Response to Reviewer #2

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.

In the study titled "A global historical twice-daily (daytime and nighttime) land surface temperature dataset produced by AVHRR observations from 1981 to 2005", the authors produce a global LST product from 1981 to 2005 at 0.05 degree using AVHRR observations. The study is potentially useful for understanding changes in surface climate over a longer time period than what we can currently examine using most existing LST products. However, I have several concerns that should be addressed before the paper is considered for publication.

Q1. The biggest issue I have is that the dataset is restricted to 2005. Given that AVHRR products have large biases compared to MODIS Aqua and use different inputs (such as the dynamic emissivity estimates used), one cannot combine MODIS and AVHRR to perform long-term analysis. Since the AVHRR is still operational, the dataset needs to be extended to more recent years.

Response: Thanks for your valuable suggestion. As emphasized in the introduction section, this study aims to fill the data gap of global satellite-derived twice-daily LST before 2000. However, considering global meteorology and climatology-related applications urgently need more than 30 years of daily LST products, there are two ways of satisfying that requirement based on GT-LST. One way is to combine GT-LST (1981-2000) with the existing satellite-derived daily LST product (2000-present), which depend on different products with the same observation period to eliminate or limit the bias between different sensors. Therefore, we extend the time span of GT-LST to 2005. Benefiting from the same observation period (i.e., 2000-2005) with MODIS LST, we will produce a global long-term (1981-present) LST data record according to the method of Liu et al. (2012), which will be primarily from the AVHRR (1981-2000) and MODIS (2000-present).

Indeed, as you mentioned, extending the time span of GT-LST to present is another way to address this issue. We have already started working on generating GT-LST products (2006-present). Although we have proposed a framework for generating GT-LST product, we still need spend a lot of time downloading global AVHRR GAC L1B data, handling large amounts of original Level-1b data, generating huge amounts of process variable data, and so on. After all data have been processed, we will upload GT-LST (2006-present) to previous URL (https://doi.org/10.5281/zenodo.7134158).

Q2. As an addendum to the previous point, since one of the most important use cases of long-term datasets is time series analysis, the long-term changes in GT-LST should be compared against equivalent changes from MODIS products. If the orbital drift has a significant impact on long-term trends, we should be very cautious about the suitability of this data product for this use case. This issue needs to be quantified more clearly instead of just discussed in text in one section. This can potentially avoid misleading results from future uses of this dataset.

Response: Thanks a lot for your comments. Indeed, one of the intentions of GT-LST is providing effective supplementary data for global long-term time series analysis. The analysis 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). It is possible to derive an estimate of the global accurate DMLST, MMLST and AMLST based on twice-daily LST product. However, 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 (Figure 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. 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. 13 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.

