

Response to Reviewer #1

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 manuscript proposes a long-term (1981-2005) AVHRR land surface temperature (LST) dataset that includes outcomes at both daytime and nighttime. The algorithm is the generalized split-window (GSW) algorithm while in the production, this dataset also considered annual land cover change. Overall, the accuracy of the proposed dataset is promising, and it filled the gaps regarding long-term global LST datasets, especially at nighttime. Therefore, I would recommend it be published on ESSD after a major revision.

Q1. Positive bias issue. Based on site validation and inter-comparison with MYD11 and the other two AVHRR LST products, the proposed GT-LST shows a clear positive bias (>1 K) nearly in all results. The authors claim the bias is due to the emissivity difference (Line 370), however, the proposed GT-LST has a clear bias than the other three products, and it seems that the emissivity used by GT-LST is not accurate. The authors mention that the dataset will be calibrated to remove the bias in the future (Line 436). I am thinking if it would be better to solve this issue in this paper as it doesn't need to be done in a separate paper.

Response: Thanks a lot for your valuable comments. First, we would like to make some explanations on the positive bias issue as follows:

(1) The GT-LST product and the global daytime AVHRR LST (GD-LST) used a dynamic emissivity method to retrieve LST. We compared GT-LST with GD-LST on January 15, April 15, July 15, and October 16, 1999, with low positive bias of 0.6 K (Figure R1).

(2) The MYD11A1 LST product and the regional twice-daily LST product over

Africa (RT-LST) are generated by the split-window (SW) algorithm. Land surface emissivities of these two products are assigned according to classification-based method that produces emissivities with fixed values for a limited number of land cover types. This method works well over densely vegetated areas and water where emissivities are relatively stable. However, cold biases of 3-5 K are often found over semi-arid and arid regions because these regions have much higher emissivity variability, and only one fixed overestimated emissivity inferred from land cover types is assigned to these regions (Coll et al., 2009; Hulley and Hook 2009; Wan et al., 2002). In order to represent the natural variation in emissivity, we used an improved NDVI threshold method to dynamically retrieve daily emissivity. Based on the analysis above, emissivity derived from dynamic methods is lower than emissivity according to classification-based method, which makes the proposed GT-LST is higher than MYD11A1 LST and RT-LST (i.e., positive bias). We note that earlier researches on this issue had similar results. Reiners et al. (2022) compared AVHRR LST product of the TIMELINE project with MYD11_L2 LST product from 2003 to 2014, the result shows that the TIMELINE dynamic emissivity is lower than the MYD11_L2 fixed emissivity and a general positive bias (i.e., bias=2.2 K) of TIMELINE LST towards MYD11_L2 LST. Martins et al. (2019) compared MSG LST and GOES-16 LST and revealed that a positive bias (MSG > GOES) of around 1.6 K persists due to the overestimation of the fixed emissivity of GOES. Mao et al. (2007) analyzed the retrieval result by radiative transfer model with neural network algorithm and MODIS product algorithm, indicating that MOD11_L2 LST product overestimates the emissivity, resulting in an underestimation of LST.

(3) To further illustrate the positive bias issue, we present an intercomparison exercise between MxD11A1 LST products (Terra and Aqua/MODIS using SW algorithm, Collection 6) with fixed emissivity and MSG LST products (MSG/SEVIRI using SW algorithm) with dynamic emissivity for 4 days (January 15, April, 15, July 15, and October 15, 2020). The criteria in Sec 3.2 were used to guarantee the reliability of the intercomparison results. The result is shown in Figure R2, indicating that a general positive bias (daytime ranges from 0.7 K to 3.3 K, nighttime ranges from 0.2 K

to 1.4 K) of MSG LST towards MxD11A1 LST for each land cover types.

(4) The comparison with in situ LST showed that a positive bias was found for all SURFRAD sites. However, only the bias of BND and FPK are large than 1 K. Similar results were obtained by Reiners et al. (2022) and Liu et al. (2019).

Therefore, we think that positive biases obtained for GT-LST and other LST products are relatively reasonable.

Next, many LST products can provide global twice-daily LST after 2000, such as ASTER LST, MODIS LST, VIIRS LST, AATSR LST and SLSTR LST. Users can obtain a relatively long-term twice-daily LST product by combining GT-LST with these LST products. However, integration of LST from different sensors is complicated. Due to the different LST inversion methods, air conditions, viewing geometries, etc., the sensors bias between GT-LST and other LST products is not constant. Therefore, developing a general method to utilize for sensor normalization is difficult and is not the key point of this paper.

