<u>Responses to the Manuscript essd-2022-83 RC2</u>:

A global dataset of spatiotemporally seamless daily mean land surface temperatures: generation, validation, and analysis

Dear reviewer #2,

The authors would like to thank you for providing us with thoughtful and outstanding comments. We have addressed all comments in detail and revised the manuscript accordingly and tracked the changes so that you can see that we have rewritten many parts of the manuscript. Point-by-point responses are provided below.

We will be very glad to receive your feedback.

Yours sincerely,

Falu Hong, Wenfeng Zhan*, Frank-M. Göttsche, Zihan Liu, Pan Dong, Huyan Fu, Fan Huang, and Xiaodong Zhang

Email: zhanwenfeng@nju.edu.cn

I. TABLE OF CONTENTS

II. ATTENTIONS

(1) In the following responses, texts contained within the red braces {…} are identical to those in our revised manuscript.

(2) In the following responses, the line numbers [Line XXX-XXX] refer to the clean version of the revised manuscript.

(3) Fig. 1, 2, and 3…, and Eq. 1, 2, and 3… refer to the figures and equations excerpted from our revised manuscript.

(4) In the following responses, all the related references are provided collectively in Part IV References.

III. Responses to Reviewer #2

Comment #1

This paper describes an improved annual and diurnal temperature cycle-based framework method to generate global spatiotemporally seamless daily mean LST products from MODIS data with the support of reanalysis data. The developed dataset performs very well against global in-situ surface observations. Overall, this new method produces a 0.5 --degree daily product of daily mean LST over the globe. Given that this data has high spatial resolution at a daily time scale, it should be a useful tool for climate studies after its flaws are addressed.

Authors' reply:

Thanks very much for your appreciation. We have addressed the flaws you mentioned. Please refer to the following point-to-point response for the details.

Major comments

Comment #2

The developed GADTC product has a spatial resolution of 0.5-degree, how to deal with the scale mismatch between the in-situ measurements and the product, the validation can be carried out at a higher spatial resolution, such as MODIS original resolution. Maybe, the authors can classify the in-situ sites to different levels according to the spatial heterogeneity of the site, to further analyze the errors at different sites.

Authors' reply:

Thank you for your comment. This comment is related to three issues: (1) addressing the scale mismatch between *in situ* measurements and generated GADTC product; (2) validating the daily mean land surface temperature (LST) product at the MODIS original resolution; (3) analyzing the errors according to the spatial heterogeneity of the sites.

(1) Addressing the scale mismatch between in-situ measurements and product

We agree with you that the scale mismatch exists between *in situ* measurement and satellite-based LST product. To avoid the scale mismatch, we validate the framework merely based on *in situ* measurement, i.e., running the IADTC framework with *in situ* measurement and then using hourly measurements for validation. The results in Section 4.1 show that the mean absolute errors (MAEs) of the IADTC framework are 1.4 K and 1.1 K for SURFRAD and FLUXNET data, respectively. The validation results merely based on *in situ* measurements are better than the validation results through comparing *in situ* measurements and the GADTC product which involves the scale mismatch uncertainty.

(2) Validating the daily mean LST product at the MODIS original resolution

According to your suggestion, we ran the IADTC framework with the MOD11A1 and MYD11A1 LST products to validate the daily mean LST product at the MODIS original resolution (~1 km). The seven SURFRAD sites in 2019 were used for validation. Table R1 shows the validation results at the MODIS original resolution are comparable with the validation results at 0.5 degree, i.e., MAE around 2.2 K, except for the DRA site where the MAE exceeds 4.5 K at the original resolution. The abnormal larger errors at DRA site have been reported by previous studies which validated the instantaneous LST product (Duan et al., 2019; Ermida et al., 2020). For clarification, the unsuitable descriptions of the validation results at DRA site in the original manuscript (Line 357-359) have been deleted.

Table R1. Validation results at the MODIS original resolution with the seven SURFRAD sites in 2019.

