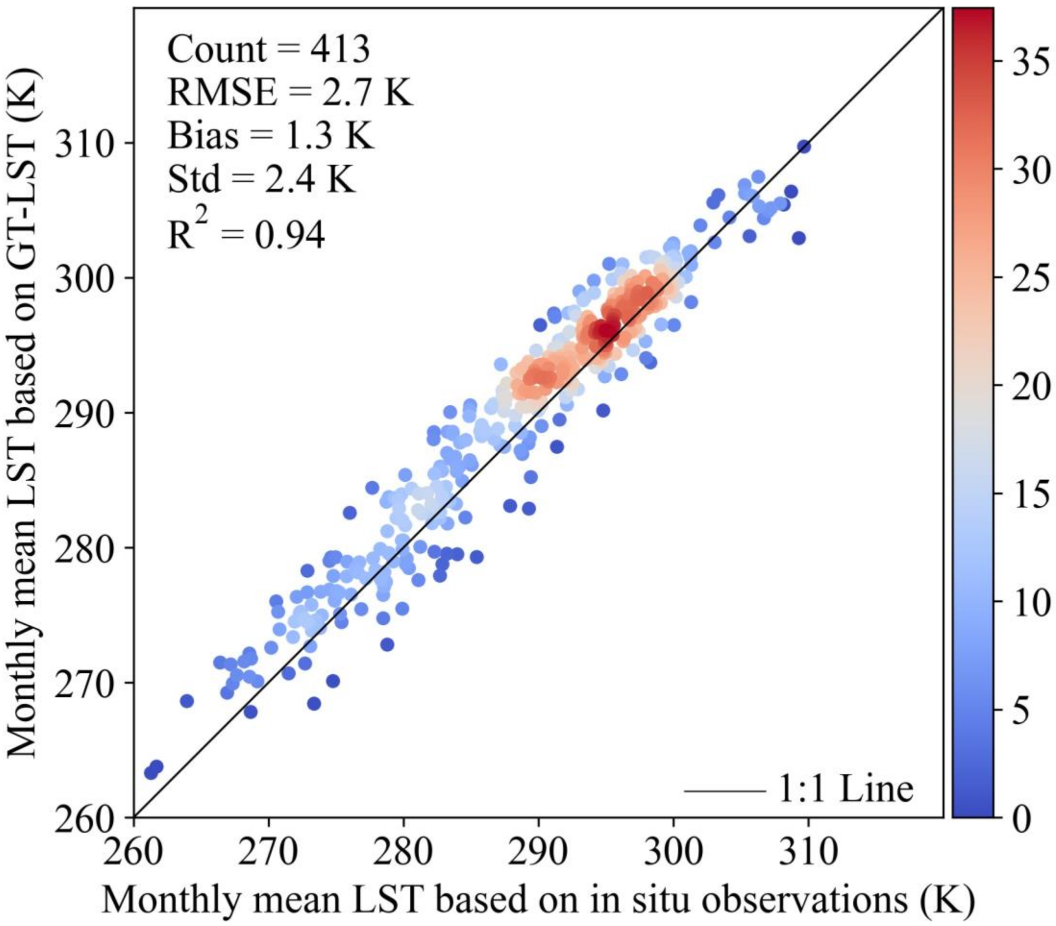


Figure 13: Monthly mean LST based on GT-LST versus monthly mean LST based on in situ LST from 1994 to 2005.

In addition, as for some details of the simple linear regression method, we have added the following descriptions in Appendix B.

"*Impacting of the NOAA satellite orbital drift, daytime and nighttime observations of NOAA afternoon satellites cannot represent maximum and minimum temperatures well. Therefore, the MMLST according to the simple average method has a significantly lower accuracy than other studies (Fig. A4). Xing et al. (2021) proposed to use 9 combinations of two to four MODIS instantaneous retrievals of which at least one daytime LST and one nighttime LST to estimate mean LSTs, and determined the weight for every moment. Inspired by the work of Xing et al. (2021), we determined to use simple linear combinations of monthly mean daytime and nighttime LST values that* *were observed at observation times for NOAA to estimate MMLST with ground-based measurement. For the combinations of two valid monthly mean LSTs (one daytime and one nighttime LST), the regression models can be written as follows*:

$$MMLST = a_1 * MMLST_{day} + a_2 * MMLST_{night} + b \qquad \text{(B1)}$$

where $MMLST$ *is the ground-based monthly mean LST,* a_1, a_2 *and* b *are the fitting coefficients,* $MMLST_{day}$ *is the monthly mean in situ LST at the NOAA daytime observation,* $MMLST_{night}$ *is the monthly mean in situ LST at the NOAA nighttime observation.*

Taking into account the observed times of NOAA satellites with orbital drift effect since 1981, combinations of two observations from these satellites contain eight cases: 13:30–17:00/01:30–05:00 local solar time in 0.5-hour interval. Based on the in situ LST measurements during the period 2003 to 2018 at 227 flux stations operating in globally diverse regions, we obtained the fitting coefficients (Table A1). Then, we calculated the MMLST of GT-LST using GT-LST monthly mean daytime and nighttime LSTs, Eq. (B1), and the fitting coefficients listed in Table A1."