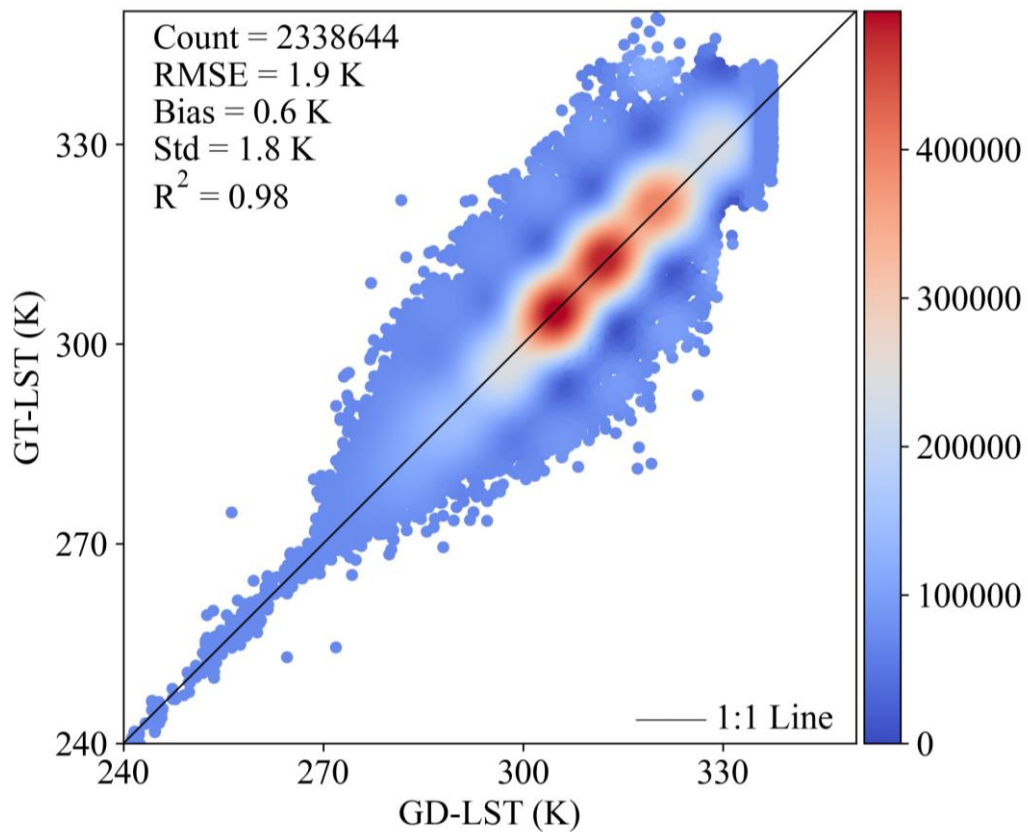
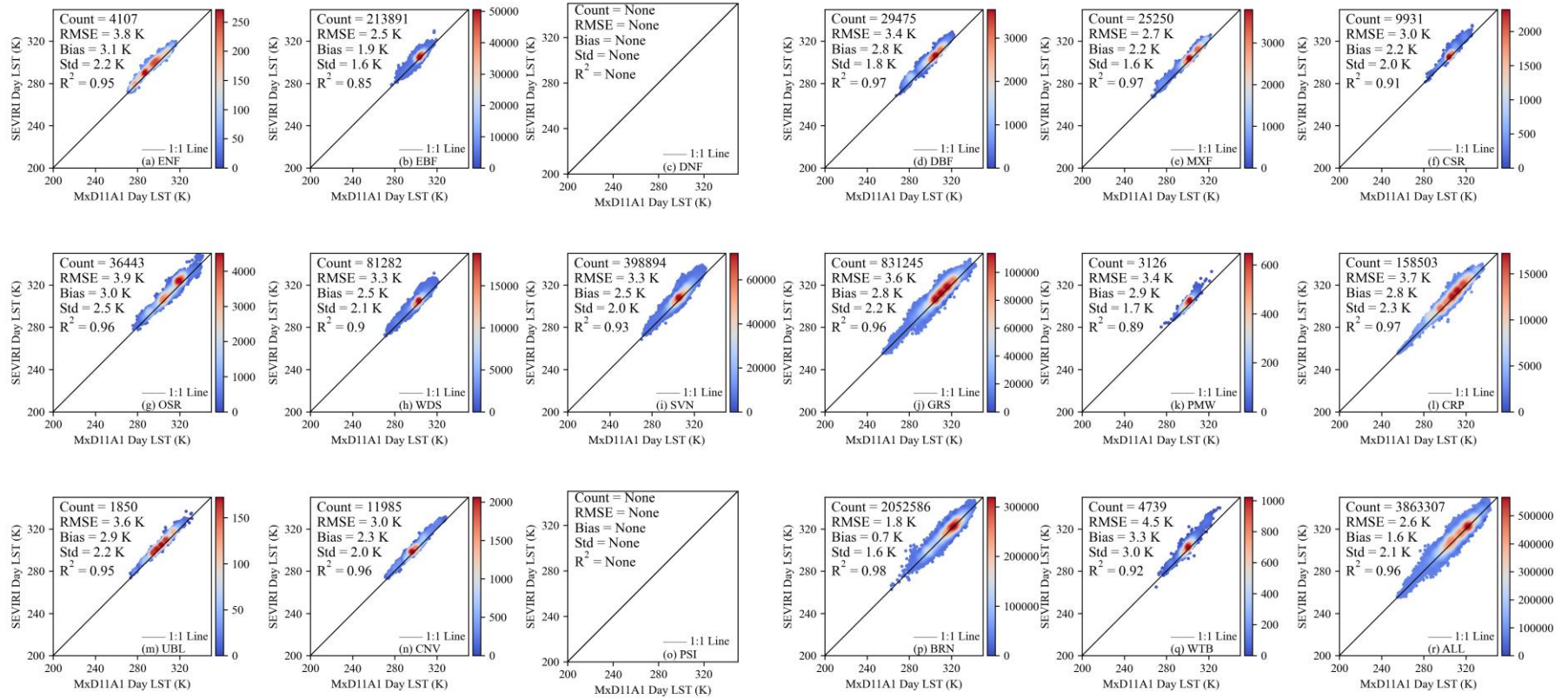


Figure R1. GT-LST versus GD-LST during the daytime on January 15, April 15, July 15, and October 16, 1999.

