A combined Terra and Aqua MODIS land surface temperature and meteorological station data product for China from 2003–2017

Bing Zhao[1,2,†], Kebiao Mao[1,3,4,†], YuLin Cai[2], Jiancheng Shi[3], Zhaoliang Li[1], Zhihao Qin[1], and Xiangjin Meng[5]

[1]Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing, 100081, China
[2]Geomatics College, Shandong University of Science and Technology, Qingdao, 266590, China
[3]State Key Laboratory of Remote Sensing Science, Jointly Sponsored by the Institute of Remote Sensing Applications of Chinese Academy of Sciences and Beijing Normal University, Beijing, 100101, China
[4]School of Physics and Electronic-Engineerring, Ningxia University, Ningchuan 750021, China
[5]School of Surveying and Geo-Informatics, Shandong Jianzhu University, Jinan, 250100, China

Correspondence to: Kebiao Mao (maokebiao@caas.cn)
†These authors contributed equally to this work and should be considered co-first authors.

Abstract. Land surface temperature (LST) is a key variable for high temperature and drought monitoring and climate and ecological environment research. Due to the sparse distribution of ground observation stations, thermal infrared remote sensing technology has become an important means of quickly obtaining ground temperatures over large areas. However, there are many missing and low-quality values in satellite-based LST data because clouds cover more than 60 % of the global surface every day. This article presents a unique LST dataset for China from 2003-2017 that filters and removes missing values and low-quality LST pixel values low-quality by clouds from raw LST images and retrieves real surface temperatures in areas with cloud coverage via a reconstruction model. We specifically describe the reconstruction model, which uses a combination of MODIS daily data, monthly data and meteorological station data to reconstruct the true LST in areas with cloud coverage, and the data performance is then further improved by establishing a regression analysis model. The validation indicates that the new LST dataset is highly consistent with in situ observations. For the six natural subregions with different climatic conditions in China, the RMSE ranges from 1.24 °C to 1.58 °C, the MAE varies from 1.23 °C to 1.37 °C, and the R^2 ranges from 0.93 to 0.99. The new dataset adequately captures the spatiotemporal variations in LST at annual, seasonal and monthly scales. From 2003-2017, the overall annual mean LST in China showed a weak increase. Moreover, the positive trend was remarkably unevenly distributed across China. The most significant warming occurred in the central and western areas of the Inner Mongolia Plateau in the Northwest Region (slope >0.10, R >0.71, P < 0.05), and a strong negative trend was observed in some parts of the Northeast Region and South China Region. Seasonally, there was significant warming in western China in winter, which was most pronounced in December. The reconstructed dataset exhibited significant improvements and can be used for the spatiotemporal evaluation of LST in high temperature and drought monitoring studies. The data are available through Zenodo at https://doi.org/10.5281/zenodo.3528024 (Zhao et al., 2019).

1 Introduction

Land surface temperature (LST), which is controlled by land–atmosphere interactions and energy fluxes, is an essential parameter for the physical processes of the surface energy balance and water cycle at regional and global scales (Li et al., 2013; Wan et al., 2014; Benali et al., 2012). Accurate surface temperature datasets are not only required for high temperature and drought research over various spatial scales but also important elements for improving global hydrological and climate prediction models. In particular, temperature changes directly influence glacier reserves and water storage on the Qinghai-Tibet Plateau (Tibetan Plateau), which is known as the "World Water Tower". In turn, these changes directly affect the living conditions of nearly 40 % of the world's population (Xu et al., 2008). Therefore, LST research at regional and global scales is crucial for further improving and refining global hydroclimatic and climate prediction models. LST data, which is conventionally measured by meteorological stations or ground surveys, have the advantages of high reliability and long time series. However, the meteorological station data collected as point data with very limited spatial coverage are often sparsely and/or irregularly distributed, especially in remote and rugged areas (Neteler, 2010; Hansen et al., 2010; Gao et al., 2018). To obtain spatially continuous LST data, various geostatistical interpolation approaches are commonly applied to achieve spatialization, such as kriging interpolation and spline function methods. However, the smoothed spatial pattern obtained after interpolation may suffer from low reliability because the ground station density is far from sufficient in most regions.

In contrast to ground-based observations with their limited availability and discrete spatial information, images captured by satellite thermal infrared instruments have become reliable alternative data sources with the advantages of detailed spatialized surfaces and near real-time data access (Vancutsem et al., 2010). For instance, for the study of uniform continuous surface temperatures over large-scale areas, such as at regional and global scales, satellite remote sensing is the only efficient and feasible method (Xu et al., 2013). Satellite remote sensing obtains global LSTs based on a variety of mature retrieval algorithms that have been proposed since the 1970s for use with data from thermal infrared channels (McMillin, 1975). Due to its optimal temporal and spatial resolution throughout the world, the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor has become an excellent data source for satellite-derived LST data, and the MODIS LST values are widely used in regional and global climate change and environmental monitoring models (Tatem et al., 2004; Wan et al., 2014). However, satellite-derived LST data are frequently and strongly affected by cloud cover, causing many data gaps and large amounts of low-quality values from undetected cloud-low-quality pixels. Cloud cover is frequent, and the locations of cloud cover are often uncertain. On average, at any one time, approximately 65 % of the global surface is obscured by clouds, leading directly to missing values over large, unevenly distributed areas in an image (Crosson et al., 2012; Mao et al., 2019). Although the integrity of the data has been greatly improved, the 8-day and monthly synthetic data still contain a number of low-quality pixels and thus contain an insufficient quantity of daily LST pixels. Cloud cover and other factors, which causes extensive amounts of missing and low-quality pixels, significantly reduces the proportion of usable LST data and poses a problem to further applications. Thus, reconstruction of these missing and low-quality low-quality LST pixels low-quality is necessary for satellite-derived LST applications.

Two categories of methods have commonly been applied to reconstruct cloud-low-quality pixels from satellite-derived data in previous studies. The first category includes methods that directly reconstruct missing and low-quality values using neighboring information with high similarity over temporal and spatial scales. Most temporal interpolation methods reconstruct missing and low-quality LST values based on the periodic behavior of data, such as time series harmonic analysis (HANTS), S-G filtering, and Fourier transform (Xu and Shen, 2013; Na et al., 2014; Scharlemann et al., 2008). Crosson (2012) used another temporal interpolation method to reconstruct the LST data from the Aqua platform (afternoon overpass) using LST data from the Terra platform (morning overpass). Regarding spatial interpolation methods, previous methods have focused on geostatistical interpolation, including kriging interpolation, spline interpolation and their variants. Some researchers have also carried out other attempts; for example, Yu et al. (2015) introduced a method using a transfer function with the most similar pixels to estimate invalid pixels. These methods, which estimate missing MODIS LST data using only LST data, take advantage of the similarity and interdependence of the available temporal/spatial attributes of neighboring pixels. To some extent, these methods have the advantage of simplicity and reliability. However, this category of methods are often not as reliable as expected especially in complex topographical regions and areas with many missing data, because it cannot obtain enough information for reconstruction. The second category of methods solves these data gap problems by establishing correlation models for cloud-low-quality pixels and corresponding auxiliary data pixels. Neteler (2010) used a digital elevation model (DEM) as an auxiliary predictor to reconstruct MODIS LST data from nine years of data on temperature gradients, which achieved reliable results in mountainous regions. Ke et al. (2013) built a regression model that included many auxiliary predictors—latitude, longitude, elevation, and the normalized difference vegetation index (NDVI)—to estimate 8-day composite LST products. Fan et al. (2014) used multiple auxiliary maps, including land cover, NDVI, and MODIS band 7, to reconstruct LST data in flat and relatively fragmented landscape regions. Other similar algorithms have drawn support by employing many factors that affect LST, including elevation, NDVI, solar radiation, land cover, distance from the ocean, slope and aspect. Considering the complexity of the terrain and the nonuniformity of the spatial distribution of large-scale LST patterns, a reconstruction model with auxiliary data that provides rich information for missing pixels can improve the accuracy of the interpolation result.

The above studies greatly improve the usability of MODIS LST data and further add value to long-term LST trend analyses. However, despite the use of various techniques to reconstruct the LST value, existing techniques focus on the retrieval of the LST value under the assumption of clear-sky conditions instead of cloudy conditions, which cannot fulfill the need to obtain the real situation at the land surface. To address this issue, some scholars have also used microwave temperature brightness (TB) data, which are mostly derived from high-frequency channels ($\geqslant$85 GHz), to obtain the real LSTs under clouds (André et al., 2015; Prigent et al., 2016). Microwave remote sensing is capable of penetrating clouds and can obtain useful radiation information for the retrieval of LST under clouds. However, the physical mechanisms of the current microwave LST retrieval models are not very mature, and the models have low resolution (Mao et al. 2007, 2018). Moreover, due to the difference in the surface properties of the land, the depth of the radiation signal detected by the microwave will differ at different locations,

and it will deviate from the retrieval results of thermal infrared remote sensing when used to estimate LST values. Thus, new reconstruction methods for LST data need to be proposed to compensate for this deficiency.

On this premise, China is used as an example due to its large coverage area, heterogeneous landscape and complex climatic conditions. This paper presents a new long-term spatially and temporally continuous MODIS LST dataset for China from 2003 to 2017 that filters out missing and low-quality pixels and reconstructs them based on multisource data. We describe a data reconstruction process that is fully integrated with the benefits of the high reliability of surface observations, consistency and high accuracy of daily valid pixels and spatial autocorrelation of similar pixels. The process compensates for the insufficiency of reconstructing pixels under clear-sky conditions instead of under clouds in previous studies. The validation using data from the China Meteorological Administration observations indicates the robustness of the LST data after interpolation. The dataset is ultimately used to capture the annual, seasonal and monthly spatiotemporal variations in the LST in six natural subregions in China. It is envisioned that this dataset will help capture changes in surface temperature and will be useful for studies on high temperatures, drought and food security.

2 Study area

Insert Figure 1 here

Figure 1: The study area is divided into six natural subregions (I, II, III, IV, V, and VI), and the spatial distribution of the meteorological stations in the subregions is shown. The red circles mark the key areas where the temperature has changed significantly, and meteorological station from Subset (2) located these areas are used to validate the accuracy of the new LST dataset (a, b, c, d, e, and f).

To explore the temporal and spatial characteristics of China's LSTs, the study area is divided into six natural subregions based on China's three major geographical divisions: climatic conditions, landform types and tectonic movement characteristics. The eastern region is topographically characterized by plains and low mountains. This region has a variety of monsoon climate zones, which, from south to north, include tropical, subtropical and temperate monsoon climate zones. Therefore, we divide the eastern region into the following four regions, as shown in Fig. 1. (I) The Northeast Region, which mainly covers the area to the east of Daxing'anling. This region has a temperate monsoon climate with an average annual precipitation of 400~1000 mm, and rain and heat are prevalent in the same period. (II) The North China Region lies to the south of the Inner Mongolia Plateau, to the north of the Qinling Mountains and Huaihe River, and to the east of the Qinghai-Tibet Plateau. The region is dominated by a temperate monsoon climate and a temperate continental climate with four distinct seasons. This area is characterized by flat plains and plateau terrain. (III) The Central-Southwest China Region extends from the eastern part of the Qinghai-Tibet Plateau to the western parts of the East China Sea and South China Sea, south to the Huaihe River - Qinling Mountains, and north to the area where the daily average temperature is greater than or equal to 10 °C. The accumulated temperature in this region is 7500 °C. This region is commonly dominated by a subtropical monsoon climate. (IV) The South China Region is located in the southernmost part of China and is characterized by a tropical and subtropical

monsoon climate with hot and humid conditions. The area has abundant rainfall, and the average annual precipitation is approximately 1900 mm.