Site ID	Bias (K)	MAE (K)	RMSE (K)	STD (K)	R-square
BON	-1.61	2.04	2.45	1.85	0.97
TBL	-0.67	2.20	2.76	2.68	0.94
DRA	-4.41	4.51	5.05	2.45	0.97
FPK	-1.07	2.20	2.86	2.65	0.97
GWN	-1.89	2.13	2.48	1.61	0.97
PSU	-2.08	2.27	2.70	1.73	0.98
SXF	-1.16	1.88	2.36	2.06	0.98

(3) Analyzing the errors according to the spatial heterogeneity of the site

We define the spatial heterogeneity of SURFRAD sites by calculating the standard deviation of the land cover types within the MODIS original resolution pixel footprint (Fig. R1 & Table R2). The land cover types were obtained from the LCMAP collection 1.1 land cover map in 2019 (Brown et al., 2020). Table R2 shows that BON and TBL sites are relatively homogeneous, and GWN and PSU sites are relatively heterogeneous. However, the validation results are not expected to be related to spatial heterogeneity. This is probably because, at MODIS original resolution (~1 km), the uncertainty of scale mismatch still exists, and other factors, such as the sensor differences and atmosphere correction uncertainties, can also affect the validation results. Due to these concerns, apart from the direct comparison between *in* situ measurement and satellite-based daily mean LST, we also validated the IADTC framework merely based on *in situ* measurement to avoid the uncertainty of scale mismatch.

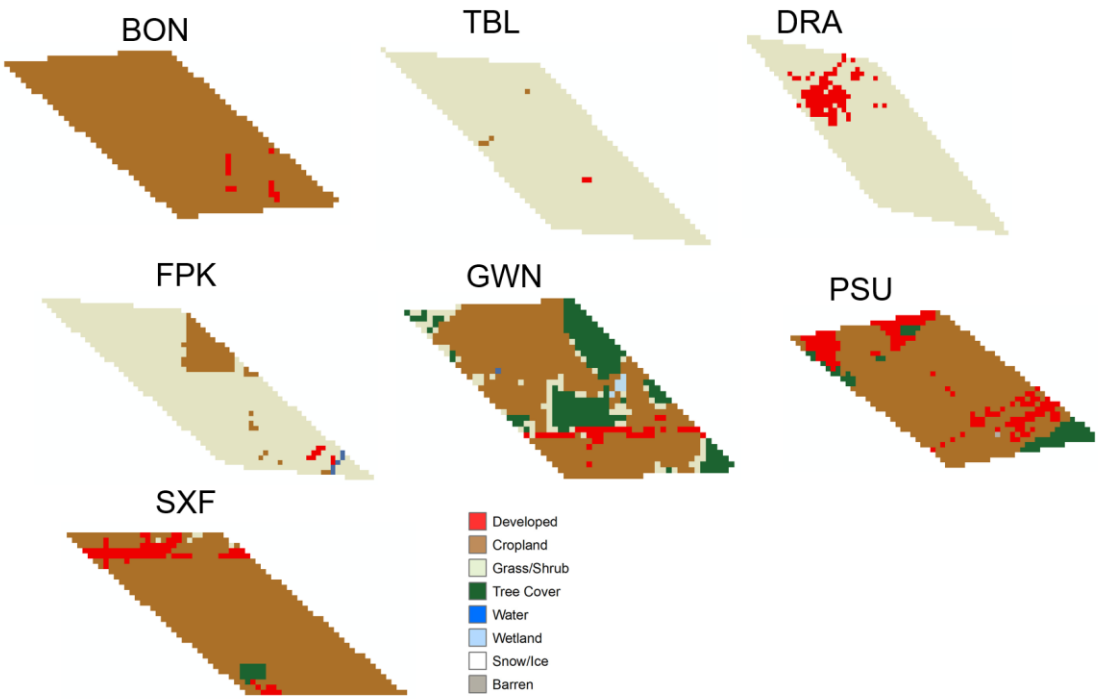

Fig. R1. The land cover types of each SURFRAD site within the MODIS 1-km pixel footprint.

Table R2. The standard deviation of the land cover types of each SUFRAD site within the MODIS pixel footprint.

Site ID	Land cover STD
BON	0.111
TBL	0.112
DRA	0.582
FPK	0.376

GWN	0.907
PSU	0.691
SXF	0.385

Comment #3

The Surfrad site only has 7 sites, Why not merge the data from the Surfrad and Fluxnet networks when validating the Tdm product. Also, in section 5.1, the ΔDTR can be obtained using the Surfrad and Fluxnet data together.

Authors' reply:

Thank you for your comment. We agree with you that the validation results using the merged SURFRAD and FLUXNET datasets should be provided for readers' reference and convenience. Therefore, we added the contents displaying the validation results using the merged SURFRAD and FLUXNET. The updated Fig. 8 and Fig. 11 in the revised manuscript display the validation results of the T_{dm} product and the determination of ΔDTR with the merged SURFRAD and FLUXNET datasets, which would be given at the end of this reply for your convenience.

