https://doi.org/10.5194/essd-2021-313
Preprint. Discussion started: 12 October 2021

Open Access Earth System Science Data Discussions

A global seamless 1 km resolution daily land surface temperature dataset (2003-2020)

Tao Zhang[1], Yuyu Zhou[1], Zhengyuan Zhu[2], Xiaoma Li[3], Ghassem R. Asrar[4]

[1] Department of Geological and Atmospheric Sciences, Iowa State University, Ames, IA, 50011, USA

[2] Department of Statistics, Iowa State University, Ames, IA, 50011, USA

[3] School of Landscape Architecture and Art Design, Hunan Agricultural University, Changsha, Hunan, 410128, China

[4] Universities Space Research Association, Columbia, MD, 21046, USA

Correspondence to: Yuyu Zhou (yuyuzhou@iastate.edu)

Abstract. Land surface temperature (LST) is one of the most important and widely used parameter for studying land surface processes. Moderate Resolution Imaging Spectroradiometer (MODIS) LST products (e.g., MOD11A1 and MYD11A1) can provide this information with high spatiotemporal resolution with global coverage. However, the broad applications of these data are hampered because of missing values caused by factors such as cloud contamination. In this study, we used a spatiotemporal gap-filling framework to generate a seamless global 1 km daily (mid-daytime and mid-nighttime) MODIS-like LST dataset from 2003 to 2020 based on standard MODIS LST products. The method includes two steps, 1) data pre-processing and 2) spatiotemporal fitting. In the data pre-processing, we filtered pixels with low data quality and filled gaps using the observed LST at another three time points of the same day. In the spatiotemporal fitting, first, we fitted the long-term trend (overall mean) of observations in each pixel (ordered by day of year). Then we spatiotemporally interpolated residuals between observations and overall mean values for each day. Finally, we estimated missing values of LST by adding the overall mean and interpolated residuals. The results show that the missing values in the original MODIS LST were effectively and efficiently filled, and there is no obvious block effect caused by large areas of missing values, especially near the boundary of tiles, which might exist in other seamless LST datasets. The cross-validation with different missing rates at the global scale indicates that the gap-filled LST data have high accuracies with the average root mean squared error (RMSE) of 1.88℃ and 1.33℃, respectively for mid-daytime (1:30pm) and mid-nighttime (1:30am). The seamless global daily (mid-daytime and mid-nighttime) LST dataset at a 1 km spatial resolution is of great use in global studies of urban systems, climate research and modeling, and terrestrial ecosystems studies. The data are available at Iowa State University's DataShare at https://doi.org/10.25380/iastate.c.5078492 (Zhang et al., 2021a).

1 Introduction

Land surface temperature (LST) is an important variable for studies of energy balance, evapotranspiration, ecosystem processes in monitoring of Earth's resources (Anderson et al., 2010; Long et al., 2020). It has been widely used in various studies such as urban heat island (Li et al., 2021; Liu et al., 2020b; Tang et al., 2017; Yue et al., 2019), hydrology (Bai et al., 2019; Zhang et al., 2017), meteorology (Anderson et al., 2010; Li et al., 2018b), ecology (Sims et al., 2008), and energy systems (Peng et al., 2012; Zhou et al., 2014b). LST varies significantly in both space and time due to the spatiotemporal variation of factors such as solar radiation, atmospheric conditions, land surface characteristics (Li et al., 2018a; Peng et al., 2014; Zhang et al., 2015). LST can be measured

https://doi.org/10.5194/essd-2021-313
Preprint. Discussion started: 12 October 2021

in situ, obtained from land surface modeling, and measured by remote sensing (Ford and Quiring, 2019; Sheffield et al., 2018). Remotely sensed LST is by far the most widely obtained/used due to its global spatial coverage, high spatiotemporal resolutions, and available long-term data records.

LST products with a variety of spatial and temporal resolutions have been developed from different sensors/satellites such as: (1) high spatial resolution of about 100m and low temporal resolution of about every 16 days from Landsat (Parastatidis et al., 2017; Roy et al., 2014) and ASTER (Hulley et al., 2015); (2) coarse spatial resolution of 3-5km but high temporal resolution sub-daily to sub-hourly from geostationary satellites (Choi and Suh, 2013; Duguay-Tetzlaff et al., 2015; Jiang and Liu, 2014; Trigo et al., 2008; Yu et al., 2009); and (3) high spatial resolution about 1 km and high temporal resolution of daily from the Moderate Resolution Imaging Spectroradiometer (MODIS) (Wan, 2013, 2014), Visible Infrared Imaging Radiometer Suite (VIIRS) (Guillevic et al., 2014), and Sea and Land Surface Temperature Radiometer (SLSTR) LST (Ghent et al., 2017). Among them, MODIS LST is the most widely used especially for regional and global studies due to its global coverage and long-term and well calibrated and documented data records (since 2000) (Aguilar-Lome et al., 2019; Li et al., 2017; Peng et al., 2012; Sandeep et al., 2021; Zhao et al., 2020b; Zhou et al., 2019). However, MODIS LST has large number of missing values due to a variety of factors such as cloud contamination, non-overlapping satellite orbits, and instrumental malfunction (Crosson et al., 2012; Li et al., 2018a; Liu et al., 2020a; Shen et al., 2015; Wan, 2013).

Filling the missing values of MODIS LST is an effective way to overcome this limitation in MODIS LST product. Several such seamless datasets have been developed in previous studies. Zhao et al. (2020a) reconstructed monthly MODIS LST data in China (2003 – 2017). Metz et al. (2017) published a global 3 arc-min (~5.6 km at the equator) monthly average, minimum and maximum LST data (2003 – 2016). Cheng (2021) published a daily (mid-daytime and mid-nighttime) 1 km seamless LST of China (2002 – 2020). Zhang et al. (2021c) produced daily 1 km all-weather land surface temperature dataset for the Chinese land-surface and surrounding areas (2000 – 2020). Zhan et al. (2021) produced a global 1 km daily average LST dataset (2003 – 2019). Li et al. (2018a) produced a 1 km daily (mid-daytime and mid-nighttime) LST dataset for urban and surrounding rural areas of conterminous United States in 2010. However, there are still limitations in these products regarding spatial coverage or accuracy, especially in areas where LST has strong spatial and temporal variations such as cities. A global 1 km daily minimum and maximum LST dataset that can be used for a variety of studies and applications by scientists and practitioners such as city planners and water resources managers is still not available.

A variety of gap-filling methods have been proposed to fill gaps in MODIS LST. These methods can be divided into four groups (Li et al., 2018a; Weiss et al., 2014; Zhang et al., 2020). The first group is based on data fusion methods, which combine LST data from different satellites or different overpasses times (e.g., morning and afternoon) of the same satellite on the same day (Crosson et al., 2012; Duan et al., 2017; Long et al., 2020; Xu and Cheng, 2021; Zhang et al., 2020, 2021b). The second group is based on empirical relationship among different methods that were used to estimate the missing values by fitting empirical relationship between LST and auxiliary data (e.g., latitude, longitude, altitude, surface moisture, normalized difference vegetation index, and ground observed LST) (Fan et al., 2014; Ke et al., 2013; Zhao et al., 2020a). The third group is based on the internal spatiotemporal relationship that predicted the missing values with the available LST using algorithms such as temporal interpolation (Kilibarda et al., 2014; Xu and Shen, 2013), spatial interpolation (Ke et al., 2013; Yang, 2004), spatiotemporal interpolation (Sun et al., 2017; Weiss et al., 2014), and multi-dimensional smoothing (Garcia, 2010, 2011; Liu et al., 2020a; Pham et al., 2019). The fourth group is hybrid method that combined several methods from previous groups mentioned above (Hong et al., 2021; Li et al., 2018a; Metz et al., 2017; Weiss et al., 2014).