The western region is divided into 2 natural subregions. (V) The Northwest Region includes the northern Qilian Mountain-Altun Mountains-Kunlun Mountains, the Inner Mongolia Plateau and the western part of the Greater Khingan Range. This region is located in the continental interior and features complex terrain conditions, dominated by plateau basins and mountainous areas. This region has a tropical dry continental climate with rare rainfall. Consequently, this area features large areas of barren land, with a desertified land area of 2.183 million km^2, accounting for 81.6 % of China's desertified land area (Deng, 2018). Moreover, the Taklimakan Desert in this region is one of the 10 largest deserts in the world. (VI) The Qinghai-Tibet Plateau Region is mainly located on the Qinghai-Tibet Plateau, which is the highest-elevation plateau in the world. This region is mainly described as having an alpine plateau climate, with relatively high temperatures and an extensive grassland meadow area.

3 Data and methods

3.1 MODIS data

MODIS is a key sensor of the Earth Observing System (EOS) program that provides a unified grid product with global coverage of the land, atmosphere and oceans. MODIS covers 36 spectral bands in the visible, near-infrared and thermal infrared ranges (from 0.4 to 14.4 um), so it is extensively used to study global marine, atmospheric, and terrestrial phenomena (Wan et al., 1997). The MODIS instruments are aboard two NASA satellites, Terra and Aqua, which were launched in December 1999 and May 2002, respectively. As both the Aqua and Terra satellites are near-polar orbit satellites with a flying height of approximately 705 km in sun-synchronous orbit, they provide data with a temporal resolution of twice daily. The Terra satellite passes through the equator from north to south at approximately 10:30 am and 10:30 pm local solar time and is called the morning star. The Aqua satellite passes through the equator in the opposite direction from south to north at approximately 1:30 am and 1:30 pm and is called the afternoon satellite (Christelle and Ceccato, 2010). The two satellites collect repeat observations every 1-2 days and transmit observation data to the ground in real time.

MODIS LST data are retrieved with two algorithms: the generalized split-window algorithm (Wan and Dozier, 1996; Wan et al., 2002) and the day/night algorithm (Wan and Li, 1997). The split-window algorithm is advantageous for removing atmospheric effects because the signal difference between the adjacent thermal and middle infrared channels (channel 31 with a wavelength of 10.78–11.28 μm and channel 32 with a wavelength of 11.77–12.27 μm) is caused by the differential absorption of radiation in the atmosphere (Wan et al., 2002). The latest LST V006 version data used in this article is obtained by another important algorithm: the day/night algorithm. Due to the complexity of surface reflectance and atmospheric conditions, multichannel data are utilized in day/night image pairs to derive the LST, including radiation calibration data, atmospheric temperature and water vapor data, cloud masks and geolocation information and L1B data in seven thermal infrared bands

(bands 20, 22, 23, 29, 31, 32, 33) from MOD07_L2 (Wan, 2007). Then, these day/night pairs of MODIS data are used to construct 14 nonlinear equations to retrieve the surface emissivity and surface temperature based on a physics-based day/night LST model from the MODIS 1B data without high accuracy atmospheric temperature and water vapor profiles (Wan and Li, 1997).

The day/night LST algorithm used in the LST V006 version products exhibits great advantages in retrieving LST: it not only optimizes atmospheric temperature and water vapor profile parameters for optimal retrieval but also does not require complete reversal of surface variables and atmospheric profiles (Ma et al., 2000, 2002). A comprehensive sensitivity and error analysis was performed for the day/night algorithm, which showed that the accuracy was very high, with an error of 1–2 K in most cases (0.4-0.5 K standard deviation over various surface temperatures for mid-latitude summer conditions) (Wan and Li, 1997, Wang and Liang, 2009; Wang et al., 2007).

Here, two MODIS products were chosen for research: daily LST data (MOD11C1 and MYD11C1) and monthly LST data (MOD11C3 and MYD11C3), which were available for the period from 2003–2017. The datasets include daytime and nighttime surface temperature data provided by NASA. These data are the new collection 6 series data provided in 2017, which has been fixed and substantially improved compared to the collection 5 data used in many previous studies. In collection 6 data, the identified cloud-low-quality LST pixels were removed from the MODIS Level 2 and Level 3 products, and the classification-based surface emissivity values were adjusted (Wan. 2014). Both datasets provide the global LSTs generated by the day/night algorithm with a spatial resolution of 0.05°×0.05° (approximately 5600 m at the equator), which is provided in an equal-area integerized sinusoidal projected coordinate system. The composited 8-day (MOD11C2/MYD11C2) and monthly (MOD11C3/MYD11C3) data are deduced from daily global data (MOD11C1/ MYD11C1) without cloud contamination. The global LST product MOD11C1/MYD11C1 (V006) was created by assembling daily MOD11B1/MYD11B1 tiles and resampling from 5600 m spatial resolution to a resolution of 7200 columns and 3600 rows of the latitude/longitude grid.

The preprocessing of the MODIS data mainly includes extraction of daytime and nighttime LST and the corresponding quality assessment (QA) data from 17 scientific datasets, projection conversion (reprojection of raw data from a sinusoidal projection (SIN) to an equal-area conical projection (Albers), WGS84 coordinate system), conversion of the data format (conversion from HDF to Geo-tiff), and clipping (selection of geographic subsets according to latitude and longitude position to improve the efficiency of the model). The brightness temperature is converted to real surface temperature by coefficient and offset correction.

3.2 Supplementary data

LST records at the hourly intervals from 2399 meteorological ground stations in China from 2003-2017 were used in this study, and they were provided and subjected to strict quality control and evaluation by the China Meteorological Administration (CMA). Meteorological station data were randomly divided into two completely independent subsets by the jackknife method (Benali et al., 2012). Subset (1): the number of ground stations used for the reconstruction process was 1919, accounting for

80 % of the total number of ground stations. Subset (2): the number of sites used for verification was 480, accounting for 20 % of the total. Then, the data used for the reconstruction process for subset (1) were created by extracting meteorological station LST data at the overpass times of Terra and Aqua (01:30, 10:30, 13:30, and 22:30). For the verification process, six key areas where warming/negative trends were the most significant (i.e., shown in the red ellipses a-f in Fig. 1 and Table 1) were selected for detailed verification to better reflect the accuracy of the LST data. All meteorological ground station data were tested for temporal and spatial consistency, which included identifying and rejecting extreme values and outliers. It is worth noting that the key areas marked by red circles contain site data from subset (1) and subset (2). Generally, there are more stations in the red circle than the sites used for verification in Table 1, especially in the Eastern China where there are a large number of stations.

Table 1 Basic information for some of the meteorological stations in key zones

Insert Table 1 here

Elevation data with 1 km resolution are available from the NASA Space Shuttle Radar Terrain Mission (SRTM) V4.1 for reconstruction of cloud-low-quality data.

3.3 LST data restoration method

For satellite-derived LST data, due to the extensive and random distribution of missing and unreliable values caused by cloud contamination in the images. It is difficult to reconstruct the cloudless LST dataset on daily scale, and it is even more difficult to reconstruct the dataset that can reflect the LST of the ground under the cloud cover. Therefore, we create a reconstruction model that combines meteorological station data and daily and monthly MODIS LST data to reconstruct a high-precision monthly dataset that can reflect the true LST under cloud coverage. The reconstruction model effectively preserves the highly accurate pixels in the original daily data, reconstructs only the low-quality daily data, and finally, replaces low-quality pixels with the composite average pixel value in the monthly data. To better describe the data processing, Fig. 2 shows a summary flowchart for the reconstruction of MODIS monthly LST data. The reconstruction model we propose is divided into two general steps: LST pixel filtering and LST data restoration. Low-quality pixel values were first identified and set to missing values for all input monthly LST images based on pixel quality filtering(see section 3.3.1 for details）. Both missing pixels and low-quality pixels are considered invalid pixels that need to be reconstructed. low-qualityFor each invalid pixel in the monthly images, we first determined the invalid pixels in daily LST images at the same location at the corresponding time. And then we reconstructed these invalid daily pixels. The reconstruction process for the invalid daily pixels is divided into three substeps low-quality(see section 3.3.2 for details）: 1) in situ LST observations at the same location, as judged from the longitude and latitude information, were used to fill the invalid pixels at the same location; 2) the geographically weighted regression (GWR) method was employed to interpolate invalid pixels based on similar pixels from multiple sources; and 3) the remaining invalid pixels were reconstructed based on regression of the elevation temperature gradient. Finally, the averages of the available daily LST pixels were calculated and filled using the corresponding monthly pixels.

Insert Figure 2 here

Figure 2: (a)The summary flowchart for reconstructing MODIS monthly LST data, (b)the detailed flowchart for reconstructing reconstruct missing daily pixels in (a).

3.3.1 Filtering of MODIS LST

MODIS LST data are retrieved from thermal infrared bands in clear-sky conditions and contain many missing values and low-quality values caused by clouds and other atmospheric disturbances. Generally, the cold top surface of a thin or subpixel cloud is difficult to detect, and the LST retrieved under such conditions usually leads to an underestimation of the temperature values in the MODIS LST data (Neteler, 2010; Markus et al., 2010; Jin and Dickinson, 2010; Benali et al., 2012). Moreover, other factors can also contaminate the observation signal and cause the data to be unavailable, such as atmospheric disturbance, observation geometry and instrumental problems (Wan, 2014). In general, the abnormally low values in LST maps caused by undetected thin clouds, together with other low-quality values, need to be identified and filtered because these values greatly reduce the accuracy of the LST data.

Cloud cover is extensive and inevitable in daily weather conditions. Statistical calculations were performed and showed that missing values for daily data reached approximately 50 % for Terra and Aqua satellites. Figure 3 shows an example representing the distribution of valid pixel values in the daytime for winter and summer. The coverage of pixels with missing data in the study area at 10:30 am during the daytime on January 1, 2017, and July 1, 2017, for the Terra platform reached 44.9 % and 51.7 %, respectively. The spatial gaps in the daily data are characterized by an arbitrary distribution and generally large aggregations. In fact, the emergence of a large number of missing values every day makes it difficult to reconstruct a high-precision daily LST images that represents real values under clouds using current techniques due to such a paucity of information, especially for areas with complex climates.

However, the random occurrence of cloud-covered areas has a much smaller impact on monthly composite products, which makes these products a reliable source for building a high-precision monthly LST dataset. Compared with daily and 8-day composite data, spatiotemporal integrity and consistency have been greatly improved in monthly composite LST data. However, for many regions, the lack of data or quality degradation caused by clouds is still common even in monthly composite data（Fig. 4）. It is necessary to identify and reconstruct low-quality pixels, which seriously affect the use and analysis of monthly LST data. A reliable method for removing low-quality pixels is implemented using the data quality control information for MODIS LST data. The data quality control information is statistically calculated and stored in the corresponding QA layer and is represented by an 8-bit unsigned integer and can be found in the original MODIS LST HDF files. Therefore, we use the quality control labels for daily and monthly files as mask layers to identify low-quality pixels to ensure the quality of the LST data. Finally, pixels with QA layer labels of “the average LST error <=1 K”, “LST produced, good quality” and “the average emissivity error <=0.01” are considered to be high-quality data, and the remaining pixels are low-quality pixels and are set to missing values. Finally, we reconstructed all the invalid pixels in monthly LST data. Quality

information is almost indicative; thus, sufficient information is provided for the filtering of low-quality pixels (Benali et al., 2012). A summary flowchart of the process used to construct the LST data model is schematically illustrated in Fig. 2.