a



b

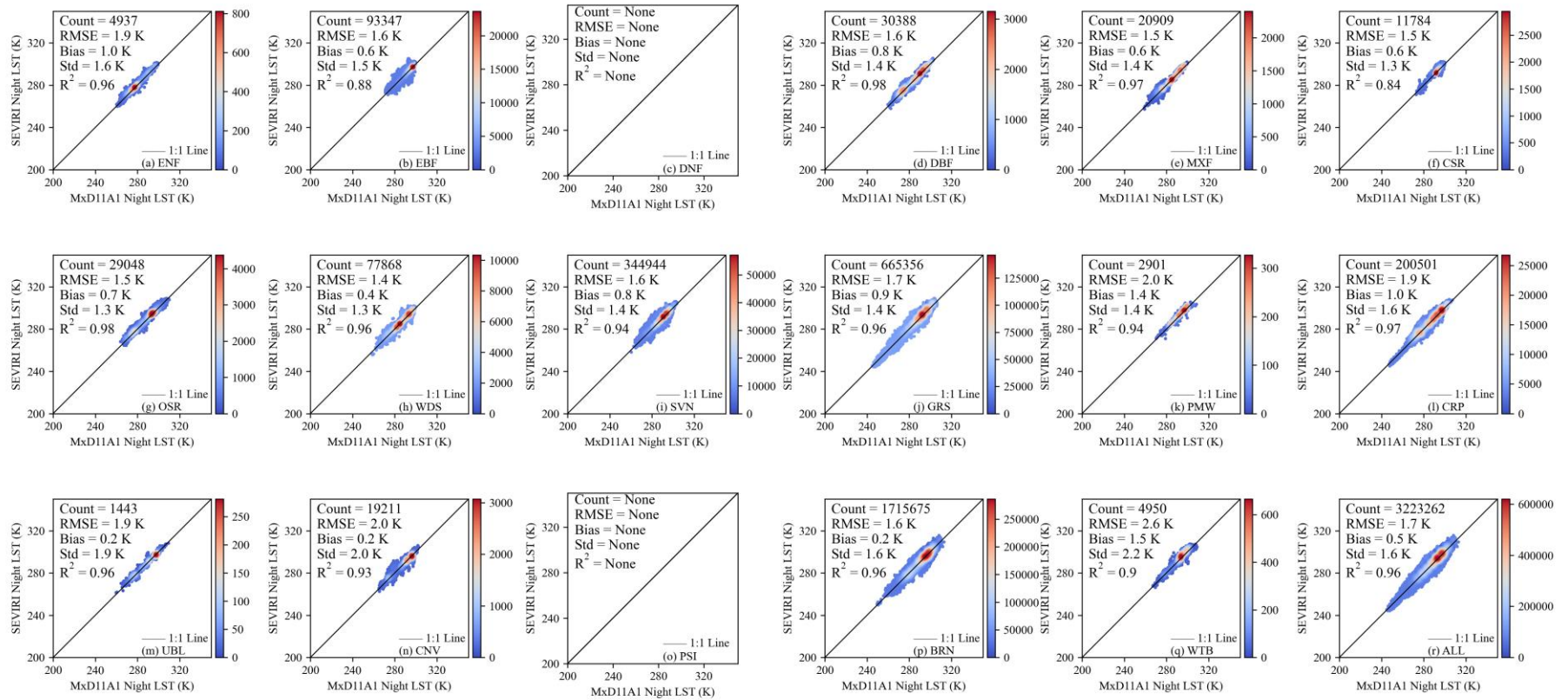


Figure R2. The bias (SEVIRI LST minus MxD11A1 LST) and RMSD between the MxD11A1 product and SEVIRI during daytime (a) and nighttime (b) over various land cover types.

Q2. Large RMSE (4.1 K) of the monthly mean LST result. The GT-LST is claimed to have the strength to generate gap-free monthly mean LST; however, the outcome has an RMSE of 4.1 K which is too large at a monthly scale compared to other studies (Line 395). This part weakened the statement of the advantage of GT-LST for temporal upscaling based on the logic chain. I would suggest either removing this part or quantifying the impact of orbit drift, in other words, comparing the accuracies of samples that have not and have suffered from orbit drift, and then claiming the potential of this data after orbit drift.

Response: Thanks for your suggestion. We used simple linear combinations of monthly mean daytime and nighttime LST values to estimate MMLST. The detailed revisions are listed 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 **1995–2005**. 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. 15 according to the simple linear regression method. For your convenience, we listed it below.

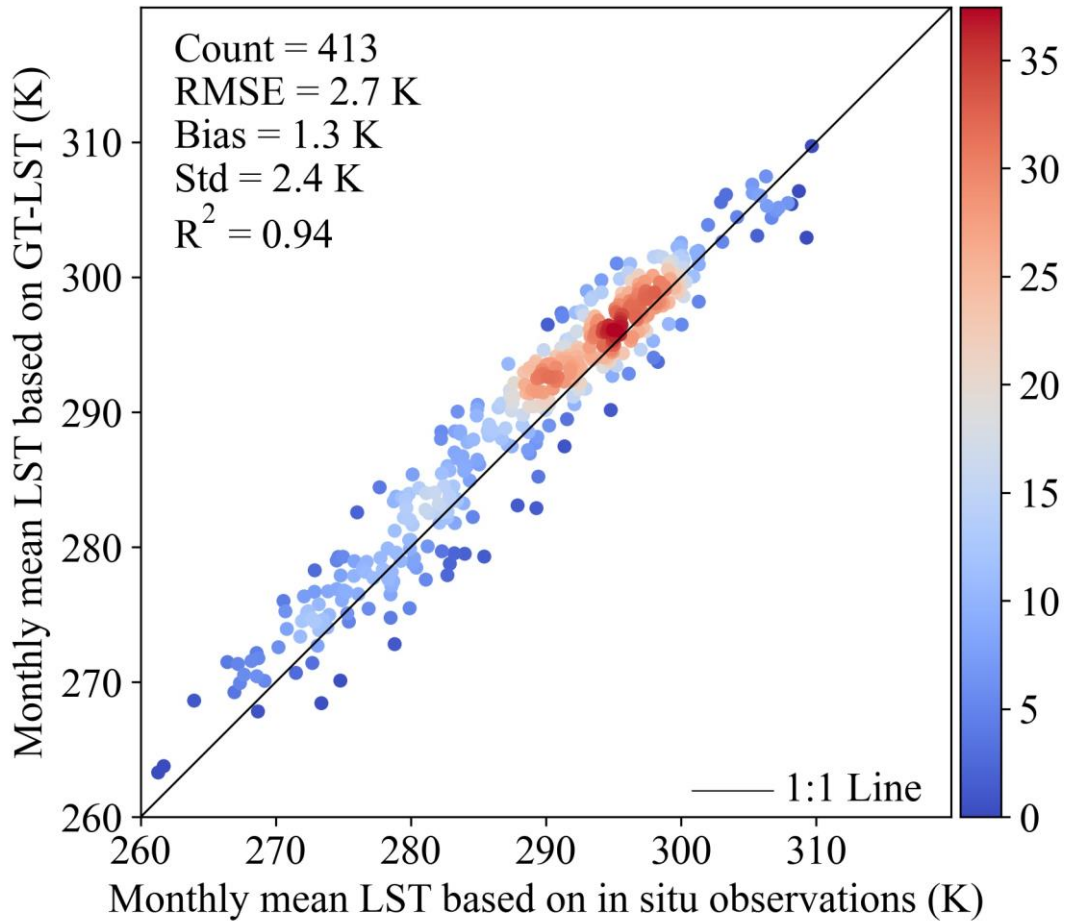


Figure 15. Monthly mean LST based on GT-LST versus monthly mean LST based on in situ LST from 1995 to 2005.

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 (Figure A3). 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:

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

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. The impact of annual land cover change. This is an interesting part of the study, whereas the study didn’t pay attention to the performance of such change. Traditionally people mainly utilized a land cover climatology map rather than annual changes to retrieve global LST. I would suggest including additional analysis to find some examples and compare with LST from Ma et al. (2020) to demonstrate the progress using annual land cover maps.