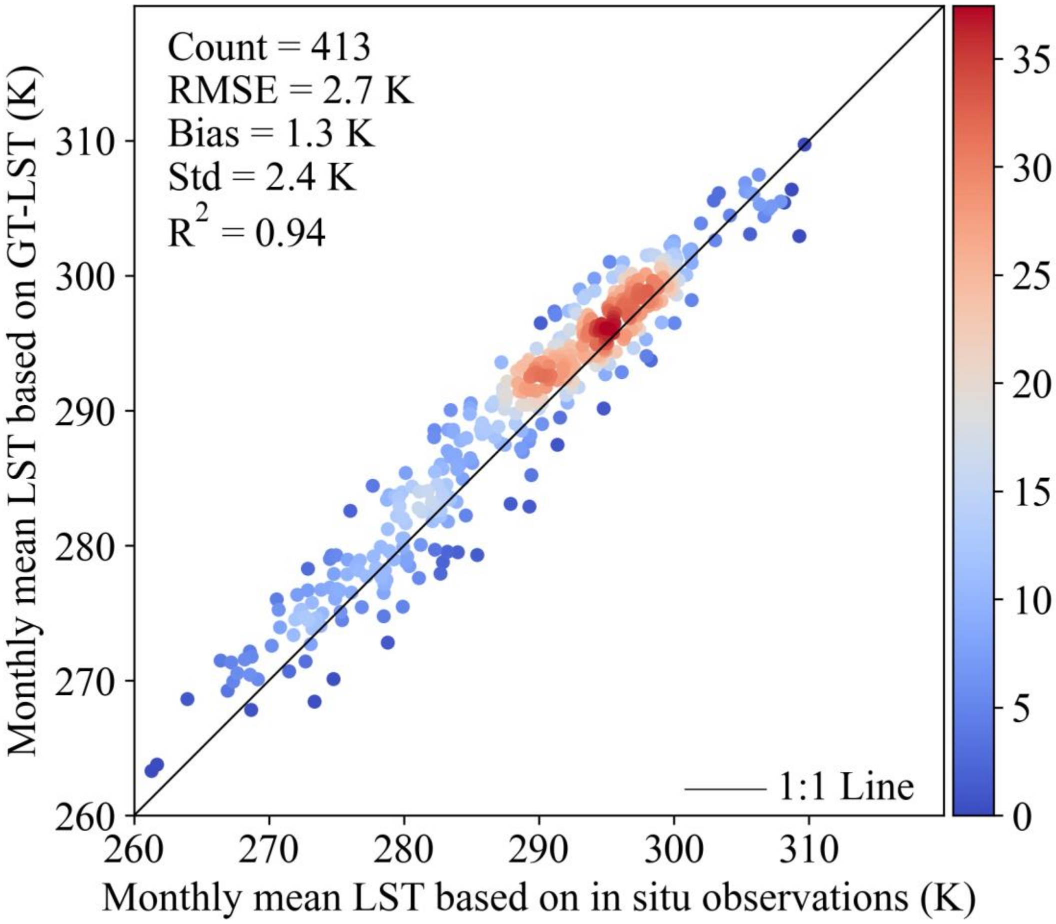

Figure 13: Monthly mean LST based on GT-LST versus monthly mean LST based on in situ LST from 1994 to 2005.

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*:

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

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

Q3. A second major source of concern is the dynamic emissivity method used. There are several vegetation-adjusted emissivity methods available, which can give different values, different enough to account for some of the biases seen. Of note, at 0.05 degree, you would start resolving larger urban areas, which is a major use case for satellite-derived LST (Voogt & Oke, 2003). Different emissivity methods perform differently over urban surfaces, which impacts this important use case (Chakraborty et al. 2021). Ideally, this issue needs to be tested further using different emissivity methods.

Response: This is a good suggestion! As you mentioned, to date, various land surface emissivity (LSE) estimation methods have been proposed with the same goal but different advantages, and limitations, e.g., classification-based method, NDVI threshold method, TES method, and physics-based day/night method and so on (Li et al., 2013). With their advantages and limitations, these methods have different accuracies and are applicable for various sensors and applications. To reflect the performance for the emissivity-retrieved methods and account for the positive bias between GT-LST and MYD11A1 LST, according to your suggestion, we compared GT-LST with MYD21A1 LST that uses the same observations with MYD11A1 but uses a physics-based algorithm to dynamically retrieve both LST and spectral emissivity. The intercomparison results of MYD21A1 LST showed a very lower bias. As for the comparison with MYD21A1, a brief explanation was analyzed below in Q4. There is no denying that it is important and significative to evaluate emissivity methods under different circumstances and for various applications. However, the goal of this study is to develop a global historical twice-daily LST product from 1981 to 2005, where LSE is only one of the key parameters. Therefore, we choose an improved NDVI threshold method to estimate LSEs from space for a global case by taking the sensor characteristics, the required accuracy, as well as computation time into account. Although evaluating emissivity methods under different circumstances is not be discussed in more detail in this study, this is a meaningful research topic. Inspired by your suggestion and the work of Chakraborty et al. (2021), we will evaluate these methods on an identical standard and to give the quality and accuracy on their applications in future work. After evaluating them we may attempt to generate the first estimates of LST at a global scale using AVHRR GAC data by combining all these approaches.