However, in the revised manuscript, we still kept the separate validation results because the differences between SURFRAD and FLUXNET networks can also provide valuable information for readers. Their differences were summarized as follows:

(1) Their data sources are different. The SURFRAD sites have been managed uniformly by National Oceanic and Atmospheric Administration (NOAA) for over 15 years, and the associated radiance measurements have been consistently quality-controlled (Augustine et al., 2000). In contrast, FLUXNET sites are managed by different principal investigators. The quality control might not be consistent as SURFRAD sites.

(2) Their observation numbers are unevenly distributed. The number of FLUXNET sites is far more than the number of SURFRAD sites (126 *vs* 7). Consequently, the number of FLUXNET observations is far more than SURFRAD observations (226220 *vs* 42600). If we merged these two datasets, the results would be determined predominantly by FLUXNET dataset which occupies the majority. In other words, the contribution of SURFRAD dataset would be largely ignored.

(3) The covering land cover types are different. FLUXNET sites are mainly

located in vegetated areas. In contrast, the land cover types of the SURFRAD sites are not limited to vegetated areas. SURFRAD sites additionally cover barren area (the DRA site). Merging them would reduce the contributions from diverse land cover types.

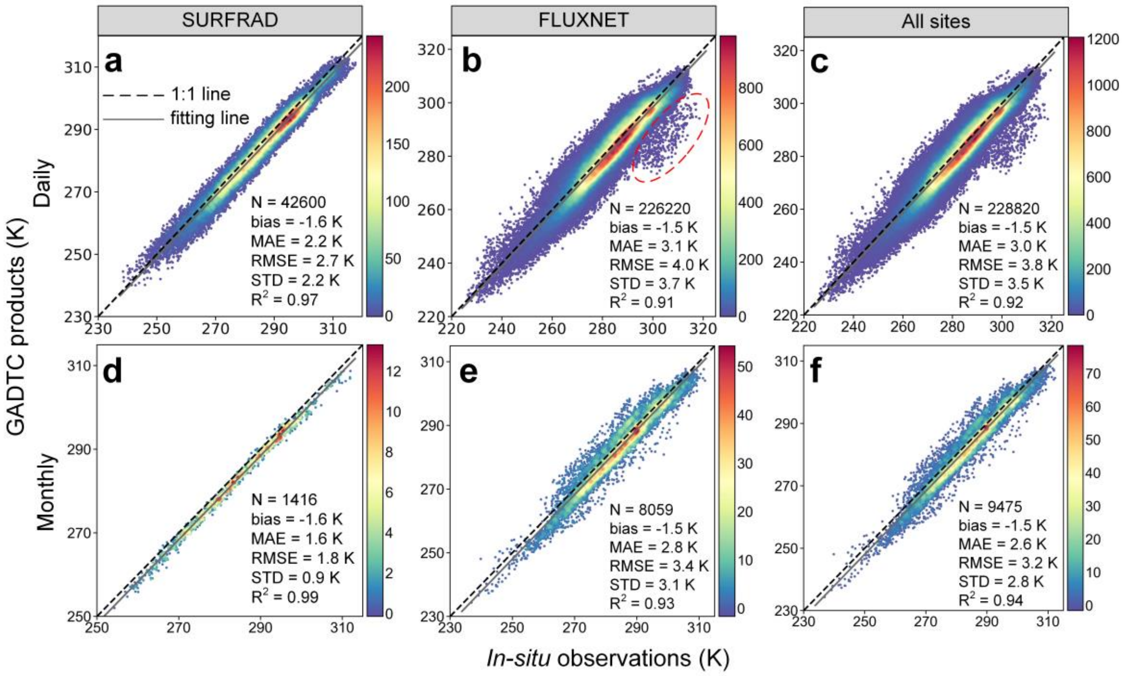

Fig. 1. GADTC products versus *in situ* observations. (a), (b), and (c) compare the daily mean LST over the SURFRAD, FLUXNET, and combined sites, respectively; and (d), (e), and (f) show the corresponding results for monthly mean LST. The biases were calculated by the GADTC products minus the *in situ* measurements. The red ellipse in (b) highlights the cases with notably large errors.