Response: This is a good suggestion! Changes in land cover have been accelerating since 1980 under the impact of climate changes and human activities. As an intrinsic property of natural materials, land surface emissivity predominantly depends on the land cover type. Therefore, using only one year of land cover data to determine long-

term emissivity is not accurate. The quantitative relationship between annual land cover change and LST is rather complex because the changes of Land surface temperature were related to many factors, including changes in land cover, land surface parameters, seasonal variation, climatic condition and economic development, etc. Furthermore, GT-LST and LST from Ma et al. (2020) used different LST retrieval algorithms and data sources, which makes it harder to analyze the impact of annual land cover change between these two LST products. However, this is a meaningful research topic, and we will further analyze the impact in future work.

Q4. Some processes were not introduced clearly.

Q4.1 why does not GT-LST cover 1981 to 2022? GAC raw data is still updating.

Response: Thanks a lot for your comment. The reasons that GT-LST only cover 1981 to 2005 are as follows:

Existing satellite-based global twice-daily LST products can only date back to 2000. Therefore, when the study began, we aimed to fill the data gap of global satellite-derived twice-daily LST before 2000. Considering global meteorology and climatology-related applications urgently need more than 30 years of daily LST products, GT-LST can be combined with the existing satellite-derived daily LST product (e.g., MODIS LST, AATSR LST and ASTER LST) after 2000 to satisfy that requirement. However, integration of LST from different sensors need to eliminate or limit the bias between the sensors. We then extend the time span of GT-LST to 2005. Benefiting from the same observation period with other LST products, these extended data can be used to calibrate the bias between GT-LST and other LST datasets. In this way, users can obtain a relatively homogeneous twice-daily LST product for 1981 to 2022. However, we will apply your suggestion to extend the time span of GT-LST to 2022 in the near future.

Q4.2 why did the authors only employ the site observations from 1995 to 2000? If you can extend it to 2005, you can include one more SURFRAD site.

Response: Thank you for your suggestion. We have extended the observations of SURFRAD sites to 2005 and employed one more SURFRAD site (i.e., SXF) observations according to your suggestion. We have redrawn Fig. 8. For your convenience, we listed it below.

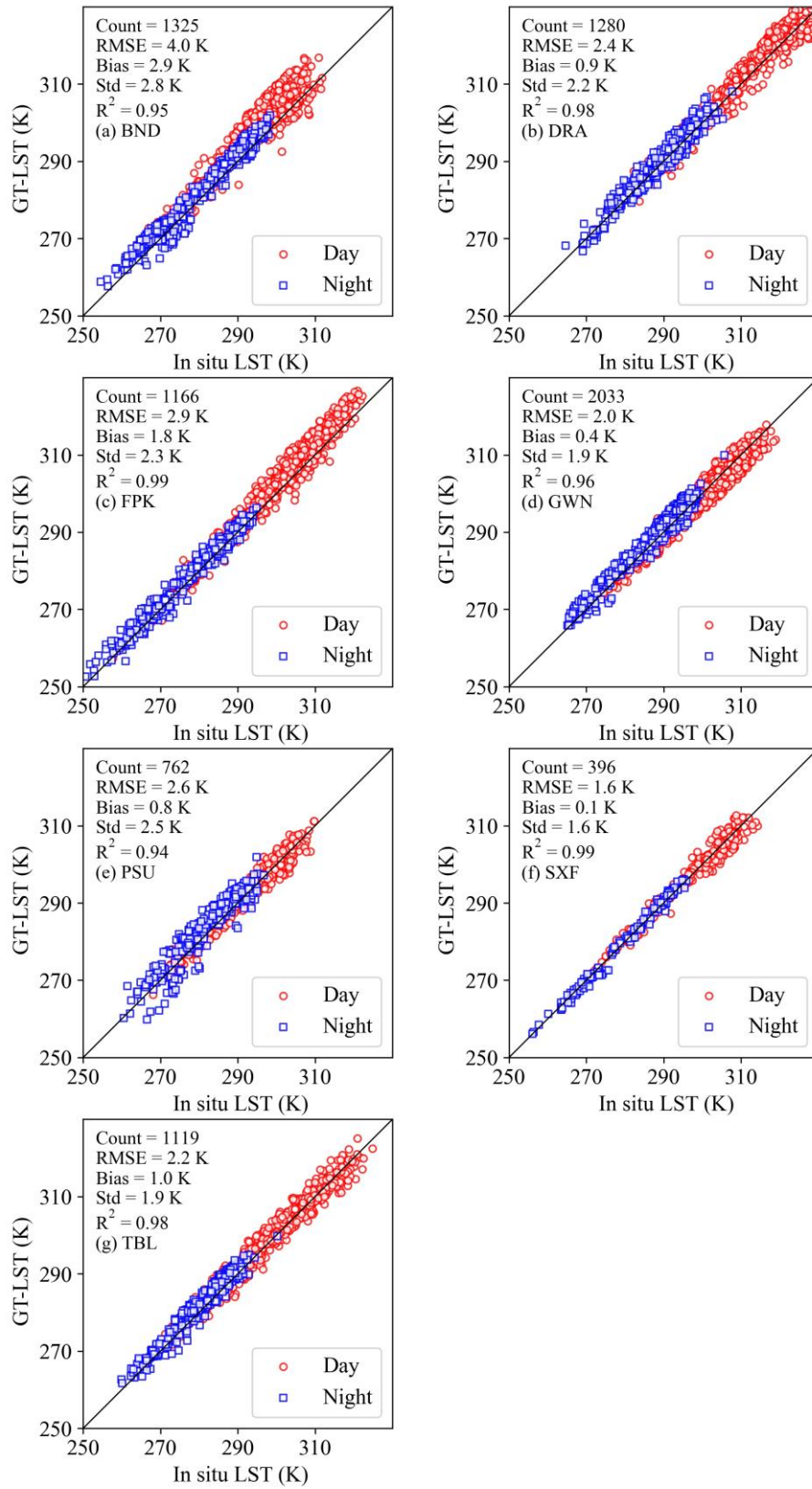


Figure 8. GT-LST versus in situ LST for 1995–2005 at (a) BND, (b) DRA, (c) FPK, (d) GWN, (e) PSU, (f) SXF, and (g) TBL sites.