Q4. The comparison with MODIS MYD11 is somewhat difficult because of the different emissivity method used. The comparison should be done against MODIS MYD21, which uses the same observations, but a temperature-emissivity separation method instead of classification-based prescribed emissivity.

Response: Thanks for your valuable comments. According to your comments, we have compared GT-LST with MYD21A1 LST (Aqua/MODIS LST product using the TES algorithm, Collection 6.1). Spatially, this intercomparison was conducted at the global scale. Temporally, it was performed on 4 months in 2004 (January, April, July, and October) which cover different seasons. The results of the daytime and nighttime comparison, in Fig. 9, are as follows: The daytime and nighttime RMSD values of 3.2 K and 2.5 K and that of bias of 0.1 K and 1.3 K. Compared to the result of MYD11A1, the significantly smaller bias was obtained for MYD21A1. The possible reason is attributed to the fact that the MYD21A1 LST uses the same observations with MYD11A1 but uses TES method to dynamically retrieve LSE. The following contents have been added in Line 184-187 and Line 405-412, respectively.

"*In this study, Collection-6.1 MYD11A1 of 2004 was selected for sensor-to-sensor comparison. MYD21A1 LST product, which uses the same observations with MYD11A1 but uses temperature–emissivity separation method to dynamically retrieve LST and emissivity, was also selected to make an intercomparison with GT-LST in this study. This inter-comparison was conducted on 4 months in 2004 (January, April, July, and October) which cover different seasons.*"

"*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). To further demonstrate this point, we compared GT-LST with MYD21A1 LST. Fig. 9 shows the daytime and nighttime RMSD values of 3.2 K and 2.5 K and that of bias of 0.1 K and 1.3 K between GT-LST and MYD21A1 LST for 4 months in 2004. Compared to the result of MYD11A1, the significantly smaller bias was obtained for MYD21A1. The possible reason is attributed to the fact that the MYD21A1 LST uses the same observations with MYD11A1 but uses a physics-based method to dynamically retrieve emissivity.*"

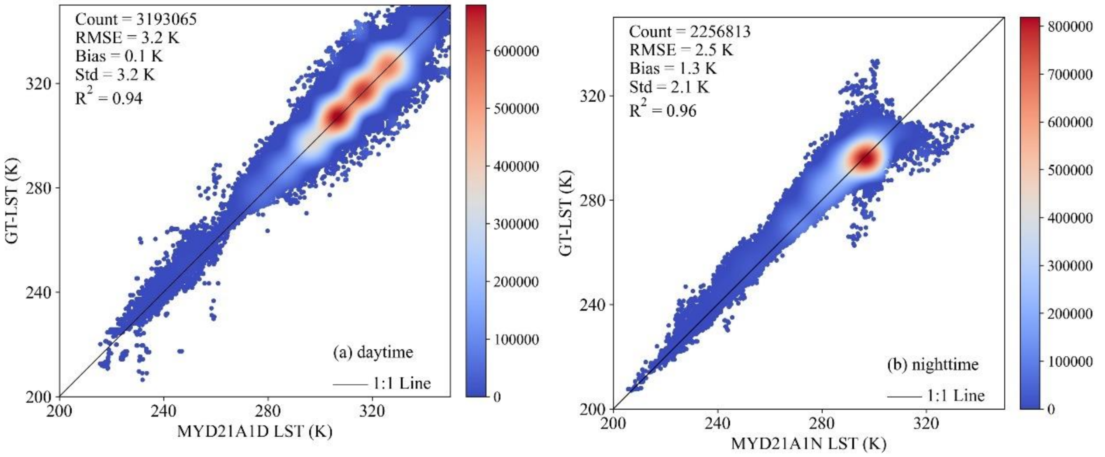

Figure 9: Intercomparison of GT-LST and MYD21A1 LST in January, April, July, and October 2004: (a) daytime; (b) nighttime.