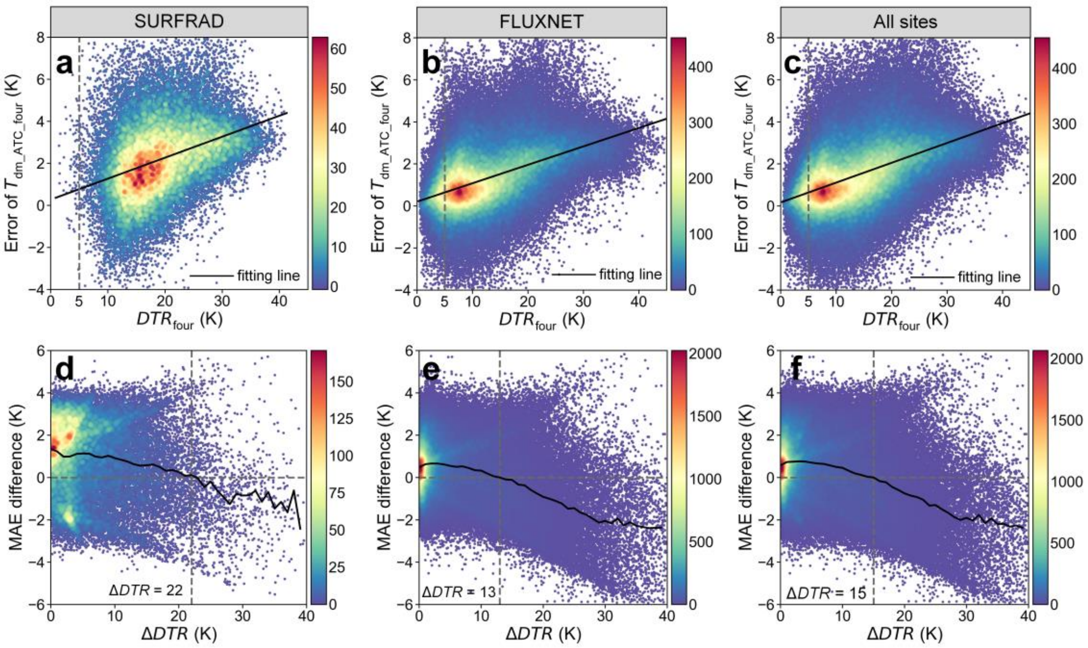

Fig. 2. Threshold determination for the two criteria in Error! Reference source not found.. **(a), (b), and (c) display the errors of $T_{dm\_ATC\_four}$ ($T_{dm\_ATC\_four}$ minus T_{dm_true}) depending on DTR_{four} for SURFRAD, FLUXNET, and combined data, respectively; and (d), (e), and (f) display the MAE differences between $T_{dm\_ATC\_four}$ and $T_{dm\_ATC\_DTC}$ (i.e., the MAE of $T_{dm\_ATC\_four}$ minus the MAE of $T_{dm\_ATC\_DTC}$) depending on the Δ*DTR* for SURFRAD, FLUXNET, and combined data, respectively. The black lines in (d), (e), and (f) denote the averaged MAE difference within every unit along the x-axis.**

Comment #4

The authors used MAE and bias, why not use the RMSE, which is typically used in the LST validation.

Authors' reply:

Thank you for your comment. We agree with you that the RMSE results should be included in the LST validation results. The updated Fig. 8, Table 1, and Table 2 are given as follows for your convenience.