The spatial distribution pattern of invalid Terra LST data after filtering by the QA layer is shown in Fig. 4. The low-quality pixel coverage rates for January and July 2017 were 23.45 % and 19.68 %, respectively. There are more missing values in winter than in summer in the northeastern region, which may be affected by the confusion resulting from large areas of snow cover and clouds in the winter. However, the missing values are mainly concentrated in southern China in summer, which is closely related to the increased cloud cover in the hot summers in South China.

Insert Figure 3 here

Figure 3: Spatial distribution of valid data for daily MODIS LST data from Terra during the daytime on (a) January 1, 2017, and (b) July 1, 2017. Areas of missing data are blank.

Insert Figure 4 here

Figure 4: Spatial distribution of valid data after pixel filtering for monthly MODIS LST data from Terra during the daytime on (a) January and (b) July. Areas of invalid data are blank.

3.3.2 LST data restoration

Missing values caused by cloud coverage have always been an important limitation of using LST from thermal infrared (TIR) data in a wide range of applications. For daily LST data, although many attempts have been made to reconstruct the LST data under clear-sky conditions, the widespread missing values make it difficult to reconstruct high-precision daily LST pixels under clouds using the limited available information. To obtain a high-precision LST dataset that can retrieve the true temperature of the land surface under cloud cover instead of clear skies, we adopted another strategy.

Given that monthly LST data are composited from the corresponding daily data, an insufficient amount of valid daily data and low availability for some pixels will lead to quality degradation in the monthly data. The contributions of multiple valid daily pixels $\bar{P}_i$, despite their good precision, are rejected along with the final low-quality values in monthly pixels. Thus, considering the inheritance of these high-quality data, we believe that these valid daily pixels $\bar{P}_i$ can help to reconstruct high-precision estimates for LST pixels under cloudy conditions. Therefore, we uses a combination of MODIS daily data, monthly data and meteorological station data to reconstruct the true LST in areas with cloud coverage.

In the reconstruction model, we first filter the monthly image, and the locations of the cloud-low-quality pixels (i.e., the missing and low-quality monthly pixels) are determined. Then, we filter all the daily pixels from the month in which the cloud-low-quality pixels occurred. The valid pixels $\bar{P}_i$ in the daily data are retained, the low-quality daily data are reconstructed, and the low-quality pixels in the monthly data are replaced with the corresponding averaged pixel values from the daily data (Fig. 2). Missing daily pixel is defined as the target pixels T_t, the image contains the target pixel T_t is target image. The reconstruction process for the the target pixels T_t is as follows.

During the daytime, the real ground LST values in pixels obscured by clouds are usually lower than the values in the adjacent unaffected pixels, and at night it is the opposite. Many influencing factors are considered for data reconstruction. The ground meteorological station values observed at the same time as the reconstruction result are the most reliable among all influencing factors, such as NDVI, elevation, longitude and latitude. Therefore, the latitude and longitude information of the invalid pixels was first used to search for the ground stations in the same location. Invalidt pixels were filled using the values from valid ground-based LST data at the same location at the same time, and these filled pixels were marked. Then, for the invalid pixels without ground meteorological station data, we used a combination of two strategies to reconstruct the missing LST data to improve the accuracy of the result. The first strategy identified the most similar pixels by using adaptive thresholds and reconstructed them by using a GWR method.

GWR is a common and reliable method for estimating missing pixels, which quantifies the contribution of each similar pixel to contaminated pixels. This method assumes that similar pixels that are spatially adjacent to the target pixel are close in the spectrum and should be given more weight. Due to the high temporal variability of thermal radiation emitted from the land surface and atmospheric state parameters, satellite sensors that measure the thermal radiation energy from different phase images at the same locations often produce different values even when the same thermal infrared sensor is used. Some of the most common regular changes in surface features, such as the vegetation spectrum changes due to seasonal changes, can be predicted using auxiliary information. However, multiple unpredictable changes that cause abrupt transformation in the thermal energy of infrared radiation are difficult to predict, such as wheat harvesting . Therefore, for the target pixel i and a valid pixel j in the same image, the three images that are temporally closest at the same overpass time are determined as reference images, and a adaptive threshold is then used in the reference image to determine whether the pixels corresponding to pixels i and j are similar. If two pixels are judged to be similar in more than one image, the valid pixel j is identified as a similar pixel for the target pixel i.

Because the factors that influence surface temperature (vegetation cover, sun zenith angle, microrelief, etc.) vary greatly among different regions and seasons, the differences of adjacent pixels in different areas may also vary greatly. Thus, there will be large deviations in the similar pixel decision criteria if a fixed similarity threshold is used. Here, we use an adaptive threshold $\varphi^{\tau}$to determine similar pixels for each invalid pixel (Eq. 3). The adaptive threshold φ^{τ}, calculated from the standard deviation, indicates the local area smoothness. Local area is a certain size area centered on similar pixel, which is located in the three reference images. The closer the pixel is, the more similar the environment is, so the smoother the local area will be. For example, the jth valid pixel in the target image is determined to be a similar pixel of the target pixel i only when the relationship described in Eq. (2) is satisfied in the reference image τ. Simultaneously, similar pixels were determined based on all valid pixels in the image rather than a sliding window because missing values are often arbitrarily clustered in a large area rather than scattered.

$$|P_s^{\tau} - P_t^{\tau}| \le \varphi^{\tau} \ , \tag{1}$$

$$\varphi^{\tau} = \sqrt{\sum_{i=1}^{n}(P_s^{\tau} - \varepsilon)^2}, \tag{2}$$

where P_s^τ and P_t^τ are the values of pixels corresponding to the position of the similar pixel and the target pixel in the reference image, respectively. φ^τ is the threshold used to determine similar pixels. ε is the mean value of all pixels in local area. τ is the reference image (value=1, 2, 3). Here, we set the range of the local area to 5 pixels by 5 pixels centered on the target pixel (Zeng et al., 2013). In this paper, the number of similar pixels of the target pixel in the target image should be greater than 4 to apply the GWR method to reduce the error due to an insufficient number of similar pixels.

After determining similar pixels, the LST values of the pixels low-quality by clouds are determined through GWR. In theory, LST data from meteorological stations are the most reliable record, even in the case of thick cloud coverage. Thus, similar pixels obtained from ground stations are the most representative and can better reflect the LST under clouds than under clear-sky conditions. In the process of reconstructing missing pixels, The regression weight coefficient of a similar pixel is determined by its Euclidean distance from the target pixel. In addition， we assign a related weight multiplier to the marked ground station data based on the GWR. After selecting some of the marked pixels as experiments, it was found that the target pixels could be more accurately estimated when the relative multiweight values of the ground stations were set to 3 in this paper. Therefore, the weighting coefficients of similar pixels are determined by Eqs. (5) and (6).

$$D = \sqrt{(x - x_t)^2 + (y - y_t)^2}, \tag{3}$$

$$W_i = \frac{\frac{M_c}{D_i}}{\sum_{i=1}^{m}\frac{M_c}{D_i}+\sum_{j=1}^{n}\frac{M_g}{D_j}}, \tag{4}$$

$$W_j = \frac{\frac{M_g}{D_j}}{\sum_{i=1}^{m}\frac{M_c}{D_i}+\sum_{j=1}^{n}\frac{M_g}{D_j}}, \tag{5}$$

where D represents the Euclidean distance from the similar pixel (i， j) to the target pixel t, x, y, x_t and y_t describe the locations of the similar pixel and target pixel. i and j represent similar pixels used to estimate the low-quality LSTs, i is a pixel that is not low-quality by clouds, and j is a pixel assigned by the ground station. W_i and W_j describe the weight that similar pixels i and j contribute to the target pixels, respectively. m is the number of similar pixels that are not low-quality by clouds, and n is the number of similar pixels that are assigned by ground stations. M_c and M_g represent the weight coefficients of clear sky pixels and ground station assignment pixels, respectively. In this paper, M_c and M_g are set at 1 and 3, respectively.

Therefore, the GWR method can be represented as follows

$$T_t = \sum_{i=1}^{m} W_i \cdot T_i + \sum_{j=m+1}^{n} W_j \cdot T_j \ , \tag{6}$$

where T_t is the reconstructed LST value of a target pixel, T_i and T_j represent LST values for the similar pixel i and j, the sum of W_i and W_j values is 1.

For another strategy, the elevation temperature gradient regression method was used to reconstruct the remaining low-quality pixels that did not have enough similar pixels. In general, the elevation factor has a particularly significant effect on the spatial

variation at the regional scale. Elevation is recognized as the most important factor that characterizes the overall spatial trend of LST (Sun, 2016). DEM data and LST data are used to construct linear regression relationships for invalid pixels based on the clear sky pixels in the neighborhood of a certain window size; these data are then used to predict the missing value pixels by linear interpolation. After several simulations of the experimental pixel window size, the noise was found to be minimized when a sliding window of 19 by 19 pixels was used, and this window size was considered to have the best complement value.

$$T_i = \alpha \times h_i + \beta, \tag{7}$$

where T_i is the surface temperature data after interpolation (units: °C); h_i is the elevation value of pixel i (units: m); α is the influence coefficient of the elevation on the surface temperature, which is the regression coefficient; β is the estimated intercept. Finally, we accurately crop the image to a China-wide image to ensure that the sliding pixel window reaches the edge of the study area.

3.4 Analysis of the LST time series trend

In this study, the slope of a linear regression describes the rate of LST cooling/warming and is calculated by Eq. (8). The slope value and correlation coefficient (R), calculated with Eq. (9), were selected to quantify the temporal and spatial patterns in the LST variations.

$$Slope = \frac{\sum_{i=1}^{n}(iT_i) - \frac{1}{n}\sum_{i=1}^{n} i \sum_{i=1}^{n} T_i}{\sum_{i=1}^{n} i^2 - \frac{1}{n}(\sum_{i=1}^{n} i)^2}, \tag{8}$$

$$R = \frac{n\sum_{i=1}^{n}(iT_i) - \sum_{i=1}^{n} i \sum_{i=1}^{n} T_i}{\sqrt{n\sum_{i=1}^{n} i^2 - (\sum_{i=1}^{n} i)^2}\sqrt{n\sum_{i=1}^{n} {T_i}^2 - (\sum_{i=1}^{n} T_i)^2}}, \tag{9}$$

where i is the number of years, T_i is the average LST of year i, and n is the length of the LST time series; here, n is 15. A positive slope indicates an increase in LST (warming); a negative slope indicate a decrease in LST (cooling). The R values range from -1 to 1. An R value greater than 0 means that the LST is positively correlated with the time series, and an R value less than 0 means that the LST is negatively correlated with the time series. Meanwhile, the larger the absolute value of R is, the stronger the correlation with the time series changes.