Q5. For comparison with SURFAD stations, did the authors check that the emissivity used to generate the LST in the ground observations is same as the LST in the GT-LST product? If they are different, would be good to adjust by the emissivity difference and check if that improves the accuracy.

Response: The GT-LST use directional measurements of AVHRR in the atmospheric window, while the SURFRAD stations provide upwelling and downwelling broadband hemispherical TIR radiances using pyrgeometers in the spectral range from 3.5 to 50 µm, from which estimates of LSTs can be derived using Stefan–Boltzmann's law. To retrieve LST using Stefan–Boltzmann's law, the surface broadband emissivity must be known a priori. In this study, these broadband emissivities were estimated from ASTER emissivity product using a spectral-to-broadband linear regression equation according to the work of Duan et al. (2019), as follows: BND(0.968), TBL(0.972), DRA(0.967), FPK(0.973), GCM(0.971), PSU(0.970), and SXF(0.970). According to the study of Liang (2005), the surface broadband emissivity of sites can be obtained from AVHRR LSE in AVHRR LSE for channel centered at 11 and 12 µm via the empirical relationship:

$$\varepsilon = 0.2489 + 0.2386\varepsilon_{11} + 0.4998\varepsilon_{12} \qquad \text{(R1)}$$

According to Eq. (R1), the surface broadband emissivities are 0.976, 0.975, 0.972, 0.973, 0.973, 0.968 and 0.974 for BND, TBL, DRA, FPK, GCM, PSU, and SXF, respectively. Different empirical relationships perform an error less than 0.01 in the broadband emissivity. According to the study of Xing et al. (2021), the emissivity changes by 0.01, and the change in in-situ LST will not exceed 0.37 K. Therefore, while this error is not negligible, it does not appear to be a dominant source of uncertainty in the ground-based validation.

Q6. Finally, given the view angle of AVHRR, a broader discussion needs to be added about thermal anisotropy (DUffour et al. 2015). Satellites only provide a 2d directional view of LST, and this is not directly comparable across satellites (Landsat vs MODIS) or against ground observations that have a downward pointing radiometer. This is of particular concern over heterogeneous terrain, such as mixed forests and over cities.

Response: Thanks a lot for your valuable comments. we would like to make some explanations on the thermal anisotropy issue as follows:

(1) Previous multi-sensor comparison studies (Guillevic et al., 2012; Trigo et al., 2008) found differences up to 12 K between MODIS and SEVIRI LST due to directional effects. Appropriate matchups significantly reduce the discrepancies induced by directional effects (Guillevic et al., 2014). In this study, to avoid the uncertainties induced by directional effects, a strict criterion of viewing geometry alignment was established to guarantee the reliability of the intercomparison results (Li et al., 2023): the difference in VZA between MYD11A1 and GT-LST is limited to be less than 15°.

(2) LST from satellite and ground measurements may differ according to their measurement methods. AVHRR use directional measurements in the atmospheric window, while ground-based longwave radiation measurements are hemispheric, wider spectrum derivations. If the surface is black body, the two LSTs are the same (Wang et al., 2005, Li et al., 2023). However, most natural objects are not black bodies. AVHRR view zenith angles were considered to be an important factor influencing the results when comparisons were made with in situ measurements. Figure R1 shows the relationship between view angles and bias in instantaneous LST at station pixels. Our result is that a high view angle does not necessarily show a high bias and a low view

angle does not necessarily always show a low bias. This means that view angle should not be a significant source for the bias of GT-LST at a 0.05° pixel size.