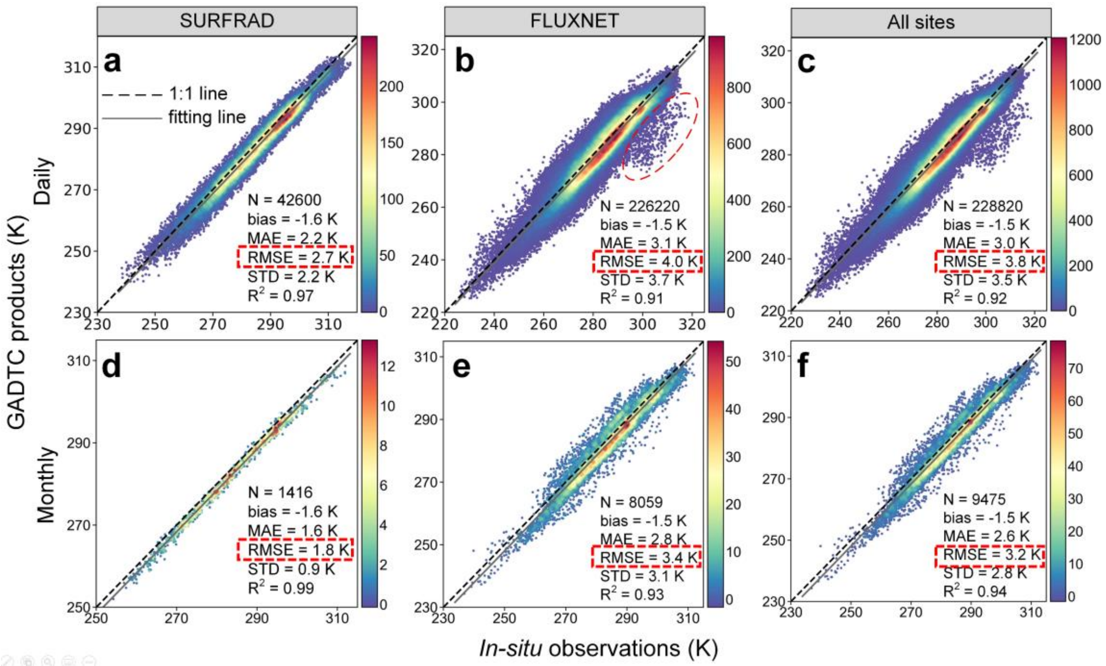

Fig. 3. GADTC products versus *in situ* observations. (a), (b), and (c) compare the daily mean LST over the SURFRAD, FLUXNET, and combined sites, respectively; and (d), (e), and (f) show the corresponding results for monthly mean LST. The biases were calculated by the GADTC products minus the *in situ* measurements. The red ellipse in (b) highlights the cases with notably large errors.

Table 1. Validation results obtained over the seven SURFRAD sites.

Site ID	Lat./Long.	IGBP	*N**	Bias (K)	MAE (K)	RMSE (K)	STD (K)
BON	40.05°/−88.37°	CRO	6153	−1.20	1.97	2.44	2.12
TBL	40.13°/−105.24°	GRA	6124	−1.37	2.30	2.89	2.54
DRA	36.62°/−116.02°	BSV	6102	−2.04	2.26	2.69	1.74
FPK	48.31°/−105.10°	GRA	6157	−1.78	2.54	3.18	2.63
GWN	34.25°/−89.87°	WSA	6144	−1.83	2.25	2.70	1.98
PSU	40.72°/−77.93°	CRO	6134	−1.30	1.85	2.24	1.82
SXF	43.73°/−96.62°	CRO	5786	−1.39	2.06	2.54	2.13

*: *N* denotes the number of days used for validation.

Table 2. Validation results for the GADTC products stratified by IGBP land cover type of the FLUXNET sites.

IGBP	Site number	*N**	Bias (K)	MAE (K)	RMSE (K)	STD (K)
MF	5	7564	−1.95	2.62	3.25	2.61

EBF	11	29588	−1.71	2.75	3.34	2.87
WET	15	14556	−0.66	2.76	4.22	4.17
DBF	19	32594	−1.78	2.89	3.56	3.08
SAV	5	10355	−2.65	3.16	3.84	2.79
CRO	14	14387	−1.59	3.26	4.10	3.78
GRA	23	45257	−1.62	3.32	4.22	3.90
ENF	25	58616	−0.81	3.38	4.18	4.10
WSA	5	7810	−2.33	3.44	4.06	3.32
OSH	3	5090	−3.34	3.62	4.33	2.75
SNO	1	403	−3.39	4.80	5.91	4.84

*: *N* denotes the number of days used for validation.

Minor comments

Comment #5

Line 67, some latest papers about the C6 MODIS LST accuracy can be added, such as DOI: 10.1109/TGRS.2020.2998945, https://doi.org/10.1016/j.jag.2018.04.006

Authors' reply:

Thank you for your reminder. We have added the reference you mentioned.

Comment #6

Line 104, the MxD11C1 was derived using the day/night algorithm and giving a reference

Authors' reply:

Thanks for your comment. We have added the reference from Wan and Li (1997) which is the representative study using the day/night algorithm to derive land surface temperature. The revised sentence in Line 103-105 is given as follows for your convenience.