However, most of the current methods have some shortcomings in accuracy and efficiency for producing globally consistent and seamless MODIS-like LST. For example, the data fusion method has the problem of mismatch between LST from different

https://doi.org/10.5194/essd-2021-313
Preprint. Discussion started: 12 October 2021

Open Access Earth System Science Data Discussions

sources and usually cannot fully fill gaps (Crosson et al., 2012). The method based on empirical relationship is computationally intensive for global applications and may not fully capture the spatial and temporal variations of LST as the auxiliary data have low temporal resolution (Fan et al., 2014; Ke et al., 2013; Zhao et al., 2020a). The temporal interpolation and multi-dimensional smoothing methods are computationally efficient but may miss short-term temporal variations of LST (Kilibarda et al., 2014; Xu and Shen, 2013). The spatial interpolation methods may lead to physically unrealistic features in the interpolated LST when there are a lot of missing observations (Ke et al., 2013; Yang, 2004). The spatiotemporal interpolation methods can capture the short-term changes of LST but are time-consuming due to the use of local moving windows for each pixel (Li et al., 2018a; Weiss et al., 2014). The hybrid methods take the advantages of methods mentioned above and carry with it some of shortcomings of these methods, and may actually amplify them in the process of merging data imputed using different methods.

We proposed a spatiotemporal gap-filling framework to gap-fill missing values in MODIS LST product with good accuracies and high efficiencies. This framework includes two key steps of preprocessing and spatiotemporal fitting. Based on this framework, we developed a global 1 km daily (mid-daytime and mid-nighttime) LST dataset from 2003 to 2020 using the 1 km daily MODIS LST product. The remainder of this paper describes the study area and data (Sect. 2), the proposed spatiotemporal gap-filling approach (Sect. 3), the results and discussion (Sect. 4), data availability (Sect. 5), and conclusions (Sect. 6).

2 Study area and data

The study area is nearly the entire global land surface, including 178 MODIS tiles (Fig. 1). The 1 km daily MODIS LST product Version 6 from 2003 to 2020 is the primary data used in this study. It was produced based on the National Aeronautics and Space Administration (NASA) Earth Observing System (EOS) satellites Terra and Aqua (MOD11A1 and MYD11A1) (Wan, 2013, 2014). There are four observations each day from the two satellites (i.e., 10:30 am and 10:30 pm for Terra (T1 and T3), 1:30 am and 1:30 pm for Aqua (T2 and T4)). Another two auxiliary datasets used are the annual MODIS land cover product (MCD12Q1) (Sulla-Menashe and Friedl, 2018) and urban and their surrounding rural areas derived from nighttime light observations (Zhou et al., 2014a, 2018). Water pixels from the MCD12Q1 product were excluded in our analysis.

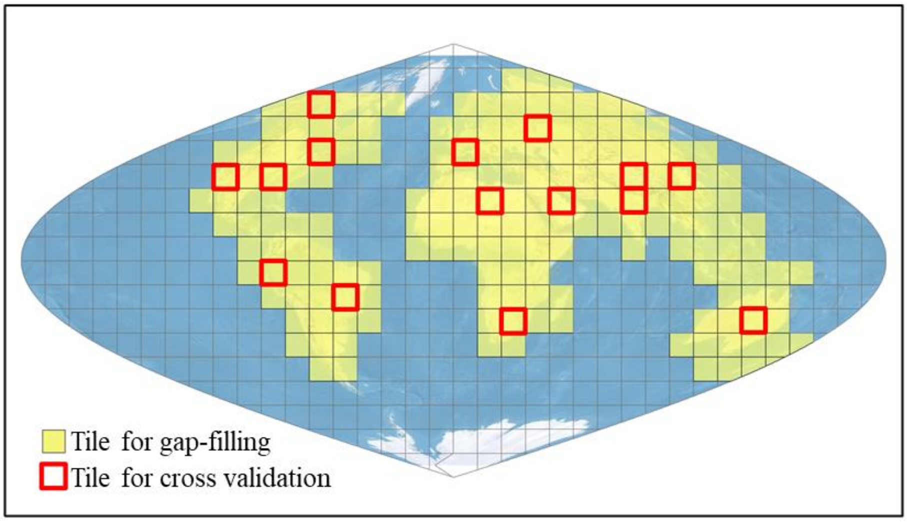


Figure 1: MODIS data tiles used in gap-filling and cross validation analysis.

3 Method

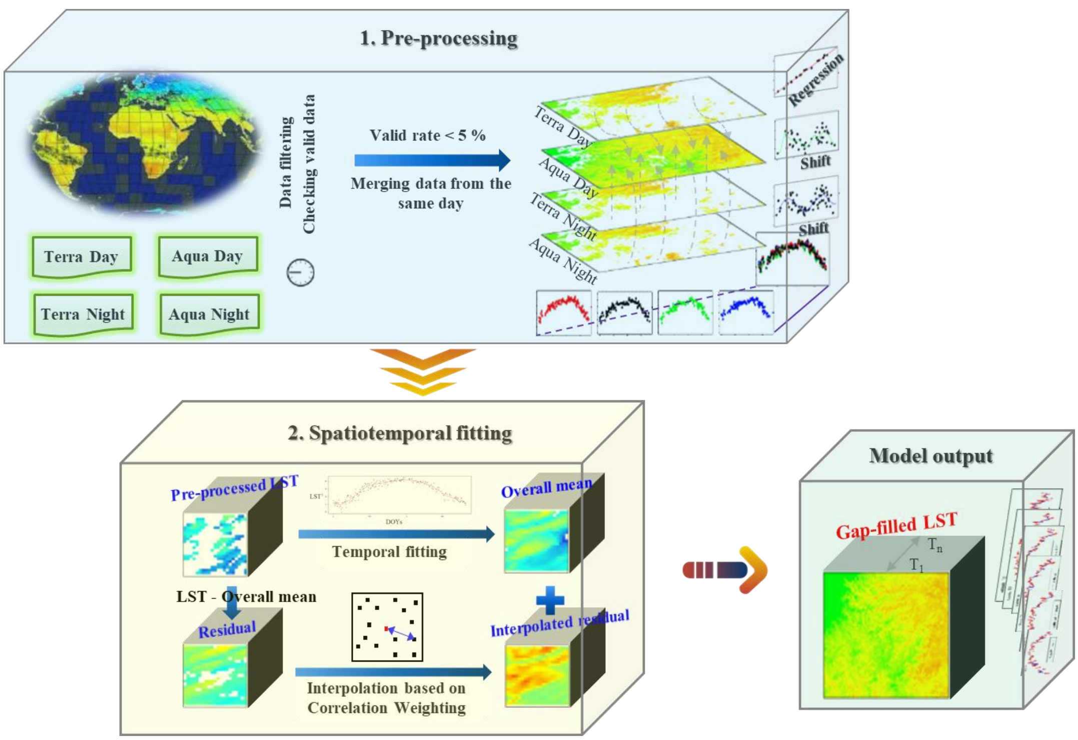
https://doi.org/10.5194/essd-2021-313
Preprint. Discussion started: 12 October 2021

Open Access Earth System Science Data Discussions

We developed a spatiotemporal gap-filling framework to gap-fill missing values in MODIS daily LST to produce a seamless 1 km spatial resolution global dataset from 2003 to 2020 (Fig. 2). The framework includes two key steps, 1) data pre-processing (Sect. 3.1) and 2) spatiotemporal fitting (Sect. 3.2). In the data pre-processing, we filtered pixels with low data quality and filled part of the gaps using the observed LST at another three time points of the same day to ensure high quality and good spatial coverage. In the spatiotemporal fitting, first, we fitted the long-term trend (overall mean) of observations in each pixel (ordered by day of year). Then we spatiotemporally interpolated residuals between observations and overall mean values for each day. Finally, we estimated missing values of LST by adding the overall mean and interpolated residuals. This gap-filling method was applied to MODIS LST at T2 (~1:30pm, Aqua Day in Fig. 2) and T4 (~1:30am, Aqua Night in Fig. 2), respectively, to build the 1 km daily LST (maximum (mid-daytime) and minimum (mid-nighttime)) data. In the sections below, we described each of these steps in detail.