4 Results

Temperature changes show significant differences at different time scales (day, night, month, season and year) and different spatial scales. Therefore, various QA methods were performed to validate the new data on the monthly and seasonal scales. Furthermore, to better understand the spatial and temporal variations in surface temperature and the reactions of different regions, we analyze the LST at yearly, seasonal and monthly scales and show the changes in temperature during the day and at night.

4.1 Annual change analysis

4.1.1 Average change

After LST data restoration data reconstruction, four overpass times of images are obtained each month. Calculated the average LST at four times to represent the LST image of the month. To obtain the overall LST trend, we averaged the LST for each year to remove seasonal effects. Fig. 5 shows the annual average LST change in China over the period from 2003-2017. The LST fluctuations in China exhibited a general weak positive trend. The sliding average of the 5-year unit also showed a weakly fluctuating positive trend. The lowest LST in China appeared in 2012 at 7.51 °C. The temperature reached its highest value in 2007 (9.26 °C), but after 2012, the LST remained high. This result coincides with the global warming stagnation period that was noticed from 1998-2012, and the LST increased significantly after 2012. After analyzing the LST on the seasonal and monthly scales, we found that the cooling in 2012 mainly occurred in the winter, as it was concentrated from January-February, and the cooling in the southern region was more significant than that in the other regions. In 2012, due to the abnormally strong East Asian winter monsoon, there was abnormal rainfall in the south in winter. We also observed a sudden decrease in LST in 2008 and a sudden increase in 2013. In 2008, severe persistent low-temperature snowstorm events in southern China in winter caused a decline in LST. The warming in 2013 was mainly affected by the abnormally high temperatures in the middle and lower reaches of the Yangtze River in summer.

Insert Figure 5 here

Figure 5: Annual mean LST changes in China from 2003 to 2017.

Insert Figure 6 here

Figure 6: Spatial dynamics of interannual change trends in LST from the slope (a) computed by Eq. (8), the correlation coefficient (b) computed by Eq. (9) and frequency distribution of the slope (c) during 2003-2017. In panel c, the different temperature trends (slope) are divided into 10 subinterval ranges corresponding to the ranges in panel a. The area to the left of the line AB represents the proportion of the area that experienced cooling, and the area to the right represents the proportion that experienced warming.

For a more detailed understanding of the spatial patterns and regional differences in the LST changes in different areas, the rate of annual average LST change per pixel from 2003 to 2017 was calculated, and the slope (Fig. 6a), correlation coefficient (R, Fig. 6b), frequency distribution of the slope (Fig. 6c) and the significance of the trend (P, Fig. A1) are shown. From 2003-2017, the annual average LST in China showed a weakly positive trend. The LST exhibited a strong positive trend in many regions in the north but negative trends in the south, and the positive trend in the west was greater than that in the east. Different regions showed significant regional variations. Most of China, accounting for 63.7 % of the study area, experienced a positive trend (slope >0) (corresponding to the pale yellow, yellow, light orange, orange, and red parts in Fig. 6c). Additionally, approximately 20.80 % of the pixels experienced significant warming (slope >0.05, R >0.6, $P < 0.05$). The areas with significant warming were mainly concentrated in the Inner Mongolian Plateau and areas to the south in the northwestern region.

In contrast, 36.25 % of the areas showed a negative trend (slope <0, depicted in green and 4 shades of blue in Fig. 6c). The area with a significant cooling pattern (slope <-0.075, R < -0.6, P < 0.05) covered 6.53 % of the study area, and these areas were mainly concentrated in the northeast. The specific spatial variation characteristics of LST in different regions are summarized as follows.

In 2013, the Intergovernmental Panel on Climate Change (IPCC) noted that climate warming is clear (IPCC et al., 2013). However, some areas of the Northeast Region (I) showed a significant warming hiatus over the past 15 years, and these areas made the greatest contribution to China's negative trend. We observed widespread and relatively strong cooling regimes in most areas (i.e., the slope value ranged from -0.06 to less than -0.12, see Fig. 6a, b for details), especially in the north of the Northeast Plain (slope < -0.1, R < -0.8, P < 0.05, see Fig. 6a, Fig. 6b and Fig. A1).

In the North China Region (II), The North China Plain and the Yangtze River Delta in the south both exhibit obvious positive trends, both of which are densely populated areas. In addition, the Central-Southwest China Region (III) and the South China Region (IV) also showed negative trends, but the negative trend was stronger in the South China Region (most area: slope < 0.75, R < -0.8, P < 0.05) than in the Central-Southwest China Region.

In the Northwest Region (V), some areas in the Tianshan Mountains and the Inner Mongolia Plateau experienced significant positive trends (slope > 0.10, R > 0.8, P < 0.01), and this area exhibited the strongest. positive trend in China over the past 15 years. In the Qinghai-Tibet Plateau Region (VI), the ecological environment is complex, and the unique plateau terrain and thermal properties of the surrounding areas play an important role in regulating the surrounding atmospheric circulation system. Because the Qinghai-Tibet Plateau is extremely sensitive to climate change, it is considered to be a key area of global climate change. Therefore, we have also paid close attention to the temperature changes on the Tibetan Plateau. As shown in Fig. 6a and b, an obvious positive trend was captured in the southern part of the Qinghai-Tibet region (slope >0.08), which should be emphasized. Additionally, the positive trend in the Qaidam Basin in the northeast is significantly higher (slope > 0.1) than that in the surrounding area.

4.1.2 Day and night change analysis

To more specifically assess the interannual changes in LST, we further analyzed the day and night trends in LST. The spatial distribution of the average annual day and night LST in the time series is shown in Fig. 7, and the corresponding significance is shown in Fig. A2. During the day, the positive trend mainly comes from the eastern part of North China, the central and western parts of the northwest, and the southern part of the Qinghai-Tibet Plateau. The annual daytime positive/negative trends of LST in most regions from 2003 to 2017 are significantly higher than those in the nighttime; thus, the average LST positive/negative trends can be attributed to changes during the daytime. The temperature difference between day and night also indicates that the trend of LST changes is more likely due to factors such as daytime human production and sunshine hours. The effects of changes in solar radiation on the near-surface thermal conditions are the most pronounced. Among these changes, the positive trend in the southern part of the Qinghai-Tibet Plateau is obvious (slope >0.09). Duan and Xiao (2015)

found that since 1998, the amount of daytime cloud cover in the southern part of the Qinghai-Tibet Plateau has decreased rapidly, resulting in an increase in sunshine hours. The increase in solar radiation during the day will directly lead to an increase in surface temperature, which is an important factor leading to an increase in daytime temperature. However, compared with the trend during the day, the interannual temperature change trend at night is relatively gentle and can be considered stable.

Insert Figure 7 here

Figure 7: Spatial dynamics of Fig. 7 a) day and Fig. 7 a) night LST change trends based on slope (a) and correlation coefficient (b).

4.1.3 Analysis of the diurnal temperature difference

Figure 8 shows the spatial distribution of the average daytime LST, average nighttime LST, and day and night temperature differences. The LST shows significant spatial variation. During the day, the distribution of LST varies with surface insolation depending on the solar zenith angle (Jin and Liang, 2006). The highest LST during the day appears in the northwestern part of V and the desert area of the Tarim Basin in the Alashan Plateau (30 °C to more than 35 °C) rather than in the southern part of the low-latitude tropics, which is different from the patterns at night. Except for the Qinghai-Tibet Plateau, the nighttime LST decreases from the southern low-latitude areas to the northern high-latitude areas, and the spatial variation is roughly consistent throughout the six subregions. This result may suggests that the spatial temperature change is related to the latitude range, and also to the smooth terrain in the east..

As shown in Fig. 8c, the largest diurnal difference in LST is concentrated mainly in the mountainous area of the Qinghai-Tibet Plateau; this difference is greater than 25 °C and is especially prominent in the Qaidam Basin. The lowest diurnal LST difference of less than 5 °C can be seen at low latitudes along the southeastern coast in South China. At the same time, affected by the height of the plateau, the temperature difference between day and night in the Yunnan-Guizhou Plateau in the western part of Central China is 11-15 °C, while that in the eastern area at the same latitude is only 6-11 °C. Furthermore, compared with the inland areas at similar latitudes, the coastal areas (II and III in the eastern part of Central China and the central and eastern parts of the South China Region) usually have small diurnal LST differences.

Insert Figure 8 here

Figure 8: Spatial dynamics of average daytime LST (a) and average nighttime LST (b) from 2003 to 2017

4.2 Seasonal change analysis

In addition to analyzing the characteristics of the interannual variation in LST, we also conducted an analysis of the seasonal variation characteristics to further reveal the LST variation patterns in detail (see Fig. 9 and Fig. A3). The variation characteristics are also described by the slope of the change and the correlation coefficient (R) proposed in Section 3.4.1. The results show that there is a significant spatial difference between the seasonal surface temperature trends, reflecting the effect of seasonal temperature changes on regional temperature changes. From 2003 to 2017, the positive trend in the four seasons was most significant in winter, which exhibited the largest warming area (accounting for 70 %), followed by that in spring and

summer, and the national average LST change in autumn basically did not change. Compared with the global warming hiatus that occurred from 1998 to 2002, the positive trends in China showed large differences in the four seasons (Li et al., 2015).

Specifically, in spring, the warming area is mainly concentrated in the northern areas (I, II, and V), while a weak negative trend is observed in the southern areas. The largest positive trend over the northern areas appears in the Inner Mongolia Plateau (slope >0.18, $P < 0.01$). In addition, rapid warming also occurred in the North China Plain in the eastern part of the North China Region (II) (especially near Beijing and some areas of Hebei Province, slope >0.12, $R > 0.6$, $P < 0.01$)f

As shown in Fig. 9, compared with the other two seasons, both summer and autumn showed weak positive trends throughout the country. In summer (Fig. 9b1, b2 and b3), there were slight increasing trends in most areas of China, while there were still negative trends in the Northeast Region (I) (details in Fig. 9). Significant increasing trends were mainly observed in the Qinghai-Tibet Plateau, North China Plain, Inner Mongolia Plateau, Tarim Basin and some areas in the north, with the largest positive trend in the Qinghai-Tibet Plateau. In autumn, the negative trends were mainly present in the Northeast Region (I) and the Northern Chinese Tianshan Mountains in the Qinghai-Tibet Plateau Region (VI). In contrast, the Qinghai-Tibet Plateau was still controlled by strong positive trends (near Lhasa city, slope=0.09, R=0.60, $P < 0.05$), especially in the southern part of the Tanggula Mountains.

In winter, 69.4 % of the areas experienced warming, which is significantly higher than in other seasons; thus, winter is the most important source of interannual increases in the average LST. The most remarkable positive trends in winter were observed in the Northwest Region (V) and the Qinghai-Tibet Plateau Region (VI).

Insert Figure 9 here

Figure 9: The interseasonal variability rates (slope) and correlation coefficients (R) of LST in spring (a), summer (b), autumn (c) and winter (d) from 2003 to 2017: a1, b1, c1 and d1 are the spatial distributions of the slopes in the four seasons; a2, b2, c2 and d2 are histograms of the slopes in the four seasons; and a3, b3, c3 and d3 are the spatial distributions of the correlation coefficients (R) in the four seasons.

5 Verification and discussion

MODIS exhibits good performance in retrieving LST data, as has been verified by various studies (Wan et al., 2004, Wan, 2008, Wan and Li, 2011, Wan, 2014). Furthermore, to better evaluate the accuracy of the new dataset, we performed verification for different regions using independent in situ data (subset (2) in Section 3.2) that was not used during the reconstruction process. Figure. 10 shows the statistical results of the difference between the two types of data in the six natural subregions (shown in blue in the scatterplot).