Q4.3 Regarding the site validation, 6 sites seem not enough to represent the accuracy of the global product. I would recommend adding some BSRN sites that also have good data quality.

Response: Thanks for your suggestion. Following your comments, we have added some BSRN sites to represent the accuracy of the GT-LST product in contrasting climatic zones. The following contents have been added in Section 2.5 and Section 4.1, respectively.

“...The BSRN has 76 stations that detect important changes in the Earth's radiation field at the Earth's surface since 1992. These stations provide high-quality surface and upper-air meteorological observations, which are important in supporting the validation and confirmation of satellite. We selected four sites with measurements of upwelling and downwelling TIR radiances before 2000 (Table 3).”

“...We further compared GT-LST data with in situ LST data at BAR, NYA, PAY, and TAT sites for 1995–2005. Fig. 9 shows the scatterplots between GT-LST and in situ LST at these four BSRN sites. The accuracy of GT-LST product at BSRN sites is relatively worse than that at SURFRAD sites, with RMSE (bias) ranges from 3.1 K (-2.7 K) to 4.0 K (2.5 K).”

Table 3. Details of the validation sites used in this study.

	Name	Elevation(m)	Land cover type	Latitude	Longitude	Valid period
SURFRAD	BND	230	Croplands	40.0519	-88.3731	1995–2005
	DRA	1007	Open shrublands	36.6237	-116.0195	1998–2005
	FPK	634	Grasslands	48.3078	-105.1017	1994–2005
	GWN	98	Cropland/natural vegetation mosaic	34.2547	-89.8729	1994–2005
	PSU	376	Cropland/natural vegetation mosaic	40.7201	-77.9309	1998–2005
	TBL	1689	Grasslands	40.1250	-105.2368	1995–2005
	SXF	473	Croplands	43.7343	-96.6233	2003–2005
BSRN	BAR	8	Tundra	71.3230	-156.6070	1995–2005
	NYA	11	Tundra	78.9227	11.9273	1999–2005
	PAY	491	Cultivated	46.8123	6.9422	1995–2005
	TAT	25	Grass	36.0581	140.1258	1996–2005

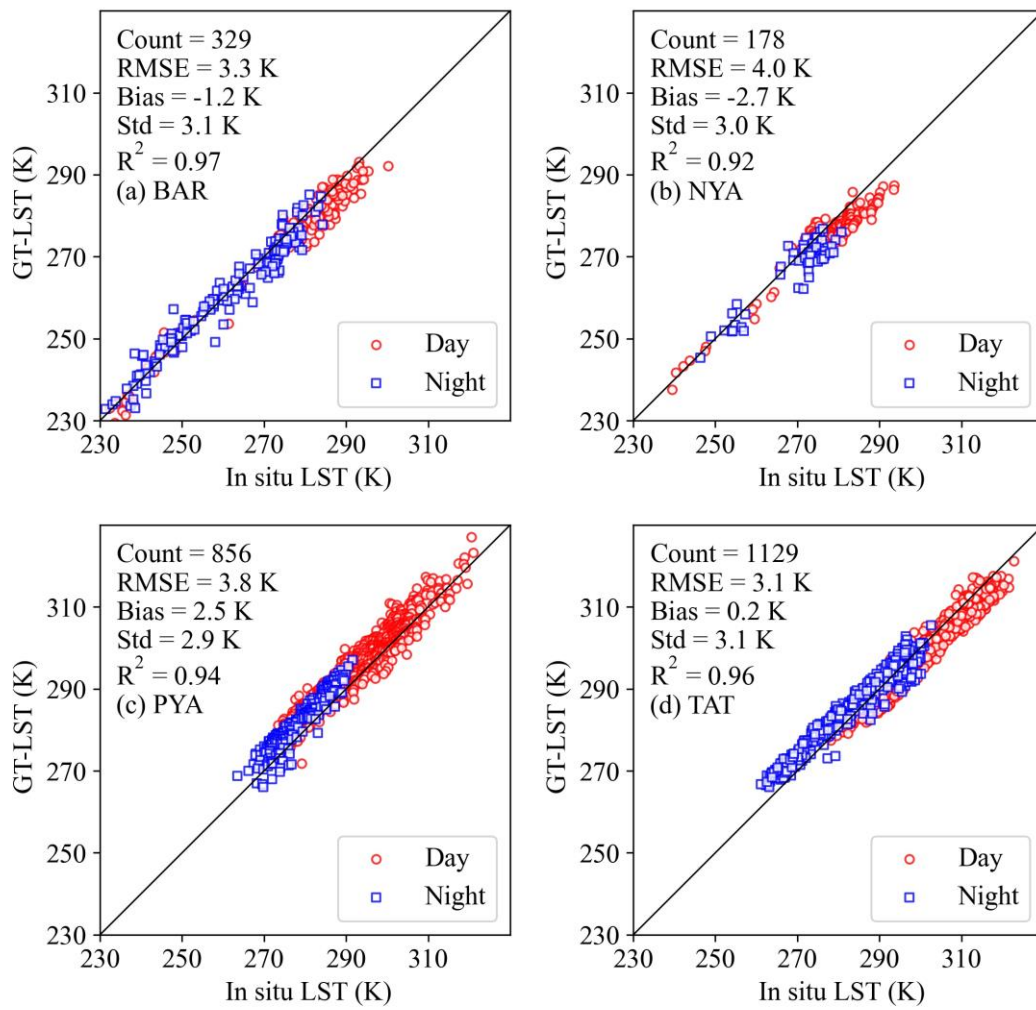


Figure 9. Scatterplots between GT-LST and in situ LST at (a) BAR, (b) NYA, (c) PYA, and (d) TAT.

Q4.4 why Fig 9(b) has some considerable scattered samples? Those cases should be discussed in the context.

Response: Thanks for your comment. We have added some discussion in Section 4.2 for the revised manuscript as follows:

“...However, as can be seen in Fig.10(b), large LST differences (GT-LST - MYD11A1 LST) more than 20 K are mostly distributed in red box. Through counting, there are 111 samples in red box, which are barren land cover type and arid climate type. Fig. A2 shows the distribution of each scattered samples in red box. 77 of 111 samples happened in Haiya, Sudan on March 31, 2004. The samples of rest happened in Taif, Saudi Arabia on April 2, 2004. For these samples, we double-checked all

variables that are essential parameters in GT-LST retrieval. The result show that all scope variables are reasonable except BT of TIR bands. Abnormal high BTs at these nighttime samples were found on March 31 and April 2, 2004 (Fig. A3), which led to extreme high LSTs. The possible reasons for abnormal high BTs are as follows: (1) These two regions may have experienced extreme events such as wars and natural disasters on March 31 or April 2, 2004. But we didn't find relevant information from historical news and documents. (2) Another factor may be instrument failure on these two days."