3.1 Data pre-processing

Data pre-processing includes two parts: 1) data filtering and 2) daily merge. We first checked the quality of original MODIS data based on its quality assurance (QA) information and removed data points with error > 3K. We applied this threshold value because a stricter (or lower) value can exclude most of LST in urban areas (Crosson et al., 2012; Metz et al., 2017). Second, we conducted a daily merge using four observations from the two satellites (Terra and Aqua) in a given day using a modified algorithm from Li et al. (2018a). Taking a pixel with missing value of T2 as an example (Fig. 2), we calculated percent of valid data (PVD) in a year for all four observations, respectively. If PVD of T2 >= 5%, did not change the data and accepted it, otherwise we gap-filled missing values based on the following order: (1) estimated LST of T2 with LST of T1 using the regression method if PVD of T1 >= 5%; (2) estimated LST of T2 with LST of T4 using the shift method if PVD of T4 >= 5%; and (3) estimated LST of T2 with LST of T3 using the shift method if PVD of T3 >= 5%. Details of the regression and shift methods can be found in Li et al. (2018a).



https://doi.org/10.5194/essd-2021-313
Preprint. Discussion started: 12 October 2021

Figure 2: An overview of the spatiotemporal gap-filling framework.

3.2 Spatiotemporal fitting

The spatiotemporal fitting algorithm includes three steps (Fig. 3). First, we fitted the overall mean of observations in each pixel (ordered by day of year) using a smoothing spline function (Green and Silverman, 1994) to capture the overall trend. The time series of daily LST in a year (e.g., LST of T2) can be divided into two components, the overall mean (trend) and daily residual with gaps (daily fluctuations). Second, we spatiotemporally interpolated residuals for each day using a correlation-based method (Details in Sect. S1), in which the missing residual of a target pixel was estimated based on the correlation between target pixel and its neighboring valid pixels (i.e., with good quality). We selected 1% of the uniformly distributed pixels (10 km intervals) as representative neighboring pixels to perform the interpolation of residuals with high efficiency without reducing the accuracy based on our experiments. Moreover, we divided the global land surface area into 9 overlapped zones to avoid possible boundary effects (Details in Sect. S2). Finally, the seamless two parts (i.e., overall mean and daily residuals) were added to obtain the gap-filled LST data.

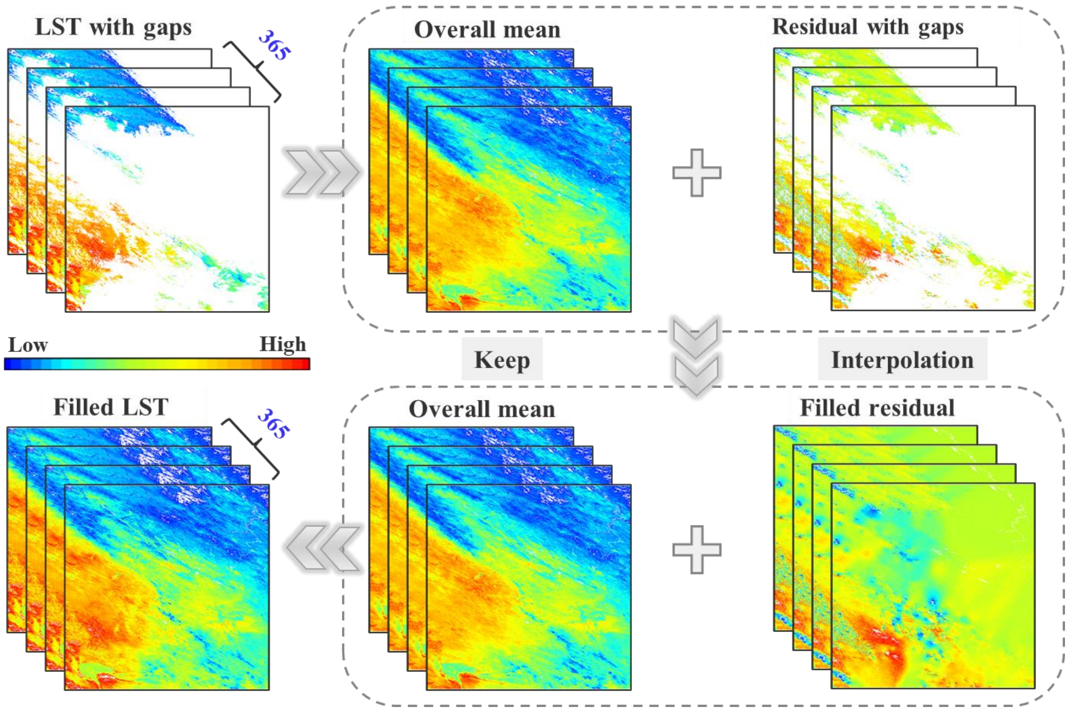


Figure 3: An example of the spatiotemporal fitting algorithm for gap-filling LST.

3.3 Accuracy assessment

We evaluated the accuracy of the gap-filled data using cross validation by randomly selecting 15 MODIS tiles in 2005, 2010, and 2015 (Fig. 1). In each of these years, we selected 19 days with the largest coverage of high quality data (coverage > 95% quantile) in cross validation. For each of selected days, we manually introduced gaps under three scenarios (i.e., excluding 25%, 50%, and 75% of valid pixels). Then we gap-filled these missing values and compared them with the originally values. We calculated root mean square error (RMSE) as the indicator of accuracy (Eq. 1).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{LST}_i - LST_i\right)^2} \tag{1}$$

https://doi.org/10.5194/essd-2021-313
Preprint. Discussion started: 12 October 2021

Where LST_i and $\widehat{LST}_i$ are original MODIS LST and gap-filled LST values of the *i*-th pixel; *n* is the number of the gap-filled pixels.

4 Results and discussion

4.1 Accuracy of gap-filled LST

The results of cross validation indicate the gap-filled LST has high accuracies (Fig. 4 and Table 1). The observed and gap-filled LSTs of representative pixels for different ratios of exclusion scattered along the 1:1 line with RMSE ranging from 2.05 to 2.31 °C and from 1.35 to 1.62 °C, respectively for daytime and nighttime (Fig. 4). As shown in Table 1, the average RMSE at tile level (i.e., calculating based on all the excluded pixels of each tile) ranges from 1.20 to 2.13 °C with an average of 1.88°C and 1.33°C, respectively for daytime and nighttime. The lowest RMSE occurs in 2010 with 25% excluding rate for nighttime, while the highest RMSE occurs in 2015 with 75% excluding rate for daytime. Compared with the accuracies at tile level, the accuracies for urban areas (i.e., calculating based on urban pixels in the excluded areas of each tile) are always higher with RMSE ranging from 1.14 to 2.06 °C (Table 1).

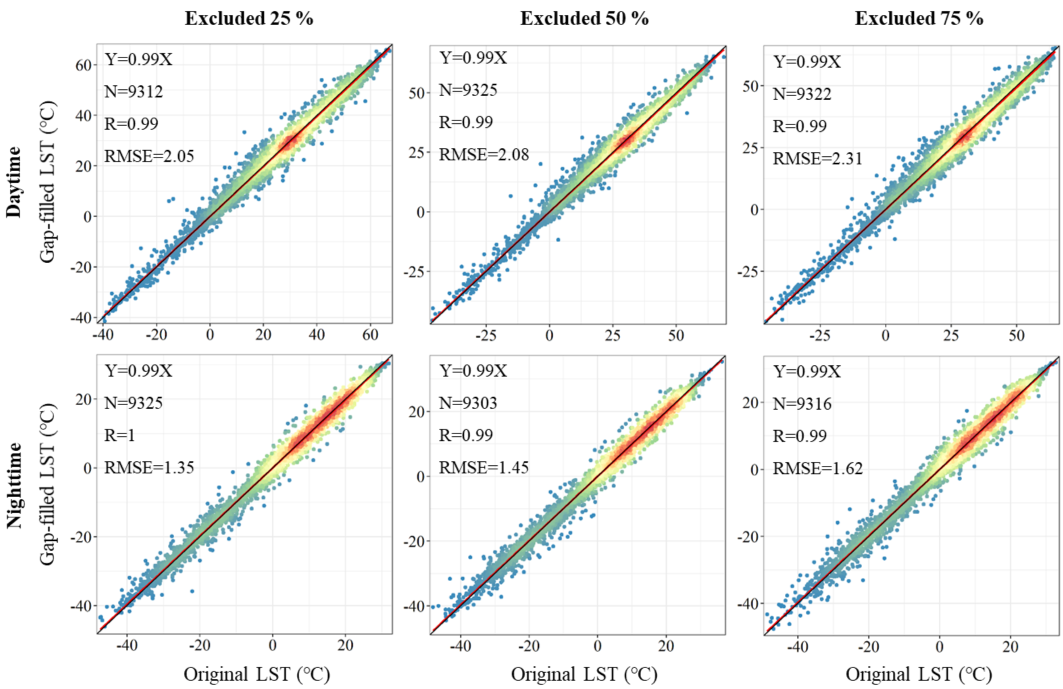


Figure 4: Scatter plots between gap-filled LST and original MODIS LSTs for daytime and nighttime in the excluded areas used for cross validation. Each point represents LST value of a pixel in the excluded area (25%, 50%, or 75%) of a tile (10 pixels per image) for cross validation analysis in years 2005, 2010, or 2015. The color of points represents the density of points, where the red points represent the highest density, and the blue points represent the lowest density. The red solid line represents the regression line, and the black line is the 1:1 line. The figures contain different number of pixels due to some of the selected pixels are water regions which are not suitable for accuracy assessment.