Line 103-105:

{The MODIS LSTs were retrieved with a refined generalized split-window algorithm, and their accuracies are mostly within 1.0 K over homogeneous surfaces (Wan and Li, 1997; Duan et al., 2019; Wan, 2014).}

Comment #7

Line 139, how to get the hourly values?

Authors' reply:

Thank you for your comment. SURFRAD *in situ* measurements can provide minutely observations and FLUXNET *in situ* measurements can provide half-hourly observations (a part of the sites provide hourly observations). To get hourly values, we aggregated minutely or half-hourly observations to hourly values. This step was to reduce the impact of short-term LST fluctuations caused by local weather variation. In Line 139-140, we mentioned the way how to get the hourly values, which were given as follows for your convenience.

Line 139-140:

{To reduce the impacts of short-term LST fluctuations on validation, we aggregated minutely observations into hourly values.}

Comment #8

Line 319, Scenarios #1 and #3, How many sites per scenario, the results can be analyzed by scenario, not by Surfrad and Fluxnet.

Authors' reply:

Thanks for your comment. We calculated the count and the percentage of each scenario for the SURFRAD and FLUXNET datasets (Table R3). In addition, we provided the accuracy results by scenario (Fig. R2). Table R3 shows that Scenarios #1, #2, and #3 covers 0.2%, 95.0%, and 4.8% for the SURFRAD datasets, and 10.2%, 82.5%, and 7.3% for FLUXNET datasets. Fig. R2 shows that for SURFRAD dataset, the MAE in Scenario #2 is the smallest, then followed by Scenario #1 and Scenario #3. For FLUXNET dataset, the order of MAE in each scenario is: Scenario #3 > Scenario #2 > Scenario #1. For both two datasets, the bias in Scenario #2 is slightly lower than zero, and the biases in Scenarios #1 and #3 are larger than zero. We should note that although the performances of IADTC framework in Scenarios #1 and #3 are not good as the performance in Scenario #2, the IADTC framework stills performs better than the OADTC framework in Scenarios #1 and #3 (refer to Fig. 6 and Fig. B1 in the manuscript).

We have added the descriptions of the percentage of each scenario for SURFRAD and FLUXNET sites. Please refer to Line 321-323 and Line 345-347, which were given as follows for your convenience. In Fig. B1 in the Appendix section, the MAEs under scenarios #1 and #3 were also provided for reader's convenience.

Line 321-323:

{The proportion of three scenarios were 0.2%, 95.0%, and 4.8%, respectively. In Scenarios #1 and #3 under which the accuracies were improved compared with the OADTC framework, the IADTC framework improves the MAE of estimated T_{dm} by around 0.45 K (from 2.80 K to 2.35 K, see Fig. B1a).}

Line 345-347:

{The proportion of each scenario is 10.2%, 82.5%, and 7.3%, respectively. Compared with the OADTC framework, in Scenarios #1 and #3 (the proportion is 17.4%) under which the accuracies are considerably improved, IADTC framework improved the MAE of the estimated Tdm by around 0.78 K (from 1.95 K to 1.17 K, refer to Fig. B1b).}

Table R3. The count and percentage of each scenario for the SURFRAD and FLUXNET datasets.

		Scenario #1	Scenario #2	Scenario #3
SURFRAD	Count	84	40820	2076
	Percentage	0.2%	95.0%	4.8%
FLUXNET	Count	19724	161095	14333
	Percentage	10.2%	82.5%	7.3%