According to the scatterplots of the ground station data and the reconstructed monthly MODIS LST data shown in blue in Fig. 10, we employed a correction model that uses the results of linear regression analysis between the two datasets to further improve the accuracy. The goal of the calibration model is to reduce or eliminate the combined error introduced by various variables. Therefore, the six subregions with different climatic conditions are corrected separately to obtain better calibration

results for the study area. Additionally, to eliminate contrasts at the boundaries among the six regions, smooth constraints are imposed on some edge pixels with significant differences to guarantee consistency among the regions. The comparisons of the corrected LST data with the ground station data are indicated by the gray points in Fig. 10. In this study, the main reason for adopting the regression analysis model is the reality that a linear model can further improve the robust reconstruction results have been obtained through a large amount of work. The results show that the model reconstruction results are highly consistent with the ground station data; thus, the problem of underestimation of MODIS LST data in some areas has been resolved.

In this study, three statistical accuracy measures are used to evaluate the accuracy of the calibration: the square root of the Pearson coefficient (R^2), root mean squared error (RMSE) and mean absolute error (MAE). All subregions showed good agreements between MODIS LST and meteorological station data. The R^2 values varied from 0.93 to 0.99, with an average of 0.97. The MAE varied from 1.23 °C to 1.37 °C, with an average of 1.30 °C. The RMSE ranged from 1.24 °C to 1.58 °C, with an average of 1.39 °C. A relatively large RMSE between the reconstructed LST and ground-based LST appeared in some sites in the Qinghai-Tibet Plateau Region, indicating that the temperature exhibited great spatial heterogeneity over the complex terrain. As shown in Fig. 1, there are relatively few meteorological stations in western China. Under the same conditions, the accuracy in western China is lower than that in areas with dense weather stations when using surface meteorological stations to reconstruct LST values under cloudy conditions. Figures 3 and 4 indicate that there are more clouds in eastern China than in western China. In this case, the number of days in which LST values can be obtained from the remote sensing images in a month is much smaller in eastern China than in western China. In this study, the accuracy evaluation is based on the monthly scale. The accuracy is mainly determined by the number of days of effective pixels on the monthly and annual scales, and our analysis indicates that the more days of available pixels corresponding to the pixels on the monthly scale, the higher the accuracy.

These results indicate that the reconstructed MODIS LST dataset is robust and accurate due to its high consistency with the in situ data. Therefore, we believe that the accuracy of LST data can be improved by this method.

Insert Figure 10 here

Figure 10: The scatter diagrams in six natural subregions (I, II, III, IV, V, and VI) between the ground station data and the monthly MODIS LST data. The blue points represent the verification results of the reconstructed MODIS LST, and the statistical accuracy measures (R^2, RMSE, and MAE) are also indicated. The results of the corrected linear model are indicated in gray.

To further understand the credibility of the data and clarify the limitations of the use of this method, we further assess the performance in terms of the seasonal bias and compare it with the original seasonal LST data. Verification ground stations in representative areas are selected to help illustrate the distribution of the error in the reconstructed data. Six key zones are identified, corresponding to the areas a, b, c, d, e, and f shown by the red circles in Fig. 1, and an overview of the ground stations can be found in Table 1. The six key zones are selected, including the three most significant regions for warming (b, d, and f), the two most significant regions for cooling (a and c), and the zone located in Xinjiang Province (see Fig. 6a for details). Zone (a) located in the Northeast Region and zone (b) located in the North China Region experienced the strongest

negative trend and significant warming, respectively. In particular, special attention has been given to the area around the Taklimakan Desert (e) in Xinjiang, which has complex terrain and extensive heterogeneity.

Seasonal-scale verification was performed using the RMSEs between the MODIS data (including the original LST and reconstructed LST data) and ground-based LSTs for comparison in the six key zones, as shown in Table 2. The original MODIS monthly LST data were used directly without filtering quality flags. For the original MODIS LST images, we averaged the LST data of the month corresponding to the season, and obtained the seasonal LST images. The pixels with missing LST values in original MODIS LST images for the corresponding months of the season were not used in the verification process. Therefore, if there is no missing value for the LST pixel corresponding to the site, each station can have a maximum of 15 values in each season. Compared with that of the original LST, the average RMSE of the reconstructed LST data decreased by 18 % from 1.79 °C to 1.46 °C. Both datasets exhibited the largest RMSE in summer and the smallest in autumn, indicating that the original and reconstructed LST data have highly consistent seasonal patterns. For the reconstructed LST data, we further found that the RMSE values at some sites in the summer were significantly higher than those at other sites. The regions that exhibited high RMSE values were mainly concentrated in the western regions (Xinjiang, Inner Mongolia and the Qinghai-Tibet Plateau), while the values in the other three regions were relatively low. The main reason for this difference may be the complex and diverse terrain and large climate differences in the western region. The average RMSE in autumn was the lowest at 1.07 °C. The winter RMSE ranged from 0.04 to 3.81, with an average of 1.45 °C.

For the reconstructed LST data, the distribution of the RMSE varied greatly between the eastern and western regions at the seasonal scale. The maximum RMSE values for all stations in the eastern typical zones (i.e., key zone a in the Northeast Region (I) and key zone b in the North China Region (II)) occurred in the cold winter, while the highest values for most sites in the western region occurred during the hot summer months (i.e., the remaining four zones). At the same time, the comparison results show that the mean RMSE between the ground-based observation data and the LST data was significantly higher in the western region than in the eastern region (mean 1.04 °C in eastern regions I and II and 1.69 °C in western regions IV, V, and VI). The large spatial variations in temperature caused by the complex terrain in the western region may be the cause of these large errors. At the same time, a large RMSE between the reconstructed LST data and the ground-based LST data appeared in some locations in Inner Mongolia (i.e., key zone e) in the Northwest Region, further indicating that the temperature over complex terrain exhibited great spatial heterogeneity.

We also note that the selected ground stations shown in Table 2 located in six key zones are examples of where the local LST warming/cooling rate changed by more than the average rate, and these areas actually include areas with greater terrain complexity. Moreover, the examples indicate that the reconstruction model proposed here is effective even in the areas of complex topography.

Table 2. RMSEs of seasonal LST between monthly LST data (including the original LST data and reconstructed LST data) and ground-based LST data (Orig. indicates original LST values at the ground stations. Recon. indicates the reconstructed LST values at the ground stations)

Insert Table 2 here

The verification results show that the dataset has reasonable consistency with the in situ measurements, indicating that the interference of cloud coverage is well eliminated. The dataset obtained after reconstruction is a large-scale, long-term, unique surface temperature dataset because it eliminates low-quality pixels caused by factors such as cloud disturbance and achieves complete coverage of the entire study region. The accuracy and spatiotemporal continuity of this dataset are much better than those of the original MODIS monthly data. Moreover, in this dataset, the true ground surface temperatures under cloud coverage are retrieved instead of reconstructing the LST under clear-sky conditions, which is better than the methods used in many previous studies.

Furthermore, the reconstruction strategy, which combines monthly data with daily data, effectively solves the problem of reconstructing real LST data under cloud coverage with very limited information and improves the accuracy of the monthly data reconstruction results. The final linear correction model improves the consistency of the LST data with terrestrial data. We believe that these datasets can be applied to research regional agricultural ecological environments and to monitor agrometeorological disasters. In a small range of practical applications, such as urban heat island monitoring, our current data may not be suitable for monitoring in great detail due to the coarse resolution, which is something we need to improve in future work.

6 Data availability

The LST dataset in China is distributed under a Creative Commons Attribution 4.0 License. The data are freely available from the Zenodo repository https://doi.org/10.5281/zenodo.3528024 (Zhao et al., 2019).

7 Summary and conclusions

Based on the Terra and Aqua MODIS land surface temperature dataset and meteorological station data, a new LST dataset over China was established for the period from 2003-2017. This dataset effectively removed approximately 20 % of the missing pixels or low-quality LST pixels low-quality by clouds in the original MODIS monthly image. A detailed comparison and analysis with the in situ measurements shows that the reconstruction results have high precision, the average RMSE is 1.39 °C, the MAE is 1.30 °C and the R^2 is 0.97. The data are freely available at https://doi.org/10.5281/zenodo.3528024 (Zhao et al., 2019). We believe that this dataset will be of great use in research related to temperature, such as high temperature and drought studies, because it effectively overcomes the limitations of reconstructing the real LST under cloudy conditions in the past and achieves good spatiotemporal coverage.

The constructed high-precision monthly LST dataset for China provides a detailed perspective of the patterns of the spatial and temporal changes in LST. The LST dataset was used to analyze the regional characteristics and capture the variations in LST at the annual, seasonal and monthly scales. Our results showed that the LST showed a slight upward trend with a slope of 0.026 (approximately 63.7 % and 20.80 % of the pixels underwent warming and significant warming, respectively). There

were great regional differences in the climate positive trend. The Northwest Region, the Qinghai-Tibet Plateau Region and the North China Plain experienced significant positive trends (i.e., the slope ranged from 0.025 to greater than 0.1). The impacts of human activities on warming, such as the increase in greenhouse gases and black carbon aerosol emissions from urbanization and industrial and agricultural development, are prominent. Greenhouse gases absorb infrared longwave radiation from the ground, which results in an increase in warming. Moreover, the coupling of greenhouse gases and monsoon systems can result in changes in the energy budget in the monsoon region, which affect the intensity of monsoon circulation. Additionally, the change in temperature in the short term may be caused by the increase in aerosols such as scattering aerosols and black carbon emitted along with other atmospheric pollutants. Black carbon aerosol pollution leads to heating of the air and a reduction of the cooling effect of solar radiation reaching the surface, resulting in local or even global climate changes (Kühn et al., 2014). However, scattering aerosols are expected to produce cooling effects by absorbing and scattering solar radiation. Consequently, the effect of human activities on global climate change is complex. The impact of human activities on temperature trends is expected to be especially pronounced in rapidly expanding urban areas, such as North China and the Yangtze River Delta.

Meanwhile, a negative trend was also observed in China: most areas of the Northeast Region and South China Region became colder, especially in the Songnen Plains in the middle of the region (i.e., slope=-0.11, R=0.61, $P < 0.05$). Some areas in South China also showed a slight negative trend ($P < 0.05$). The interannual temperature changes indicated that the daytime temperature changed more intensely than the nighttime temperature, which may be closely related to changes in solar radiation and the release of large amounts of greenhouse gases from human activities. Seasonal changes are primarily driven by Earth's rotation but are also affected by monsoon changes, ocean currents and other factors. The LST trends showed significant changes among the different seasons. The positive trend in winter was more significant than that in the other three seasons, especially in the northwestern region of the arid and semiarid zone and the Qinghai-Tibet Plateau. As a key parameter for different research fields, such as simulating land surface energy and water balance systems, LST provides important information for monitoring and understanding high-temperature and drought conditions, which must be taken into consideration for agricultural production and meteorological research. Therefore, we believe that the LST dataset produced in this study can be useful for drought research and monitoring and can be further used for agricultural production and climate change research.

Author contributions. KM and YC designed the research and developed the methodology; BZ and XM supervised the downloading and processing of satellite images; BZ wrote the manuscript; and JS, ZL, ZQ and all other authors revised the manuscript.