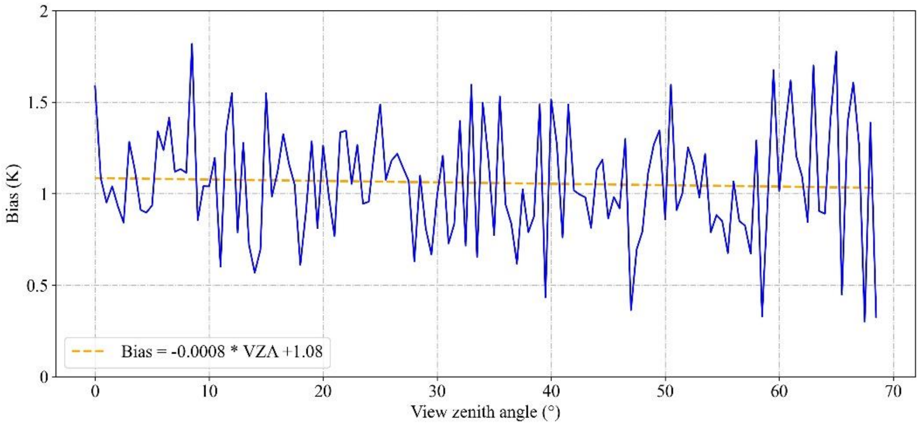

Figure R1: The relationship between bias of land surface temperature (GT-LST minus in situ observations) and view zenith angle.

References for the above responses are listed below:

Chakraborty, T. C., Lee, X., Ermida, S., and Zhan, W.: On the land emissivity assumption and Landsat-derived surface urban heat islands: A global analysis, Remote Sens. Environ., 265, 112682, https://doi.org/10.1016/j.rse.2021.112682, 2021.

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.

Duan, S.-B., Li, Z.-L., Li, H., Göttsche, F.-M., Wu, H., Zhao, W., Leng, P., Zhang, X., Coll, C.: Validation of Collection 6 MODIS land surface temperature product using in situ measurements, Remote Sens. Environ. 225, 16–29, https://doi.org/10.1016/j.rse.2019.02.020, 2019.

Guillevic, P. C., Biard, J. C., Hulley, G. C., Privette, J. L., Hook, S. J., Olioso, A., Göttsche F. M., Radocinski, R., Román, M. O., Yu, Y., and Csiszar, I.: Validation of

Land Surface Temperature products derived from the Visible Infrared Imaging Radiometer Suite (VIIRS) using ground-based and heritage satellite measurements, Remote Sens. Environ., 154, 19-37, https://doi.org/10.1016/j.rse.2014.08.013, 2014.

Guillevic, P. C., Privette, J. L., Coudert, B., Palecki, M. A., Demarty, J., Ottlé, C., and Augustine, J. A.: Land Surface Temperature product validation using NOAA's surface climate observation networks—Scaling methodology for the Visible Infrared Imager Radiometer Suite (VIIRS), Remote Sens. Environ., 124, 282-298, https://doi.org/10.1016/j.rse.2012.05.004, 2012.

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.

Li, Z.-L., Wu, H., Wang, N., Qiu, S., Sobrino, J. A., Wan, Z., Tang, B.-H., and Yan, G.: Land surface emissivity retrieval from satellite data, Int. J. Remote Sens., 34, 3084-3127, https://doi.org/10.1080/01431161.2012.716540, 2013.

Liang, S.: Estimation of Surface Radiation Budget: I. Broadband Albedo, in Quantitative Remote Sensing of Land Surfaces, John Wiley & Sons, Inc., Hoboken, NJ, USA, 310–344, 2005.

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.

Liu, Y., Liu, R., and Chen, J. M. (2012). Retrospective retrieval of long-term consistent global leaf area index (1981–2011) from combined AVHRR and MODIS data, J. Geophys. Res-Biogeo., 117(G4), 1-14, https://doi.org/10.1029/2012JG002084, 2012.

Trigo, I. F., Monteiro, I. T., Olesen, F., and Kabsch, E.: An assessment of remotely sensed land surface temperature. J. Geophys. Res-Atmos., 113, D17, https://doi.org/10.1029/2008JD010035, 2008.

Wang, K., Wan, Z., Wang, P., Sparrow, M., Liu, J., Zhou, X., and Haginoya, S.: Estimation of surface long wave radiation and broadband emissivity using Moderate Resolution Imaging Spectroradiometer (MODIS) land surface temperature/emissivity products, J. Geophys. Res-Atmos., 110, D11, https://doi.org/10.1029/2004JD005566, 2005.

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.