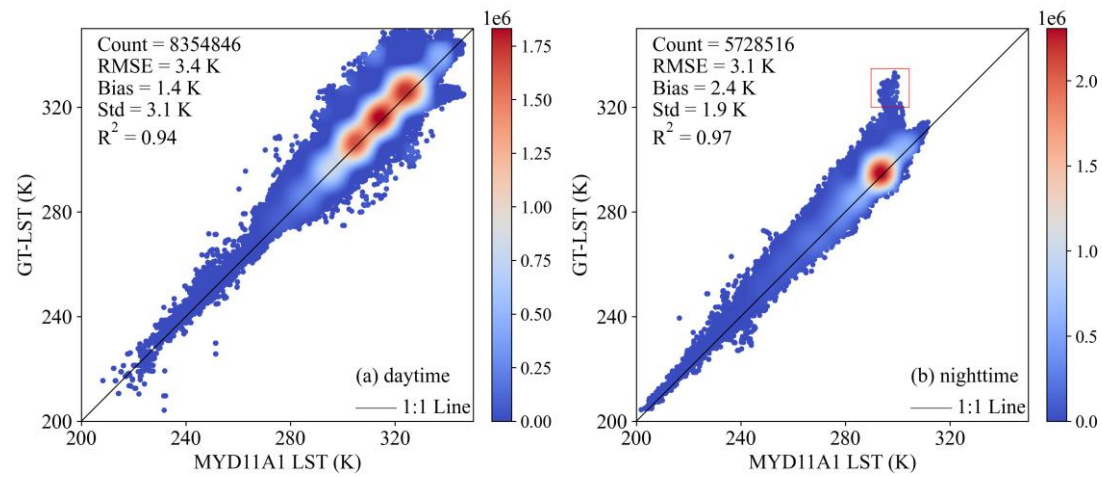
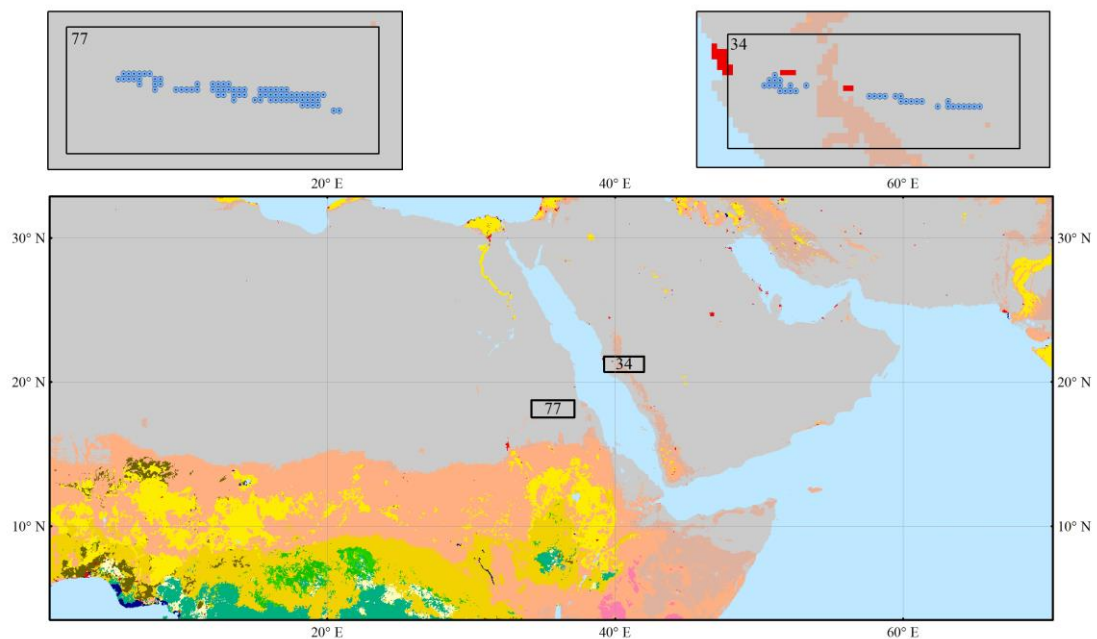


Figure 10. Inter-comparison of GT-LST and MYD11A1 LST in 2004: (a) daytime; (b) nighttime. Red box indicates considerable scattered samples.



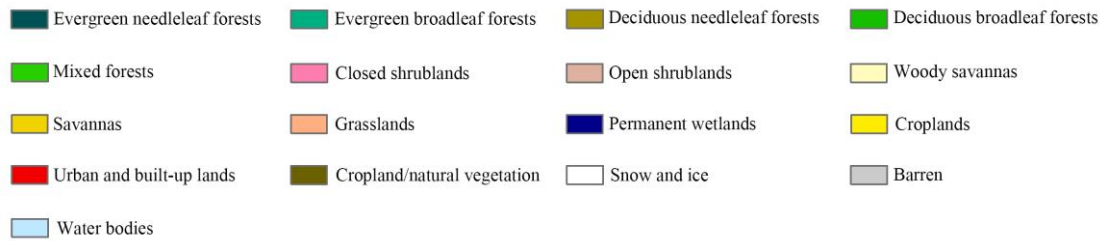


Figure A1. Distribution of the 111 scattered samples.