Table A1. Statistics for the relationship between the regressions of the eight combinations and actual monthly mean LST.

Case	Time	a_1	a_2	b	RMSE	R^2	Number
1	13:30/01:30	0.3844	0.5783	10.3446	2.0	0.97	12095
2	14:00/02:00	0.4010	0.5621	10.2042	1.9	0.98	12241
3	14:30/02:30	0.4235	0.5451	8.6172	1.9	0.98	12381
4	15:00/03:00	0.4490	0.5211	8.2652	1.8	0.98	12303
5	15:30/03:30	0.4816	0.4840	9.5710	1.8	0.98	12165
6	16:00/04:00	0.5250	0.4349	11.2284	2.0	0.97	11818
7	16:30/04:30	0.5663	0.3884	12.8572	2.2	0.96	10992
8	17:00/05:00	0.6040	0.3621	9.7302	2.4	0.96	9765

Some minor Comments:

1. Line 125: Is there a writing mistake on this sentence? "we used 54 land surface emissivity spectra to represent different land surface types, including 41 soil types, four vegetation types, four water body 125 types and five ice/snow types were selected."

Response: We appreciate your careful reading, the phrase "were selected" was removed.

2. Line 165: "The instrumental error of the SURFRAD station give rise to uncertainty in the retrieved LST value of less than 1 K". Should be "gives rise to".

Response: Corrected as suggested.

3. Line 266 to 268: "Therefore, to obtain relatively accurate emissivity values, we developed an improved method that consider annual changes in land cover from the GLASS-GLC dataset and combines ASTER GED data with the NDVI threshold method to estimate the emissivity" The verb forms need to be unified.

Response: Thank you for your careful reading. Following your suggestion, we have checked the whole manuscript and corrected this issue.

4. Line 313 to 315: This sentence seems redundant, please write it in a more explicit way.

Response: Thank you for the suggestion. We have rewritten this part according to your suggestion.:

"In contrast to the ground-based validation and satellite product inter-comparison mentioned above, the comparisons for AVHRR LST products were performed using different strategies. Concretely, GT-LST compared with GD-LST using a strategy that compares GT-LST and GD-LST with same SURFRAD measurements concurrently with the satellite overpass, to evaluate the difference in the absolute accuracy of these two products."

5. You can use either RMSE or RMSD, but keep consistency throughout the paper and all figures.

Response: Corrected as suggested.

6. Please unify the format of all references.

Response: Thank you for the suggestion. Following your suggestion, we have checked the whole manuscript and corrected this issue.

References for the above responses are listed below:

Chen, X., Su, Z., Ma, Y., Cleverly, J., Liddell, M.: An accurate estimate of monthly mean land surface temperatures from MODIS clear-sky retrievals, J. Hydrometeorol., 18, 2827-2847, https://doi.org/10.1175/JHM-D-17-0009.1,2017.

Li, Z.-L., Wu, H., Duan, S.-B., Zhao, W., Ren, H., Liu, X., Leng, P., Tang R., Ye, X., Zhu, J., Sun, Y., Si, M., Liu, M., Li, J., Zhang, X., Shang, G., Tang, B.-H., Yan, G., and Zhou, C.: Satellite remote sensing of global land surface temperature: Definition, methods, products, and applications, Rev. Geophys., 61, e2022RG000777, https://doi.org/10.1029/2022RG000777, 2023.

Liu, X., Li, Z.-L., Li, J.-H., Leng, P., Liu, M., and Gao. M.: Temporal upscaling of MODIS 1-km instantaneous land surface temperature to monthly mean value: Method evaluation and product generation, IEEE Trans. Geosci. Remote Sens., https://doi.org/10.1109/TGRS.2023.3247428, 2023.

Xing, Z., Li, Z. L., Duan, S. B., Liu, X., Zheng, X., Leng, P., Gao, M., Zhang, X., Shang, G.: Estimation of daily mean land surface temperature at global scale using pairs of daytime and nighttime MODIS instantaneous observations, ISPRS J. Photogramm., 178, 51-67, https://doi.org/10.1016/j.isprsjprs.2021.05.017, 2021.