Table 1. Average RMSEs of 15 tiles used in cross-validation analysis of efficacy of the gap-filling method (Unit: °C)

Time	Year	RMSE in excluded area (±standard deviation)					
		25%		50%		75%	
		Tile level	Urban area	Tile level	Urban area	Tile level	Urban area
Daytime	2005	1.77 (±0.41)	1.68 (±0.46)	1.78 (±0.44)	1.74 (±0.55)	2.06 (±0.38)	2.01 (±0.52)

https://doi.org/10.5194/essd-2021-313
Preprint. Discussion started: 12 October 2021

	2010	1.74 (±0.49)	1.67 (±0.71)	1.76 (±0.50)	1.69 (±0.74)	1.91 (±0.52)	1.87 (±0.69)
	2015	1.90 (±0.53)	1.82 (±0.72)	1.91 (±0.54)	1.95 (±0.64)	2.13 (±0.55)	2.06 (±0.61)
Nighttime	2005	1.21 (±0.36)	1.15 (±0.35)	1.30 (±0.35)	1.23 (±0.35)	1.45 (±0.35)	1.42 (±0.44)
	2010	1.20 (±0.30)	1.14 (±0.32)	1.29 (±0.29)	1.29 (±0.39)	1.43 (±0.35)	1.38 (±0.44)
	2015	1.28 (±0.42)	1.19 (±0.33)	1.37 (±0.39)	1.25 (±0.36)	1.48 (±0.41)	1.47 (±0.44)

Note: 'Urban' means the urban and their surrounding rural areas. The tile level means the accuracy was calculated based on all the excluded pixels of each tile; urban area means the accuracy was calculated based on urban pixels in the excluded areas of each tile. Each RMSE value is the mean of RMSEs from all selected days in 15 selected MODIS tiles.

When the number of missing values in original LST increases, the gap-filled LST data tends to reduce in accuracy (Fig. 4, Table 1). As shown in Fig. 4, when the excluding rate increases from 25% to 75%, the RMSE of LST for daytime and nighttime increases from 2.05 to 2.31℃ and 1.35 to 1.62℃ for daytime and nighttime, respectively. This is also true for all years at tile level and in urban area in Table 1. However, the RMSE values are still within reasonable ranges. When the excluding rate is 75%, the RMSEs 2.31℃ and 1.62℃, respectively for daytime and nighttime (Fig. 4). Meanwhile, 88.9% of the RMSE in Table 1 is lower than 2 ℃.

4.2 Spatial and temporal patterns of LST

The examples of global LST data illustrate that the missing values in the original MODIS LST have been effectively gap-filled using the proposed gap-filling algorithm (Fig. 5). In the original MODIS LST, the continuously missing values mainly occur in Eastern Asia, South Asia, and Central Africa for both daytime and nighttime, in the example date (Fig. 5). In the gap-filled data, the missing values in these regions were fully gap-filled.

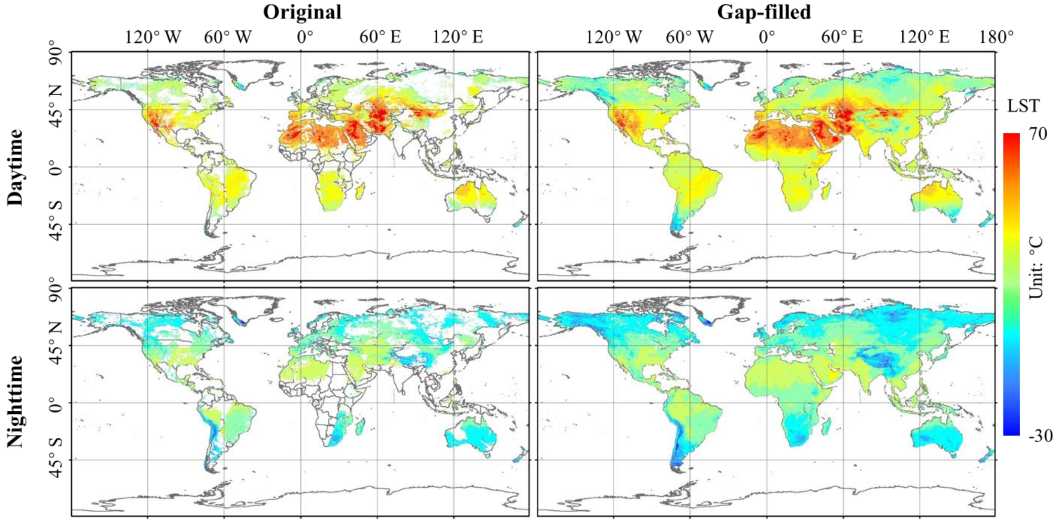


Figure 5: Spatial pattern of original and gap-filled LSTs at global scale in day 200 of year 2020.

The comparisons of spatial patterns between gap-filled and original MODIS LST in representative cities around the world (Fig. 6) illustrate that the missing values in the original MODIS LST have been effectively gap-filled at the city scale. As shown in Fig. 6, there is no missing value in the entire land surface area of the gap-filled data (water pixels were masked as NA). The gap-filled data capture well urban heat island (UHI) phenomenon (i.e., high temperature in urban than that of the surrounding rural areas). The spatial pattern of the gap-filled LST is reasonable with transition from urban to rural areas and there are no obvious boundary effects (more details in Sect. S3 and Sect. S4). For example, there is no obvious boundary effect between two MODIS tiles in the gap-filled LST data in Huston area, which suggests the interpolation of residuals (Sect. S2) in the proposed method are reliable.

https://doi.org/10.5194/essd-2021-313
Preprint. Discussion started: 12 October 2021

The gap-filled LST in the Pearl River Delta region shows a number of small speckles because this region is an agglomeration of sub-areas undergoing rapid urbanization.