Competing interests. The authors declare no conflicts of interest.

Acknowledgments. This research was funded by the National Key R&D Program Key Project (Meteorological Disaster Risk Maps and Integrated Systems for China and Typical Regions/Watersheds (2019YFC1510203), the Second Tibetan Plateau

Scientific Expedition and Research Program (STEP) (Grant No. 2019QZKK0206)), Fundamental Research Funds for Central Non-profit Scientific Institution (Grant No. 1610132020014), and Open Fund of State Key Laboratory of Remote Sensing Science (Grant No. OFSLRSS201910). The authors would also like to thank the National Aeronautics and Space Administration (NASA) for their support by providing the LST product and elevation data. We also thank the China Meteorological Administration for providing the meteorological LST data.

References

André, C., Ottlé, C., Royer, A., and Maignan, F.: Land surface temperature retrieval over circumpolar arctic using ssm/i–ssmis and modis data. Remote Sens. Environ., 162, 1–10, https://doi.org/10.1016/j.rse.2015.01.028, 2015.

Benali, A., Carvalho, A. C., Nunes, J. P., Carvalhais, N., and Santos, A.: Estimating air surface temperature in Portugal using MODIS LST data. Remote Sens. Environ., 124, 108-121, https://doi.org/10.1016/j.rse.2012.04.024, 2012.

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

Deng, M. J.: "Three Water Lines" strategy: Its spatial patterns and effects on water resources allocation in northwest China. J Geogr., 73, 1189-1203, https://doi.org/10.11821/dlxb201807001, 2018 (in Chinese).

Duan, A. and Xiao, Z.: Does the climate warming hiatus exist over the Tibetan Plateau? Sci Rep., 5, 13711, https://doi.org/10.1038/srep13711, 2015.

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

Gafurov, A. and Bárdossy, A.: Cloud removal methodology from MODIS snow cover product, Hydrol. Earth Syst. Sci. Data., 13, 1361–1373, https://doi.org/10.5194/hess-13-1361-2009, 2009.

Gao, L., Wei, J., Wang, L., Bernhardt, M., and Schulz, K.: A high-resolution air temperature dataset for the Chinese Tian Shan in 1979–2016. Earth Syst. Sci. Data., 10, 2097-2114, https://doi.org/10.5194/essd-10-2097-2018, 2018.

Hansen, J., Ruedy, R., Sato, M., and Lo, K.: Global surface temperature change. Rev Geophys., 48, RG4004, https://doi.org/10.1029/2010RG000345, 2010.

Jin, M. and Liang, S. L.: An improving land surface emissively parameter of land surface models using global Remote Sens observations J. Clim., 19, 2867–81, https://doi.org/10.1175/jcli3720.1, 2006.

Jin, M. L. and Dickinson, R. E.: Land surface skin temperature climatology: benefitting from the strengths of satellite observations. Environ. Res. Lett., 5, 044004, https://doi.org/10.1088/1748-9326/5/4/044004, 2010.

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, 1602-1606, https://doi.org/10.1109/LGRS.2013.2263553, 2013.

Li, B., Chen, Y., and Shi, X.: Why does the temperature rise faster in the arid region of northwest China? J Geophys Res-Atmos., 117, 1-7, https://doi.org/10.1029/2012JD017953, 2012.

Li, Q., Yang, S., Xu, W., Wang, X. L., Jones, P., Parker, D., Zhou, L. M., Feng, Y., and Gao, Y.: China experiencing the recent warming hiatus. Geophys Res Lett., 42, 889-898, https://doi.org/ 10.1002/2014GL062773, 2015.

Liu, H., Li, X., Fischer, G., and Sun, L.: Study on the Impacts of Climate Change on China's Agriculture. Climatic Change., 65, 125-148, https://doi.org/125–148,10.1023/B:CLIM.0000037490.17099.97, 2004.

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

Ma, X. L., Wan, Z. M., Moeller, C. C., Menzel, W. P., Gumley, L. E., and Zhang, Y.: Retrieval of geophysical parameters from Moderate Resolution Imaging Spectroradiometer thermal infrared data: evaluation of a twostep physical algorithm. Applied Optics., 39, 3537 – 3550, https://doi.org/10.1364/AO.39.003537, 2000.

Ma, X. L., Wan, Z. M., Moeller, C. C., Menzel, W. P., and Gumley, L. E.: Simultaneous retrieval of atmospheric profiles and land-surface temperature, and surface emissivity from Moderate Resolution Imaging Spectroradiometer thermal infrared data: extension of a two-step physical algorithm. Applied Optics., 41, 909 – 924, https://doi.org/10.1364/AO.41.000909,2002.

Mao, K. B., Shi, J. C., Li, Z. L., Qin, Z. H., Li, M. C., and Xu, B.: A physics-based statistical algorithm for retrieving land surface temperature from AMSR-E passive microwave data,Sci China Ser D-Earth Sci., 50, 1115-1120, https://doi.org/ 10.1007/s11430-007-2053-x, 2007.

Mao, K. B., Zuo, Z.Y., Shen, X. Y., Xu, T. R., Gao, C.Y., and Liu,G.: Retrieval of Land-surface Temperature from AMSR2 Data Using a Deep Dynamic Learning Neural Network, Chinese Geogr Sci., 28, 1-11, https://doi.org/1–11,10.1007/s11769-018-0930-1, 2018.

Mao, K. B., Yuan, Z.J., Zuo, Z.Y., Xu, T. R., Shen, X. Y., and Gao, C.Y.: Changes in Global Cloud Cover Based on Remote Sens Data from 2003 to 2012, Chinese Geogr Sci., 29, 306–315, https://doi.org/CNKI:SUN:ZDKX.0.2019-02-011, 2019.

Markus, M., Duccio, R and Markus, N.: Surface Temperatures at the Continental Scale: Tracking Changes with Remote Sensing at Unprecedented Detail. Remote Sens., 2, 333-351, https://doi.org/10.3390/rs6053822,2010.

McMillin, L. M.: Estimation of sea surface temperature from two infrared pixel window measurements with different absorptions. J. Geophys. Res., 80, 5113–5117, 1975.

Na, F., Gaodi, X., Wenhua, L., Yajing, Z., Changshun, Z., and Na, L.: Mapping air temperature in the lancang river basin using the reconstructed modis LST data. J. Res. Ecol., 5, 253–262, https://doi.org/10.5814/j.issn.1674-764x.2014.03.008, 2014.

Neteler, M.: Estimating daily land surface temperatures in mountainous environments by reconstructed MODIS LST data, Remote Sens., 2, 333–351,https://doi.org/10.3390/rs1020333, 2010.

Oku, Y., H. Ishikawa, S. Haginoya., and Y. Ma. : Recent trends in land surface temperature on the Tibetan Plateau, J. Clim., 15, 2995– 3003,https://doi.org/10.1175/JCLI3811.1, 2006.

Prigent, C., Jimenez, C., and Aires, F.: Toward “all weather” long record, and real-time land surface temperature retrievals from microwave satellite observations. J. Geophys. Res. Atmos., 121, 5699–5717. 24, 2016.

Scharlemann, J.P., Benz, D., Hay, S.I., Purse, B.V., Tatem, A.J., Wint, G.R., and Rogers, D.J.: Global data for ecology and epidemiology: A novel algorithm for temporal Fourier processing MODIS data. PLoS One., 3: e1408-10.1371, https://doi.org/10.1371/journal.pone.0001408, 2008.

Shi, Y.F., Shen, Y.P., Kang, E., Li, D.L., Ding, Y.J., Zhang, G.W., and Hu, R.J.: Recent and future climate change in northwest China. Climatic Change, 80, 379–393. https://doi.org/10.1007/s10584-006-9121-7, 2007.

Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V, and Midgley, P.M. (Eds.): IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

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, 10–20, https://doi.org/10.1016/j.cageo.2017.04.007, 2017.

Sun, X., Ren, G., Ren, Y., Fang, Y., Liu, Y., Xue, X., and Zhang, P.: A remarkable climate warming hiatus over Northeast China since 1998. Theor Appl Climatol., 133, 579–594. https://doi.org/10.1007/s00704-017-2205-7, 2017.

Sun, Z., Wang, Q., Batkhishig, O., and Ouyang, Z.: Relationship between evapotranspiration and land surface temperature under energy-and water-limited conditions in dry and cold climates. Adv. Meteorol., 1-9, 1835487, https://doi.org/10.1155/2016/1835487, 2016.

Tatem, A.J., Goetz, S.J., and Hay, S.I.: Terra and Aqua: new data for epidemiology and public health. Int. J. Appl. Earth Obs. Geoinf., 6, 33-46, https://doi.org/10.1016/j.jag.2004.07.001, 2004.

Vancutsem, C., Ceccato, P., Dinku, T., and Connor, S. J.: Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. Remote Sens. Environ., 114, 449–465, https://doi.org/10.1016/j.rse.2009.10.002, 2010.

Wan, Z. M.: Collection-5 MODIS Land Surface Temperature Products Users' Guide [EB/OL], 2007.

Wan, Z. M.: New refinements and validation of the MODIS landsurface temperature/emissivity products. Remote Sens. Environ., 112, 59–74 https://doi.org/ 10.1016/j.rse.2006.06.026, 2008.

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

Wan, Z. M. and Dozier, J.: A generalized split-window algorithm for retrieving land-surface temperature from space. IEEE T Geosci Remote., 34, 892–905, https://doi.org/ 10.1109/36.508406, 1996.

Wan, Z. M. and Li, Z. L.: A physics-based algorithm for retrieving landsurface emissivity and temperature from EOS/MODIS data. IEEE T Geosci Remote., 35, 980– 996, https://doi.org/10.1109/36.602541, 1997.

Wan, Z. M. and Li, Z. L.: MODIS land surface temperature and emissivity. Remote Sens. and Digital Image Proc., 11, 563-577, https://doi.org/ 10.1007/978-1-4419-6749-7_25, 2011.

Wan, Z. M., Zhang, Y., Zhang, Q., and Li, Z. l.: Validation of the land-surface temperature products retrieved from Terra Moderate Resolution Imaging Spectroradiometer data. Remote Sens. Environ., 83, 163–180, https://doi.org/10.1016/j.rse.2009.10.002, 2002.

Wan, Z. M., Zhang Y, Zhang Q., and Li Z L.: Quality assessment and validation of the MODIS global land surface temperature. Int J Remote Sen S., 25, 261–274 https://doi.org/ 10.1080/0143116031000116417, 2004.

Wang, K. and S. Liang.: Evaluation of ASTER and MODIS land surface temperature and emissivity products using long-term surface longwave radiation observations at SURFRAD sites, Remote Sens. Environ., 113, 1556–1565, https://doi.org/10.1016/j.rse.2009.03.009, 2009.

Wang, K., Wan, Z., Wang, P., Sparrow, M., Liu, J., and Haginoya, S.: Evaluation and improvement of the MODIS land surface temperature/emissivity products using ground-based measurements at a semi-desert site on the western Tibetan Plateau, Int. J. Remote Sens., 28, 2549–2565, https://doi.org/10.1080/01431160600702665, 2007.

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

Xu, X., Lu, C., Shi, X., and Gao, S.: World water tower: an atmospheric perspective. Geophysical Research Letters., 35, L20815, https://doi.org/10.1029/2008GL035867, 2008.