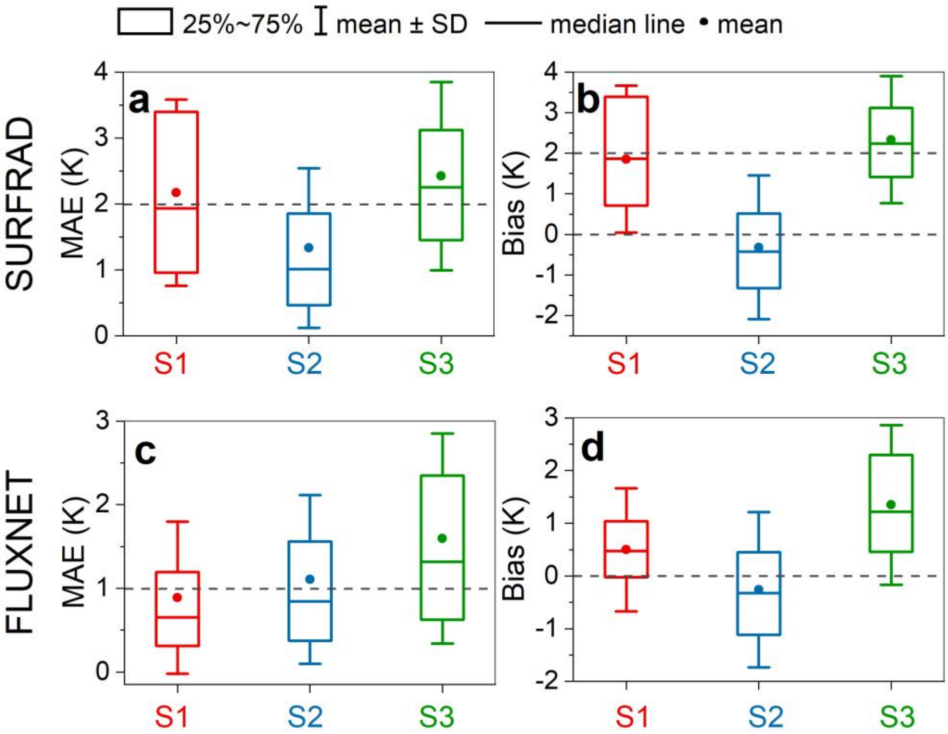

Fig. R2. Boxplot of errors of $T_{\mathrm{dm\_IADTC}}$ for each scenario. (a) and (b) display the boxplot of mean absolute error (MAE) and bias based on SURFRAD dataset, respectively; and (c) and (d) display are the same as (a) and (b), but for FLUXNET dataset.

Comment #9

Line 360, Fig.8, combines data from the two networks.

Authors' reply:

Thank you for your comment. This reply is related to Comment #3. We have added the figures showing the validation results using the combined data from the two networks in the revised Fig. 8.

Comment #10

Line 373, how to prove the large errors at these sites are related to the high spatial heterogeneity

Authors' reply:

Thank you for your comment. We need to clarify that spatial heterogeneity is one of the many possible reasons for causing large errors. Other factors, such as spatial representativeness and erroneous observations can also cause large errors. In Comment #2, the validation results at SURFRAD sites show that the errors could be large in the homogeneous sites, for example, the DRA site.

For the AU-Wac, CH-Fru, SJ-Adv, and US-Orv sites which have the top 4 largest RMSE ($\geq$ 8.0 K) among the selected 126 FLUXNET sites, we have checked their google earth image within the 0.5 $\times$ 0.5 degree and found that their observation field is quite different from their located 0.5-degree grids (Fig. R3). Therefore, we speculate that the larger errors at these sites are related to the high spatial heterogeneity. We clarified this point in Line 375-378, which was given as follows for your convenience.

Line 375-378:

{The relatively larger errors at several FLUXNET sites (e.g., AU-Wac, SJ-Adv, and CH-Fru sites, with MAEs larger than 8.0 K; refer to the red ellipse in Fig. 1e) partly account for the lower accuracy. The relatively large errors at these sites might be related to the erroneous *in situ* measurements as well as the high spatial heterogeneity around these sites.}