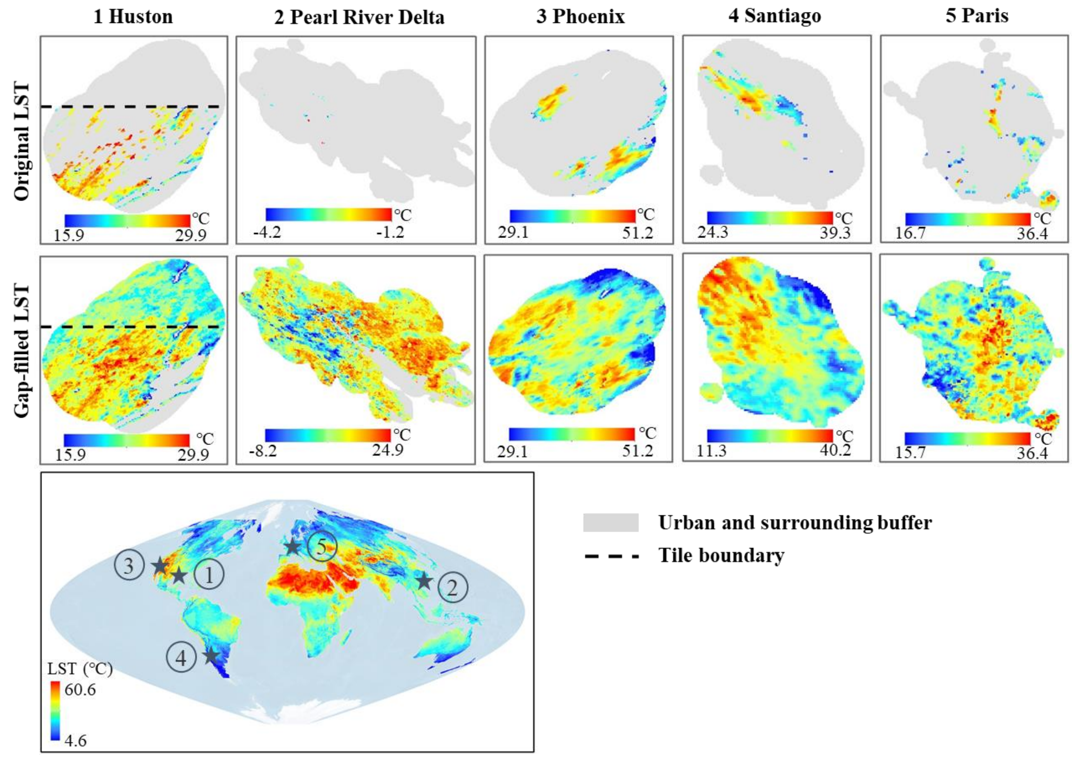


Figure 6: Spatial pattern of original and gap-filled LSTs in five representative cities. NA (gray color) in the gap-filled LST is water pixels.

The comparison of temporal pattern between gap-filled and original MODIS LSTs in a mega-city (Fig. 7) illustrates that the missing data in the original MODIS LST can be effectively and completely gap-filled for the entire period. As shown in Fig. 7, there are several days with limited valid (high quality) observations in original MODIS LST in Beijing, China in daytime in 2010, and these missing values were fully gap-filled in our data for the entire period. When there are only a few missing values in original LST data (days 28 and 130 in Fig. 7), the gap-filled and original LSTs show similar spatial pattern with significant UHI phenomenon. When there are large number of missing values in original LST data (days 219 and 293 in Fig. 7), the gap-filled LSTs can also illustrate the UHI phenomenon but different LST magnitudes with that of the former cases. Therefore, we may get more accurate estimation of annual average LST based on the gap-filled LST than original LST data.

https://doi.org/10.5194/essd-2021-313
Preprint. Discussion started: 12 October 2021

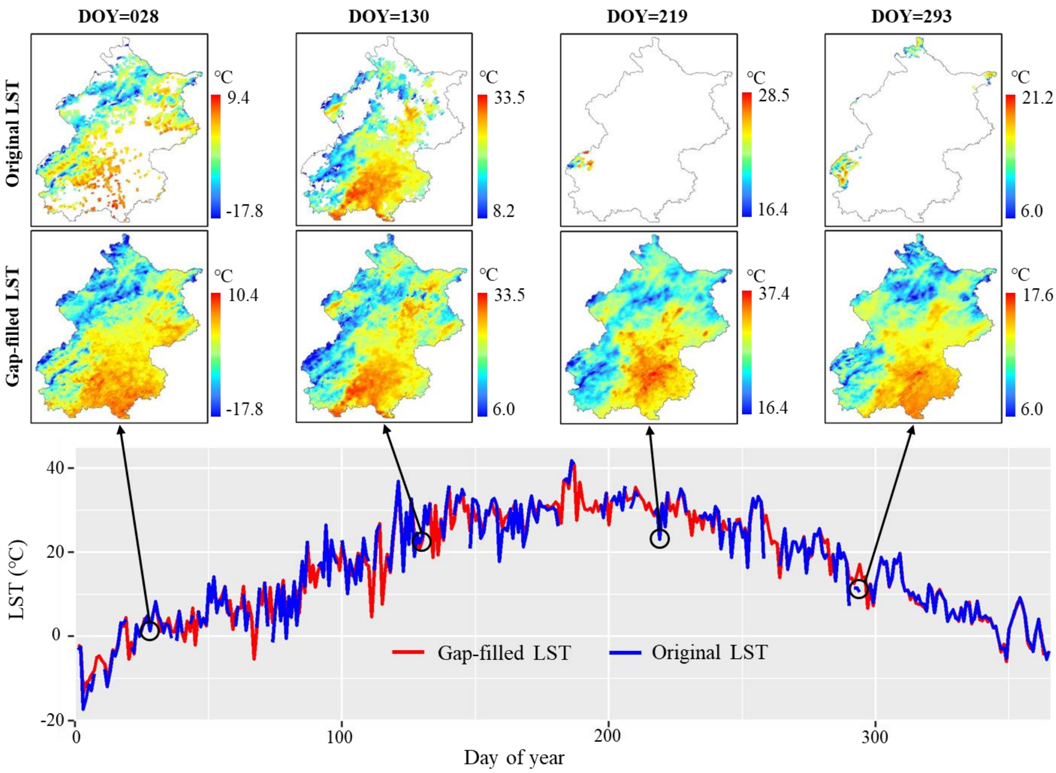


Figure 7: Temporal pattern of average daytime LST from original and gap-filled data in Beijing in year 2010. The black circles are example days showing maps of original and gap-filled LST data.

4.3 Comparison with existing seamless LST data

The accuracy of the resulting gap-filled LST from this study is comparable or better when compared with other reported seamless LST datasets. Our gap-filled LST data shows higher accuracies compared with the gap-filled LSTs based on the hybrid spatiotemporal gap-filling method proposed by Li et al. (2018a). These two datasets are most comparable because of the use of similar accuracy evaluation method (cross validation at the global scale) in both studies. In the hybrid method proposed by Li et al. (2018a), about 11% to 60% of the valid values (personal communication) were excluded for cross validation purpose in the urban areas at the global scale, and the average RMSE is 3.29℃ and 2.68℃, for daytime and nighttime, respectively. In this study, the average RMSE is 1.83℃ and 1.28℃ in the urban and surrounding areas for daytime and nighttime, respectively (Table 1). The accuracies of other seamless LST datasets were generally evaluated based on a limited number of in-situ LST observations (Zhang et al., 2019; Zhou et al., 2017), which are not exactly the same as satellite LSTs (Hong et al., 2021), and the evaluation in these studies are not directly comparable with our study.

The gap-filled LST in this study does not have the issue of boundary effect that might exists in the previous methods. Li et al. (2018a) combined several techniques including data fusion (Crosson et al., 2012), spatiotemporal interpolation (Gerber et al., 2018; Weiss et al., 2014), and temporal interpolation methods (Xu and Shen, 2013) to reconstruct daily (mid-daytime and mid-nighttime) LST. The systematic differences between neighboring regions with the use of different gap-filling techniques in the hybrid method may lead to boundary effects (personal communication). The gap-filled LST data in this study using the novel framework consisting of two key steps (Sect. S2) can mitigate boundary effects between neighboring regions (Fig. S2), neighboring tiles (Fig. S3), and within a given tile (Figs. 6 and Sect. S3).

https://doi.org/10.5194/essd-2021-313
Preprint. Discussion started: 12 October 2021

The gap-filled global 1km daytime and nighttime LST data have advantages regarding spatiotemporal resolutions (i.e., daily minimum and maximum) or coverage (i.e., global) and has significant potential for use in many disciplines of Earth system science and applications. In the existing seamless LST datasets, Zhan et al. (2021) produced a global daily average 1 km resolution LST dataset from 2003 to 2019, without resolving by daytime and nighttime. Zhao et al. (2020a) developed monthly average LST with 5.6km spatial resolution for China from 2003 to 2017. Cheng (2021) published a daily (mid-daytime and mid-nighttime) 1 km seamless LST of China from 2002 to 2020. Zhang et al. (2021c) generated daily (daytime and nighttime) 1 km all-weather LST dataset for China and its surrounding areas for 2000 to 2020. Li et al. (2018a) produced a 1 km daily (mid-daytime and mid-nighttime) LST dataset only in urban and surrounding rural areas of United States. The LST data in this study have a spatial resolution of 1 km and include daily LST at mid-daytime and mid-nighttime with a global coverage from 2003 to 2020, which has higher spatiotemporal resolutions or coverage than other existing seamless LST datasets.