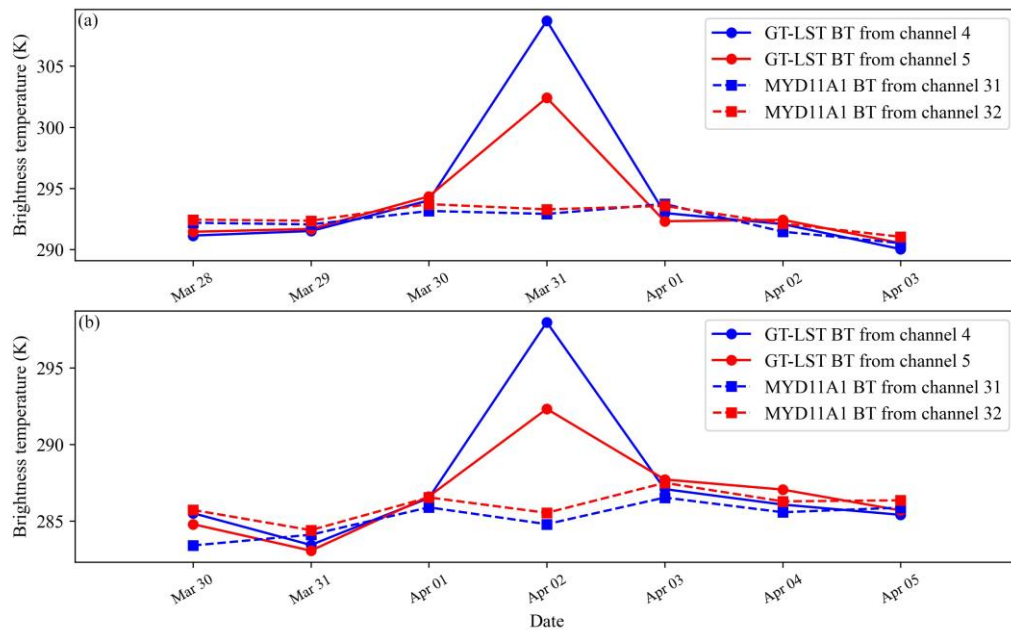


Figure A2. An example of abnormal high BTs on (a) March 31, 2004 and (b) April 2, 2004.

Q4.5 Line 350: as MODIS has been spatially aggregated to match with GT-LST, why spatial heterogeneity is still an issue here?

Response: Thanks for your comment. We have deleted this erroneous expression.

Q4.6 Fig10: I would suggest changing Fig10 to another format: consider RMSE and bias as the two dimensions of the plot, and mark each dot by their names as using color to show the bias is not easily quantified.

Response: Thanks for your valuable suggestion. We have redrawn Fig. 11 according to your suggestion. For your convenience, we listed it below.

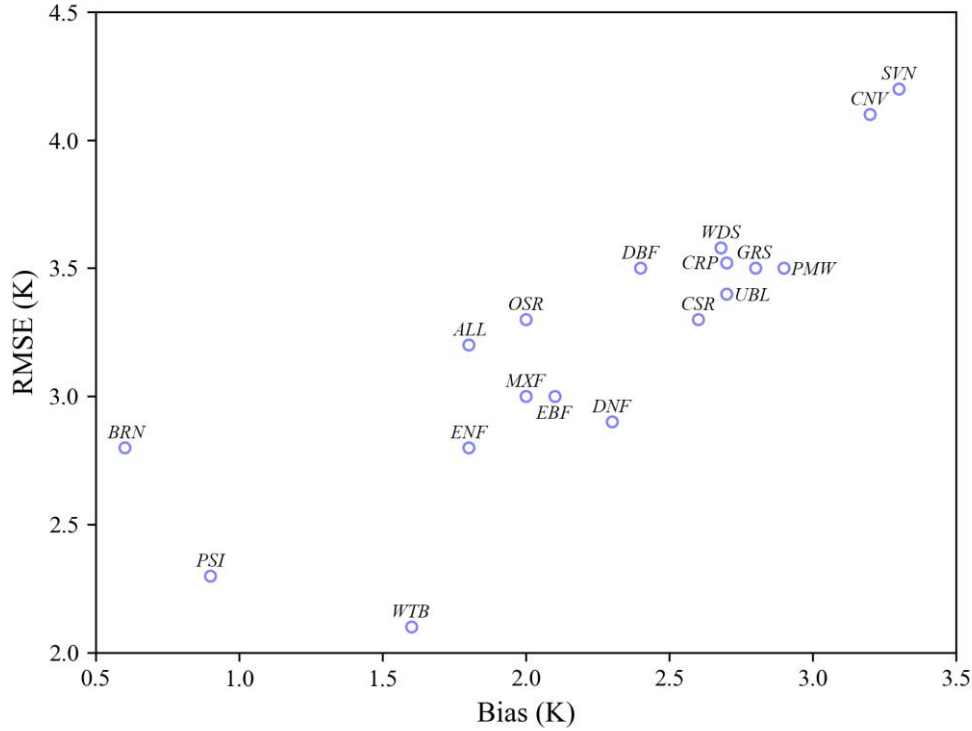


Figure 11. RMSD and bias between GT-LST and MYD11A1 LST in 2003 for various land cover types. ENF: evergreen needleleaf forests, EBF: evergreen broadleaf forests, DNF: deciduous needleleaf forests, DBF: deciduous broadleaf forests, MXF: mixed forests, CSR: closed shrublands, OSR: open shrublands, WDS: 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.

Q4.7 Line 357: why do savannas and cropland show considerable bias?

Response: Fig. R3 shows relatively large disparities between GT-LST and MYD11A1 LST over savannas (i.e., woody savannas and savannas) and croplands (i.e., cropland/natural vegetation mosaics and croplands) for the intercomparison. We would like to make some explanations on large disparities between these two products as follows:

According to NDVI threshold method, the daily emissivity of an AVHRR pixel can be derived using the following formula:

$$\varepsilon = \varepsilon_{veg} * FVC + \varepsilon_{soil} * (1 - FVC)$$

Here, ε is the emissivity, ε_{veg} is the vegetation emissivity, ε_{soil} is the bare soil emissivity, and FVC is the fraction of vegetation cover.

For a vegetation pixel, its FVC is less than 1 due to the influence of natural and

human factors, which leads to the underestimation of emissivity comparing with fixed emissivity, resulting in an overestimation of LST. The situation is particularly evident over croplands and savannas. Specially, natural disasters (e.g., drought and pests) and agricultural activities (e.g., harvest, cropland lies fallow) can significantly decrease cropland density and result in higher exposure of the soil. It leads to a decrease in cropland emissivity, resulting in an overestimation of LST. The emissivity for savannas decreases because of the increasing proportion of soil by grazing, fire and annually a long period in which moisture inadequate, resulting in an overestimation of LST.