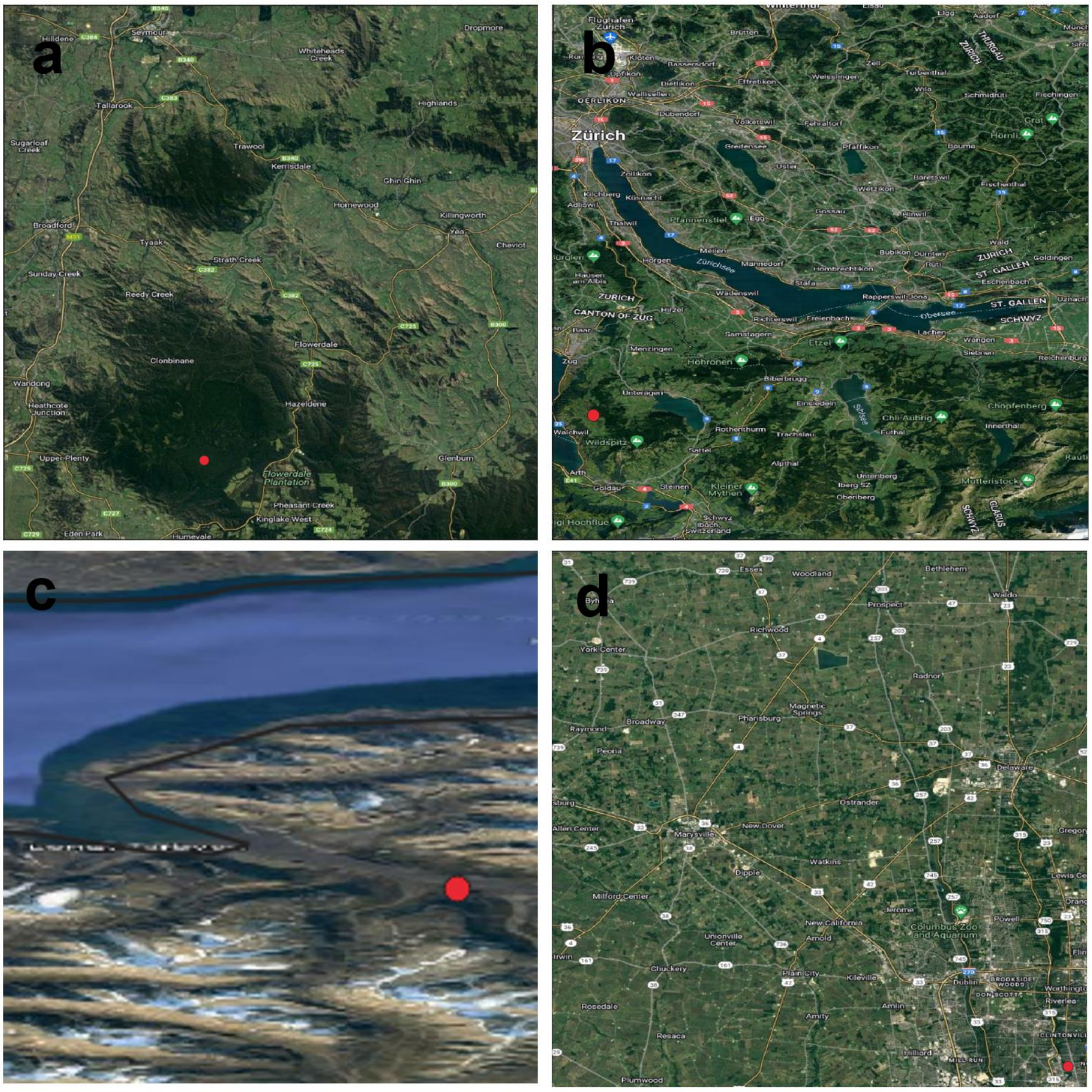

Fig. R3. The google earth images for the AU-Wac (a), CH-Fru (b), SJ-Adv (c), and US-Orv (d) sites. The image boundary is around 0.5 by 0.5 degree.

IV. References

Augustine, J. A., DeLuisi, J. J., and Long, C. N.: SURFRAD–A National Surface Radiation Budget Network for Atmospheric Research, Bull. Am. Meteorol. Soc., 81, 2341-2358, doi:10.1175/1520-0477(2000)081%3C2341:SANSRB%3E2.3.CO;2, 2000.

Brown, J. F., Tollerud, H. J., Barber, C. P., Zhou, Q., Dwyer, J. L., Vogelmann, J. E., Loveland, T. R., Woodcock, C. E., Stehman, S. V., Zhu, Z., Pengra, B. W., Smith, K., Horton, J. A., Xian, G., Auch, R. F., Sohl, T. L., Sayler, K. L., Gallant, A. L., Zelenak, D., Reker, R. R., and Rover, J.: Lessons learned implementing an operational continuous United States national land change monitoring capability: The Land Change Monitoring, Assessment, and Projection (LCMAP) approach, Remote Sens. Environ., 238, doi:10.1016/j.rse.2019.111356, 2020.

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

Ermida, S. L., Soares, P., Mantas, V., Göttsche, F.-M., and Trigo, I. F.: Google Earth Engine open-source code for land surface temperature estimation from the Landsat series, Remote Sens., 12, 1471, doi:10.3390/rs12091471, 2020.

Wan, Z. and Li, Z. L.: A physics-based algorithm for retrieving land-surface emissivity and temperature from EOS/MODIS data, IEEE Trans. Geosci. Remote Sens., 35, 980-996, doi:10.1109/36.602541, 1997.