The gap-filling framework proposed in this study can be efficiently implemented and has advantages regarding computing time compared to other algorithms/methods. For example, the gap-filling method proposed by Zhao et al. (2020a) were used for monthly 5.6km resolution LST data, and it may require significant computation time for high spatiotemporal resolution (daily, 1 km) LST data because it needs to calculate the distance between similar valid pixels and each target pixel (with missing or low quality value) based on a geographically weighted regression method. The gap-filling method proposed by Zhang et al. (2021b) is also complex and time consuming due to involvement of multi-source data and complex parameterization process on a pixel-by-pixel basis. The daily average LST data produced by Zhan et al. (2021) were calculated based on the nonlinear annual temperature cycle (ATC) and diurnal temperature cycle (DTC) modelling on a pixel-by-pixel basis, which is time-consuming for global scale applications (Hong et al., 2021). The hybrid gap-filling method proposed by Li et al. (2018a) is time consuming due to the use of spatiotemporal interpolation (Gerber et al., 2018; Weiss et al., 2014) algorithm, in which the missing value of a pixel at a specific time and location was interpolated by using a quantile regression in the corresponding local spatial and temporal window. In the proposed method in this study, the interpolation of the residual for a pixel at a specific time was also implemented by using the time series data of neighboring pixels in the corresponding local window. However, only 1% of pixels at fixed locations (10 km intervals) were used as neighboring pixels for interpolation of residual (Sect. S2), and the relevant parameters between target pixel and its neighboring pixels were calculated only one time for the entire period (365 days for a year). Our scheme can significantly improve the efficiency for global applications without reducing the accuracy according to our experiments.

5 Data availability

Data described in this manuscript can be accessed at Iowa State University's DataShare at https://doi.org/10.25380/iastate.c.5078492 (Zhang et al., 2021a). The dataset contains 36 sub datasets (one for each year in daytime and nighttime from 2003 to 2020). Each sub dataset contains LST data of a specific time (daytime or nighttime) and specific year (2003 – 2020) and is organized by day of year. The data are in GeoTIFF with the georeferenced information embedded. Each file keeps the MODIS Ellipse Sinusoidal projection with a spatial resolution of 1 km. The unit of LST in Geotiff is 0.1 Celsius temperature (0.1 ºC), and the naming rule can be found in the file of "README.pdf".

6 Conclusions

We propose a framework for filling the gap in long-term Earth observations and geophysical data records that are used by many Earth system science disciplines and applications. We used the proposed method to generate a globally consistent and 1 km daily (mid-daytime and mid-nighttime) MODIS-like LST data from 2003 to 2020 using MODIS LST datasets (MOD11A1 and

MYD11A1). The resulting dataset filled all existing gaps those resulting from elimination of poor-quality data seamlessly with high accuracies based on a cross validation under different rates of missing values for both daytime and nighttime. The average RMSE of gap-filled LST for daytime and nighttime ranges from 1.80 to 2.03 ºC and 1.23 to 1.45 ºC, respectively, when different percentages of the data were excluded. The results show that the missing values in the original MODIS LST were effectively and efficiently filled, and there is no obvious block effect caused by large areas of missing values, especially near the boundary of tiles, which might exist in other seamless LST datasets. The gap-filled global 1 km daily LST dataset can provide better data source for multidisciplinary applications such as urban heat island, air temperature estimation, soil moisture estimation, evapotranspiration, and drought monitoring (Phan and Kappas, 2018). However, it is worth noting that the accuracy of the gap-filled LST can be influenced by the rate of missing values, indicating that uncertainties might increase with the increase of missing values in the original dataset. Moreover, future work can focus on diurnal changes of LST by increasing observations within a day.

Author contributions.

Yuyu Zhou designed the research; Tao Zhang implemented the research and wrote the original manuscript; Yuyu Zhou and Zhengyuan Zhu supervised the research. All co-authors revised the manuscript and contributed to the writing.

Competing interests.

The authors declare that they have no conflict of interest.

Acknowledgements.

This research was supported by the College of Liberal Arts and Science's (LAS) Dean's Emerging Faculty Leaders award at the Iowa State University and the National Science Foundation (2041859).

References

Aguilar-Lome, J., Espinoza-Villar, R., Espinoza, J. C., Rojas-Acuña, J., Willems, B. L. and Leyva-Molina, W. M.: Elevation-

https://doi.org/10.5194/essd-2021-313
Preprint. Discussion started: 12 October 2021

Open Access Earth System Science Data Discussions

dependent warming of land surface temperatures in the Andes assessed using MODIS LST time series (2000–2017), Int. J. Appl. Earth Obs. Geoinf., 77, 119–128, doi:10.1016/j.jag.2018.12.013, 2019.

Anderson, M., Norman, J., Hain, C., Space, M. and Mecikalski, J. R.: Mapping daily evapotranspiration at field to global scales using geostationary and polar orbiting satellite imagery, Hydrol. Earth Syst. Sci. Discuss., 7(4), 5957–5990, doi:10.5194/hessd-7-5957-2010, 2010.

Bai, L., Long, D. and Yan, L.: Estimation of Surface Soil Moisture With Downscaled Land Surface Temperatures Using a Data Fusion Approach for Heterogeneous Agricultural Land, Water Resour. Res., 55, 1105–1128, doi:10.1029/2018WR024162, 2019.

Cheng, J.: 1km seamless land surface temperature dataset of China (2002-2020), edited by N. T. P. D. Center, Natl. Tibet. Plateau Data Cent., doi:10.11888/Meteoro.tpdc.271657, 2021.

Choi, Y. Y. and Suh, M. S.: Development of Himawari-8/Advanced Himawari Imager (AHI) land surface temperature retrieval algorithm, Remote Sens., 10(12), doi:10.3390/rs10122013, 2013.

Crosson, W. L., Al-Hamdan, M. Z., Hemmings, S. N. J. and Wade, G. M.: A daily merged MODIS Aqua-Terra land surface temperature data set for the conterminous United States, Remote Sens. Environ., 119, 315–324, doi:10.1016/j.rse.2011.12.019, 2012.

Duan, S. B., Li, Z. L. and Leng, P.: A framework for the retrieval of all-weather land surface temperature at a high spatial resolution from polar-orbiting thermal infrared and passive microwave data, Remote Sens. Environ., 195, 107–117, doi:10.1016/j.rse.2017.04.008, 2017.

Duguay-Tetzlaff, A., Bento, V. A., Göttsche, F. M., Stöckli, R., Martins, J. P. A., Trigo, I., Olesen, F., Bojanowski, J. S., da Camara, C. and Kunz, H.: Meteosat land surface temperature climate data record: Achievable accuracy and potential uncertainties, Remote Sens., 7(10), 13139–13156, doi:10.3390/rs71013139, 2015.

Fan, X. M., Liu, H. G., Liu, G. H. and Li, S. B.: Reconstruction of MODIS land-surface temperature in a flat terrain and fragmented landscape, Int. J. Remote Sens., 35(23), 7857–7877, doi:10.1080/01431161.2014.978036, 2014.

Ford, T. W. and Quiring, S. M.: Comparison of Contemporary In Situ, Model, and Satellite Remote Sensing Soil Moisture With a Focus on Drought Monitoring, Water Resour. Res., 55, 1565–1582, doi:10.1029/2018WR024039, 2019.

Garcia, D.: Robust smoothing of gridded data in one and higher dimensions with missing values, Comput. Stat. Data Anal., 54(4), 1167–1178, doi:10.1016/j.csda.2009.09.020, 2010.

Garcia, D.: A fast all-in-one method for automated post-processing of PIV data, Exp. Fluids, 50(5), 1247–1259, doi:10.1007/s00348-010-0985-y, 2011.

Gerber, F., De Jong, R., Schaepman, M. E., Schaepman-Strub, G. and Furrer, R.: Predicting Missing Values in Spatio-Temporal Remote Sensing Data, IEEE Trans. Geosci. Remote Sens., 56(5), 2841–2853, doi:10.1109/TGRS.2017.2785240, 2018.

Ghent, D. J., Corlett, G. K., Göttsche, F. M. and Remedios, J. J.: Global Land Surface Temperature From the Along-Track Scanning Radiometers, J. Geophys. Res. Atmos., 122(22), 12,167-12,193, doi:10.1002/2017JD027161, 2017.