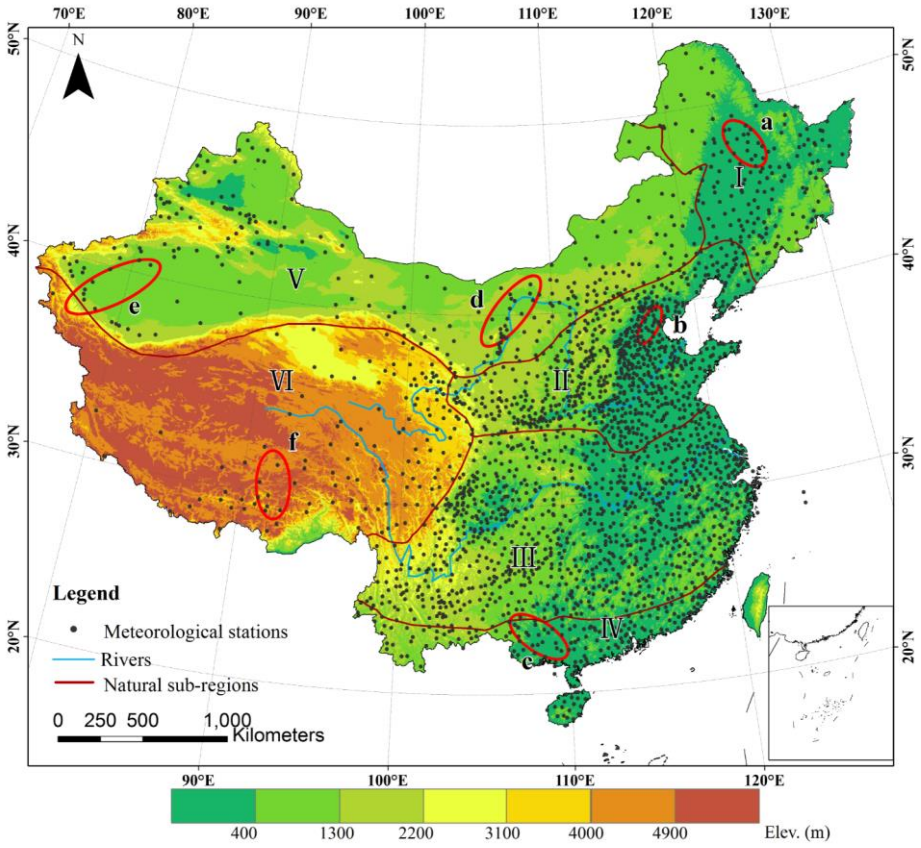
Yang, X., Lin, E., Ma, S. M., Ju, H., Guo, L. P., Xiong., Li, Y., and Xu, Y. L.: Adaptation of agriculture to warming in Northeast China. Climatic Change., 84, 45-58, https://doi.org/10.1007/s10584-007-9265-0, 2007.

Yu, W. J., Nan, Z. T., Wang, Z. W., Chen, H., Wu, T. H., and Zhao, L.: An Effective Interpolation Method for MODIS Land Surface Temperature on the Qinghai–Tibet Plateau. IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., 8, 4539-4550, https://doi.org/10.1109/JSTARS.2015.2464094, 2015.

Zeng, C., Shen, H., and Zhang, L.: Recovering missing pixels for Landsat ETM+ SLC-off imagery using multitemporal regression analysis and a regularization method, Remote Sens. Environ., 131, 182–194, https://doi.org/10.1016/j.rse.2012.12.012, 2013.

Zhao, B., Mao, K. B., Cai, Y. L., Shi, J. C., Li, Z. L., Qin, Z. H., and Meng, X. J.: A combined Terra and Aqua MODIS land surface temperature and meteorological station data product for China from 2003–2017 [Dataset], Zenodo, https://doi.org/10.5281/zenodo.3378912, 2019.

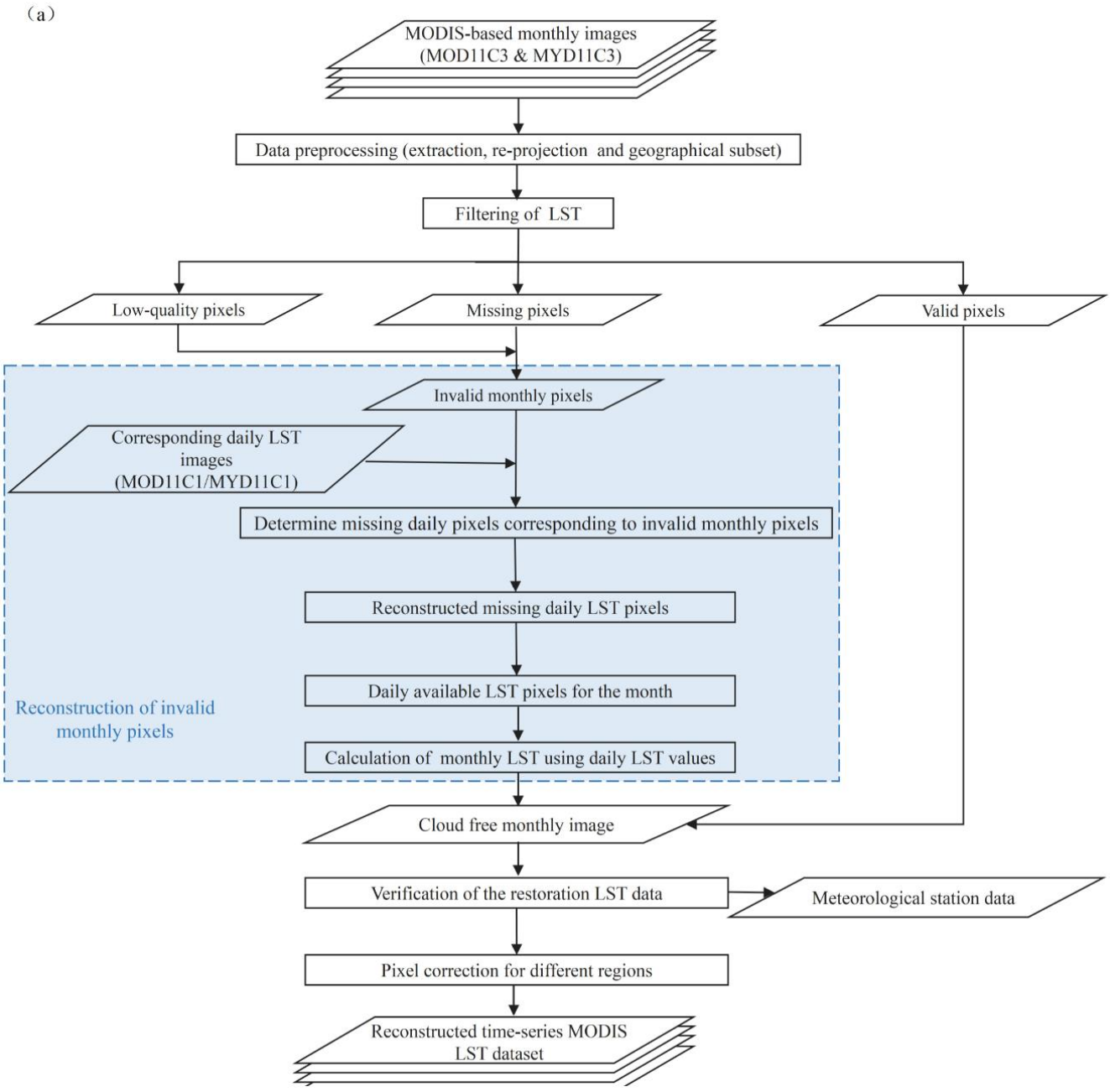
Zhao, X., Tan, K., Zhao, S., and Fang, J.: Changing climate affects vegetation growth in the arid region of the northwestern China. J Arid Environ, 75, 946–952. https://doi.org/10.1016/j.jaridenv.2011.05.007, 2011.



765

Figure 1: The study area is divided into six natural subregions (I, II, III, IV, V, and VI), and the spatial patterns of the meteorological stations in the subregions is shown. The red circles mark the key areas where the temperature has changed significantly, and meteorological station from Subset (2) located these areas are used to validate the accuracy of the new LST dataset (a, b, c, d, e, and f).

770


(a)
MODIS-based monthly images (MOD11C3 & MYD11C3)
Data preprocessing (extraction, re-projection and geographical subset)
Filtering of LST
Low-quality pixels
Missing pixels
Valid pixels
Invalid monthly pixels
Corresponding daily LST images (MOD11C1/MYD11C1)
Determine missing daily pixels corresponding to invalid monthly pixels
Reconstructed missing daily LST pixels
Daily available LST pixels for the month
Calculation of monthly LST using daily LST values
Reconstruction of invalid monthly pixels
Cloud free monthly image
Verification of the restoration LST data
Meteorological station data
Pixel correction for different regions
Reconstructed time-series MODIS LST dataset

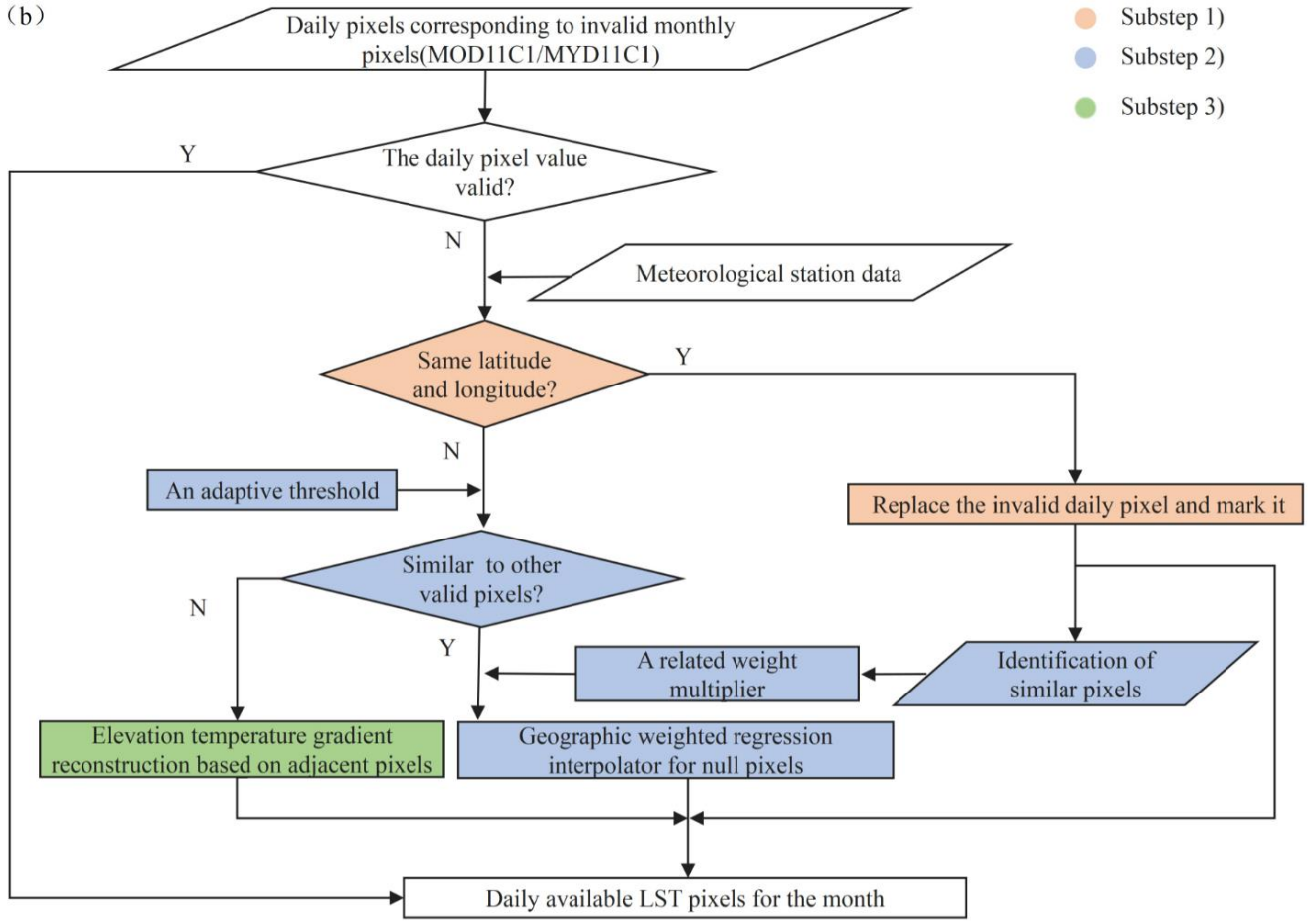


Figure 2: (a)The summary flowchart for reconstructing MODIS monthly LST data, . (b) The detailed flowchart for reconstruct missing daily pixels in (a).