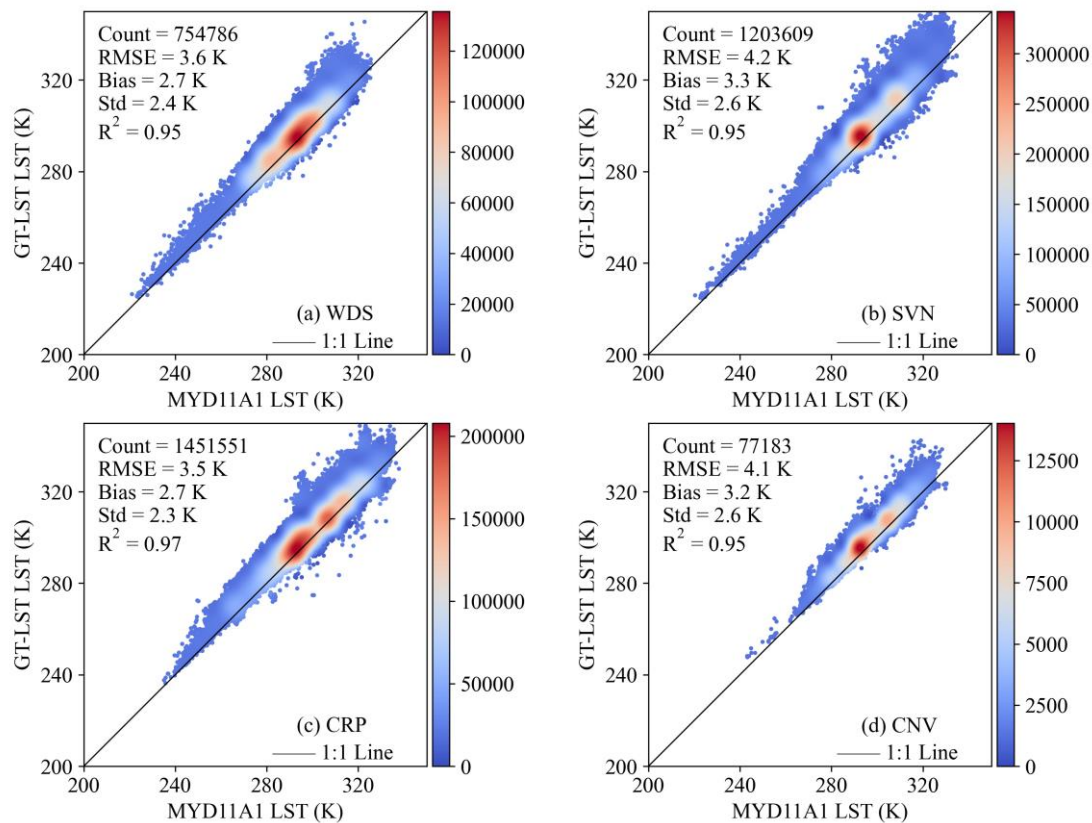


Figure R3. Scatterplots of GT-LST versus MYD11A1 LST during 2004 over WDS (a), SVN (b), CRP (c), and CNV (d). WDS: woody savannas, SVN: savannas, CRP: croplands, and CNV: cropland/natural vegetation mosaics.

Minor:

1. Line 35: Some of them used surface air temperature rather than LST to detect climate change and it should be not mixed.

Response: Thank you for your careful reading. We have removed the reference (i.e., Keenan and Riley, 2018) in the revised manuscript.

2. Line 71: remove ‘the’

Response: Corrected as suggested.

3. Line 94: polar-orbiting

Response: Corrected as suggested.

4. line 101: the first

Response: Corrected as suggested.

5. Line 179: Especially

Response: Corrected as suggested.

6. Line 298: identifier

Response: Corrected as suggested.

7. Line 301: difference

Response: Corrected as suggested.

8. Line 317: due to -> because

Response: Corrected as suggested.

9. Line 327: RMSEs

Response: Corrected as suggested.

10. Line 403: remove ‘in’

Response: Corrected as suggested.

11. Line 404: 'due to' should be followed by a noun rather than a sentence, suggest revising the whole manuscript for this issue.

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

12. Line 411: considers

Response: Corrected as suggested.

13. Line 446: open-source

Response: Corrected as suggested.

14. Line 451: cloud mask

Response: Corrected as suggested.

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, doi:10.1175/JHM-D-17-0009.1, 2017.

Coll, C., Wan, Z., Galve, J. M.: Temperature-based and radiance-based validations of the V5 MODIS land surface temperature product, J Geophys. Res-Atmos., 114, 1-15, doi:10.1029/2009jd012038, 2009.

Hulley, G. C., Hook, S. J.: The North American ASTER land surface emissivity database (NAALSED) version 2.0, Remote Sens. Environ., 113, 1967-1975, doi:10.1016/j.rse.2009.05.005, 2009.

Liu, X., Tang, B.-H., Yan, G., Li, Z.-L., Liang, S.: Retrieval of Global Orbit Drift Corrected Land Surface Temperature from Long-term AVHRR Data, *Remote Sens.*, 11, 2843, doi:10.3390/rs11232843, 2019.

Mao, K., Shi, J., Li, Z. L., Tang, H.: An RM-NN algorithm for retrieving land surface temperature and emissivity from EOS/MODIS data, *J Geophys. Res-Atmos.*, 112, 1-17, doi: 10.1029/2007JD008428, 2009.

Martins, J. P., Coelho e Freitas, S., Trigo, I. F., Barroso, C., Macedo, J.: Copernicus Global Land Operations-Lot I “Vegetation and Energy” Algorithm Theoretical Basis Document. Land Surface Temperature—LST, 1, 2019.

Reiners, P., Asam, S., Frey, C., Holzwarth, S., Bachmann, M., Sobrino, J., Göttsche, F., Bendix, J. Kuenzer, C.: Validation of AVHRR Land Surface Temperature with MODIS and In Situ LST—A TIMELINE Thematic Processor, *Remote Sens.*, 13, 3473, doi:10.3390/rs13173473, 2021.

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, doi:10.1016/j.isprsjprs.2021.05.017, 2021.

Wan, Z., Zhang, Y., Zhang, Q., Li, Z. L.: Validation of the land-surface temperature products retrieved from Terra Moderate Resolution Imaging Spectroradiometer data, *Remote Sens. Environ.*, 83, 163-180, doi:10.1016/S0034-4257(02)00093-7, 2002.