Green, P. J. and Silverman, B. W.: Nonparametric regression and generalized linear models: a roughness penalty approach, Crc Press., 1994.

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, doi:10.1016/j.rse.2014.08.013, 2014.

Hong, F., Zhan, W., Göttsche, F.-M., Lai, J., Liu, Z., Hu, L., Fu, P., Huang, F., Li, J., Li, H. and Wu, H.: A simple yet robust framework to estimate accurate daily mean land surface temperature from thermal observations of tandem polar orbiters, Remote

https://doi.org/10.5194/essd-2021-313
Preprint. Discussion started: 12 October 2021

Sens. Environ., 264, 112612, doi:10.1016/j.rse.2021.112612, 2021.

Hulley, G. C., Hook, S. J., Abbott, E., Malakar, N., Islam, T. and Abrams, M.: The ASTER Global Emissivity Dataset (ASTER GED): Mapping Earth's emissivity at 100 meter spatial scale, Geophys. Res. Lett., 42(19), 7966–7976, doi:10.1002/2015GL065564, 2015.

Jiang, G. M. and Liu, R.: Retrieval of sea and land surface temperature from SVISSR/FY-2C/D/E measurements, IEEE Trans. Geosci. Remote Sens., 52(10), 6132–6140, doi:10.1109/TGRS.2013.2295260, 2014.

Ke, L., Ding, X. and Song, C.: Reconstruction of time-series modis lst in central qinghai-tibet plateau using geostatistical approach, IEEE Geosci. Remote Sens. Lett., 10(6), 1602–1606, doi:10.1109/LGRS.2013.2263553, 2013.

Kilibarda, M., Hengl, T., Heuvelink, G. B. M., Gräler, B., Pebesmatadić, E. P. and Bajat, B.: Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution, J. Geophys. Res., 119(5), 2294–2313, doi:10.1002/2013JD020803, 2014.

Li, H., Zhou, Y., Jia, G., Zhao, K. and Dong, J.: Quantifying the response of surface urban heat island to urbanization using the annual temperature cycle model, Geosci. Front., 101141, doi:10.1016/j.gsf.2021.101141, 2021.

Li, X., Zhou, Y., Asrar, G. R., Mao, J., Li, X. and Li, W.: Response of vegetation phenology to urbanization in the conterminous United States, Glob. Chang. Biol., 23(7), 2818–2830, doi:10.1111/gcb.13562, 2017.

Li, X., Zhou, Y., Asrar, G. R. and Zhu, Z.: Creating a seamless 1 km resolution daily land surface temperature dataset for urban and surrounding areas in the conterminous United States, Remote Sens. Environ., 206, 84–97, doi:10.1016/j.rse.2017.12.010, 2018a.

Li, X., Zhou, Y., Asrar, G. R. and Zhu, Z.: Developing a 1 km resolution daily air temperature dataset for urban and surrounding areas in the conterminous United States, Remote Sens. Environ., 215, 74–84, doi:10.1016/j.rse.2018.05.034, 2018b.

Liu, H., Lu, N., Jiang, H., Qin, J. and Yao, L.: Filling Gaps of Monthly Terra / MODIS Daytime Land Surface Temperature Using Discrete Cosine Transform Method, Remote Sens., 12(3), 361, 2020a.

Liu, X., Zhou, Y., Yue, W., Li, X., Liu, Y. and Lu, D.: Spatiotemporal patterns of summer urban heat island in Beijing, China using an improved land surface temperature, J. Clean. Prod., 257, 120529, doi:10.1016/j.jclepro.2020.120529, 2020b.

Long, D., Yan, L., Bai, L., Zhang, C., Li, X., Lei, H., Yang, H., Tian, F., Zeng, C., Meng, X. and Shi, C.: Generation of MODIS-like land surface temperatures under all-weather conditions based on a data fusion approach, Remote Sens. Environ., 246, 111863, doi:10.1016/j.rse.2020.111863, 2020.

Metz, M., Andreo, V. and Neteler, M.: A New Fully Gap-Free Time Series of Land Surface Temperature from MODIS LST Data, Remote Sens., 9(12), 1333, doi:10.3390/rs9121333, 2017.

Parastatidis, D., Mitraka, Z., Chrysoulakis, N. and Abrams, M.: Online global land surface temperature estimation from landsat, Remote Sens., 9(12), 1–16, doi:10.3390/rs9121208, 2017.

Peng, S., Piao, S., Ciais, P., Friedlingstein, P., Ottle, C., Bréon, F. M., Nan, H., Zhou, L. and Myneni, R. B.: Surface urban heat island across 419 global big cities, Environ. Sci. Technol., 46(2), 696–703, doi:10.1021/es2030438, 2012.

Peng, S. S., Piao, S., Zeng, Z., Ciais, P., Zhou, L., Li, L. Z. X., Myneni, R. B., Yin, Y. and Zeng, H.: Afforestation in China cools local land surface temperature, Proc. Natl. Acad. Sci. U. S. A., 111(8), 2915–2919, doi:10.1073/pnas.1315126111, 2014.

Pham, H. T., Kim, S., Marshall, L. and Johnson, F.: Using 3D robust smoothing to fill land surface temperature gaps at the continental scale, Int. J. Appl. Earth Obs. Geoinf., 82, 101879, doi:10.1016/j.jag.2019.05.012, 2019.

Phan, T. N. and Kappas, M.: Application of MODIS land surface temperature data: a systematic literature review and analysis, J. Appl. Remote Sens., 12(04), 1, doi:10.1117/1.jrs.12.041501, 2018.

Roy, D. P., Wulder, M. A., Loveland, T. R., C.E., W., Allen, R. G., Anderson, M. C., Helder, D., Irons, J. R., Johnson, D. M.,

https://doi.org/10.5194/essd-2021-313
Preprint. Discussion started: 12 October 2021

Kennedy, R., Scambos, T. A., Schaaf, C. B., Schott, J. R., Sheng, Y., Vermote, E. F., Belward, A. S., Bindschadler, R., Cohen, W. B., Gao, F., Hipple, J. D., Hostert, P., Huntington, J., Justice, C. O., Kilic, A., Kovalskyy, V., Lee, Z. P., Lymburner, L., Masek, J. G., McCorkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne, R. H. and Zhu, Z.: Landsat-8: Science and product vision for terrestrial global change research, Remote Sens. Environ., 145, 154–172, doi:10.1016/j.rse.2014.02.001, 2014.

Sandeep, P., Obi Reddy, G. P., Jegankumar, R. and Arun Kumar, K. C.: Monitoring of agricultural drought in semi-arid ecosystem of Peninsular India through indices derived from time-series CHIRPS and MODIS datasets, Ecol. Indic., 121, 107033, doi:10.1016/j.ecolind.2020.107033, 2021.

Sheffield, J., Wood, E. F., Pan, M., Beck, H., Coccia, G., Serrat-Capdevila, A. and Verbist, K.: Satellite Remote Sensing for Water Resources Management: Potential for Supporting Sustainable Development in Data-Poor Regions, Water Resour. Res., 54(12), 9724–9758, doi:10.1029/2017WR022437, 2018.

Shen, H., Li, X., Cheng, Q., Zeng, C., Yang, G., Li, H. and Zhang, L.: Missing Information Reconstruction of Remote Sensing Data: A Technical Review, IEEE Geosci. Remote Sens. Mag., 3(3), 61–85, doi:10.1109/MGRS.2015.2441912, 2015.

Sims, D. A., Rahman, A. F., Cordova, V. D., El-Masri, B. Z., Baldocchi, D. D., Bolstad, P. V., Flanagan, L. B., Goldstein, A. H., Hollinger, D. Y., Misson, L., Monson, R. K., Oechel, W. C., Schmid, H. P., Wofsy, S. C. and Xu, L.: A new model of gross primary productivity for North American ecosystems based solely on the enhanced vegetation index and land surface temperature from MODIS, Remote Sens. Environ., 112(4), 1633–1646, doi:10.1016/j.rse.2007.08.004, 2008.

Sulla-Menashe, D. and Friedl, M. A.: User Guide to Collection 6 MODIS Land Cover Dynamics (MCD12Q2) Product, User Guid., 6, 1–18, 2018.

Sun, L., Chen, Z., Gao, F., Anderson, M., Song, L., Wang, L., Hu, B. and Yang, Y.: Reconstructing daily clear-sky land surface temperature for cloudy regions from MODIS data, Comput. Geosci., 105(April), 10–20, doi:10.1016/j.cageo.2017.04.007, 2017.

Tang, J., Di, L., Xiao, J., Lu, D. and Zhou, Y.: Impacts of land use and socioeconomic patterns on urban heat island, Int. J. Remote Sens., 38(11), 3445–3465, doi:10.1080/01431161.2017.1295485, 2017.

Trigo, I. F., Peres, L. F., DaCamara, C. C. and Freitas, S. C.: Thermal land surface emissivity retrieved from SEVIRI/Meteosat, IEEE Trans. Geosci. Remote Sens., 46(2), 307–315, doi:10.1109/TGRS.2007.905197, 2008.

Wan, Z.: Collection-6 MODIS Land Surface Temperature Products Users' Guide, Earth Research Institute, University of California, Santa Barbara., 2013.

Wan, Z.: New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product, Remote Sens. Environ., 140, 36–45, doi:10.1016/j.rse.2013.08.027, 2014.

Weiss, D. J., Atkinson, P. M., Bhatt, S., Mappin, B., Hay, S. I. and Gething, P. W.: An effective approach for gap-filling continental scale remotely sensed time-series, ISPRS J. Photogramm. Remote Sens., 98, 106–118, doi:10.1016/j.isprsjprs.2014.10.001, 2014.

Xu, S. and Cheng, J.: A new land surface temperature fusion strategy based on cumulative distribution function matching and multiresolution Kalman filtering, Remote Sens. Environ., 254, 112256, doi:10.1016/j.rse.2020.112256, 2021.

Xu, Y. and Shen, Y.: Reconstruction of the land surface temperature time series using harmonic analysis, Comput. Geosci., 61, 126–132, doi:10.1016/j.cageo.2013.08.009, 2013.

Yang, J. S.: Estimation of Land Surface Temperature Using Spatial Interpolation and Satellite-Derived Surface Emissivity, J. Environ. Informatics, 4(1), 40–47, doi:10.3808/jei.200400035, 2004.

Yu, Y., Tarpley, D., Privette, J. L., Goldberg, M. D., Rama Varma Raja, M. K., Vinnikov, K. Y. and Xu, H.: Developing algorithm for operational GOES-R land surface temperature product, IEEE Trans. Geosci. Remote Sens., 47(3), 936–951, doi:10.1109/TGRS.2008.2006180, 2009.

Yue, W., Liu, X., Zhou, Y. and Liu, Y.: Impacts of urban configuration on urban heat island: An empirical study in China mega-cities, Sci. Total Environ., 671, 1036–1046, doi:10.1016/j.scitotenv.2019.03.421, 2019.

Zhan, W., Hong, F. and Chen, Y.: Global daily average 1km resolution land surface temperature dataset from 2003 to 2019[DB/OL], Chinese Ecosyst. Sci. Data Cent., doi:10.12199/nesdc.ecodb.2016YFA0600200.01.005, 2021.

Zhang, L., Jiao, W., Zhang, H., Huang, C. and Tong, Q.: Studying drought phenomena in the Continental United States in 2011 and 2012 using various drought indices, Remote Sens. Environ., 190, 96–106, doi:10.1016/j.rse.2016.12.010, 2017.

Zhang, T., Zhou, Y., Zhu, Z., Li, X. and Asrar, G. R.: A global seamless 1 km resolution daily land surface temperature dataset (2003 – 2020), , doi:10.25380/iastate.c.5078492, 2021a.

Zhang, X., Pang, J. and Li, L.: Estimation of Land Surface temperature under cloudy skies using combined diurnal solar radiation and surface temperature evolution, Remote Sens., 7(1), 905–921, doi:10.3390/rs70100905, 2015.

Zhang, X., Zhou, J., Göttsche, F. M., Zhan, W., Liu, S. and Cao, R.: A method based on temporal component decomposition for estimating 1-km all-weather land surface temperature by merging satellite thermal infrared and passive microwave observations (IEEE Transactions on Geoscience and Remote Sensing (2019) DOI: 10.1109/TG, IEEE Trans. Geosci. Remote Sens., 57(8), 6254, doi:10.1109/TGRS.2019.2899770, 2019.

Zhang, X., Zhou, J., Liang, S., Chai, L., Wang, D. and Liu, J.: Estimation of 1-km all-weather remotely sensed land surface temperature based on reconstructed spatial-seamless satellite passive microwave brightness temperature and thermal infrared data, ISPRS J. Photogramm. Remote Sens., 167, 321–344, doi:10.1016/j.isprsjprs.2020.07.014, 2020.

Zhang, X., Zhou, J., Liang, S. and Wang, D.: A practical reanalysis data and thermal infrared remote sensing data merging (RTM) method for reconstruction of a 1-km all-weather land surface temperature, Remote Sens. Environ., 260, 112437, doi:10.1016/j.rse.2021.112437, 2021b.

Zhang, X., Zhou, J., Tang, W., Ding, L., Ma, J. and Zhang, X.: Daily 1-km all-weather land surface temperature dataset for the Chinese landmass and its surrounding areas (TRIMS LST; 2000-2020), edited by N. T. P. D. Center, Natl. Tibet. Plateau Data Cent., doi:10.11888/Meteoro.tpdc.271252, 2021c.

Zhao, B., Mao, K., Cai, Y., Shi, J., Li, Z., Qin, Z., Meng, X., Shen, X. and Guo, Z.: A combined Terra and Aqua MODIS land surface temperature and meteorological station data product for China from 2003 to 2017, Earth Syst. Sci. Data, 12(4), 2555–2577, doi:10.5194/essd-12-2555-2020, 2020a.

Zhao, J., Yu, L., Xu, Y., Li, X., Zhou, Y., Peng, D., Liu, H., Huang, X., Zhou, Z., Wang, D., Ren, C. and Gong, P.: Exploring difference in land surface temperature between the city centres and urban expansion areas of China's major cities, Int. J. Remote Sens., 41(23), 8963–8983, doi:10.1080/01431161.2020.1797216, 2020b.

Zhou, D., Xiao, J., Bonafoni, S., Berger, C., Deilami, K., Zhou, Y., Frolking, S., Yao, R., Qiao, Z. and Sobrino, J. A.: Satellite remote sensing of surface urban heat islands: Progress, challenges, and perspectives, Remote Sens., 11(1), 1–36, doi:10.3390/rs11010048, 2019.

Zhou, J., Zhang, X., Zhan, W., Göttsche, F. M., Liu, S., Olesen, F. S., Hu, W. and Dai, F.: A Thermal Sampling Depth Correction Method for Land Surface Temperature Estimation From Satellite Passive Microwave Observation Over Barren Land, IEEE Trans. Geosci. Remote Sens., 55(8), 4743–4756, doi:10.1109/TGRS.2017.2698828, 2017.

Zhou, Y., Smith, S. J., Elvidge, C. D., Zhao, K., Thomson, A. and Imhoff, M.: A cluster-based method to map urban area from DMSP/OLS nightlights, Remote Sens. Environ., 147, 173–185, doi:10.1016/j.rse.2014.03.004, 2014a.

Zhou, Y., Clarke, L., Eom, J., Kyle, P., Patel, P., Kim, S. H., Dirks, J., Jensen, E., Liu, Y., Rice, J., Schmidt, L. and Seiple, T.: Modeling the effect of climate change on U.S. state-level buildings energy demands in an integrated assessment framework, Appl. Energy, 113, 1077–1088, doi:10.1016/j.apenergy.2013.08.034, 2014b.

https://doi.org/10.5194/essd-2021-313
Preprint. Discussion started: 12 October 2021

Zhou, Y., Li, X., Asrar, G. R., Smith, S. J. and Imhoff, M.: A global record of annual urban dynamics (1992–2013) from nighttime lights, Remote Sens. Environ., 219, 206–220, doi:10.1016/j.rse.2018.10.015, 2018.