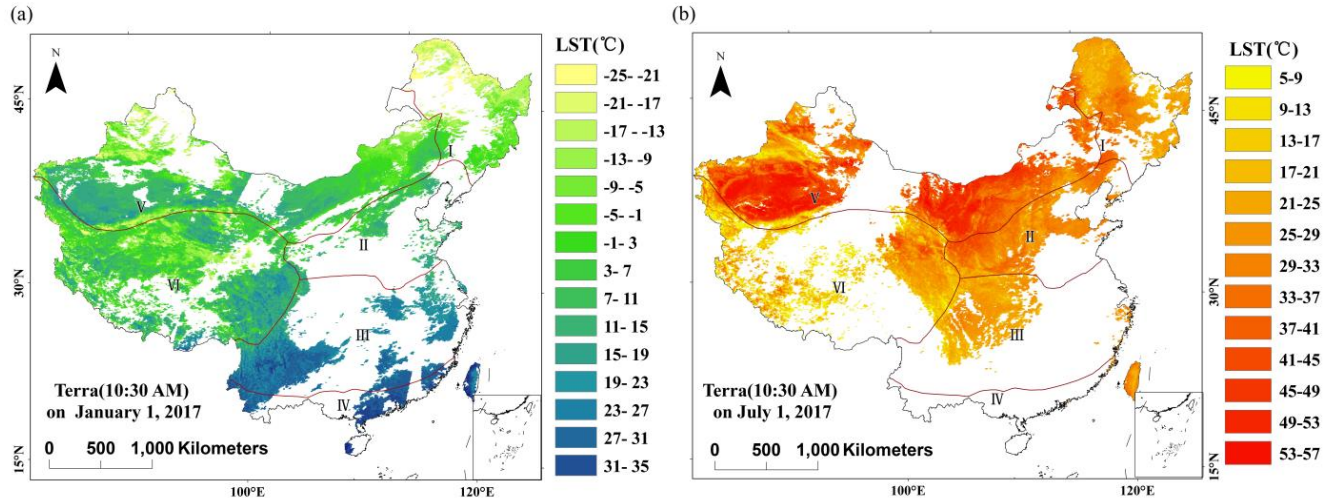


Figure 3: Spatial distribution of valid data for daily MODIS LST data from Terra during the daytime on (a) January 1, 2017, and (b) July 1, 2017. Areas of missing data are blank.

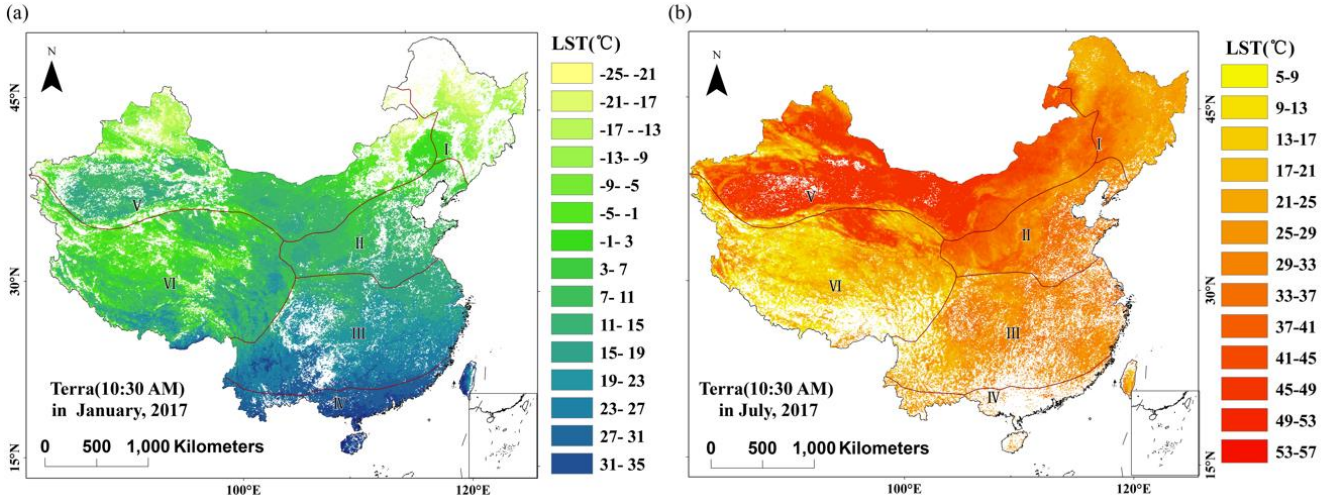


Figure 4: Spatial distribution of valid data after pixel filtering for monthly MODIS LST data from Terra during the daytime on (a) January and (b) July. Areas of invalid data are blank.

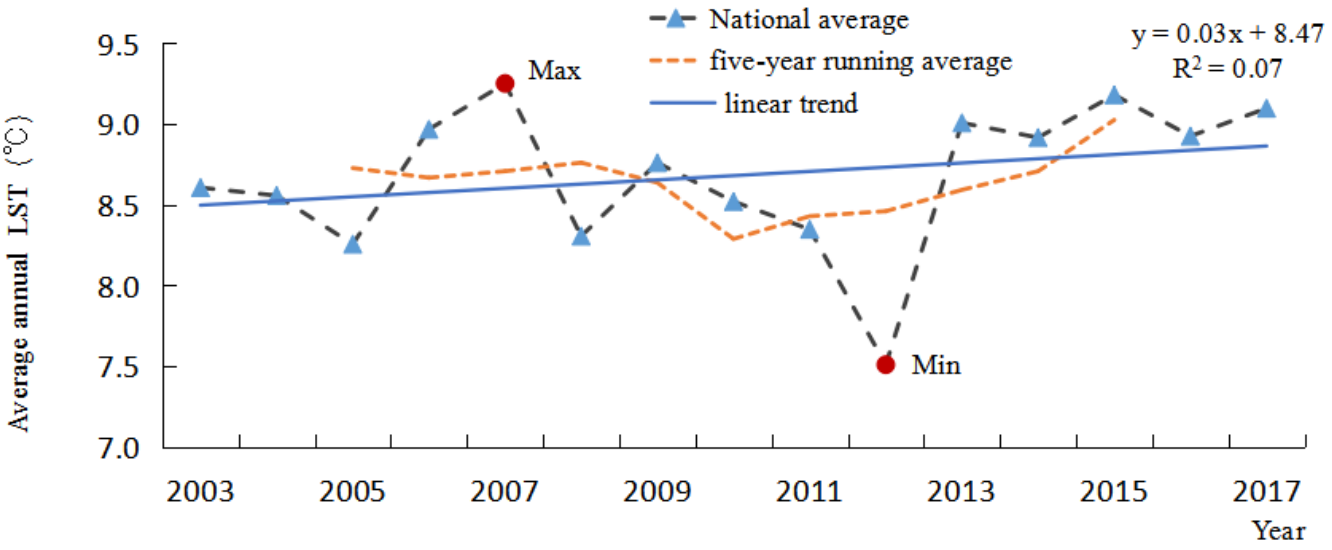
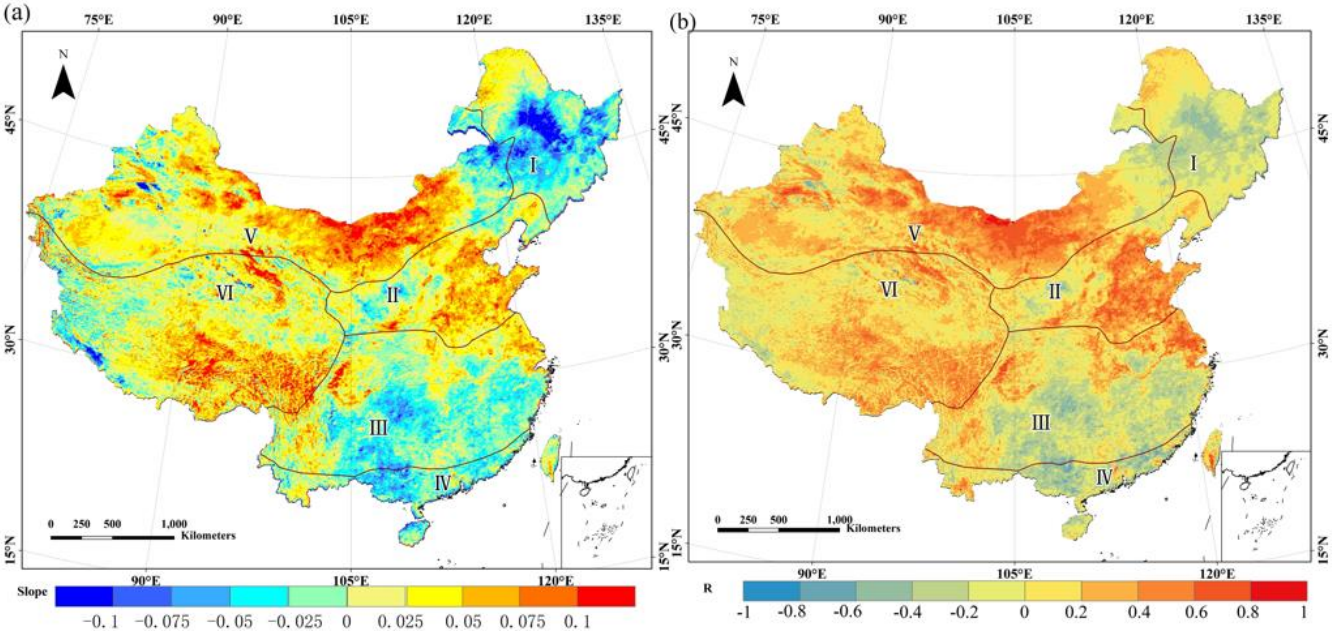


Figure 5: Time series of annual mean LST (unit: ºC) for the period of 2003-2017 based on corrected MODIS LST data for China. The solid blue line indicates the linear trend. The orange dashed line shows the five-year running average trend.



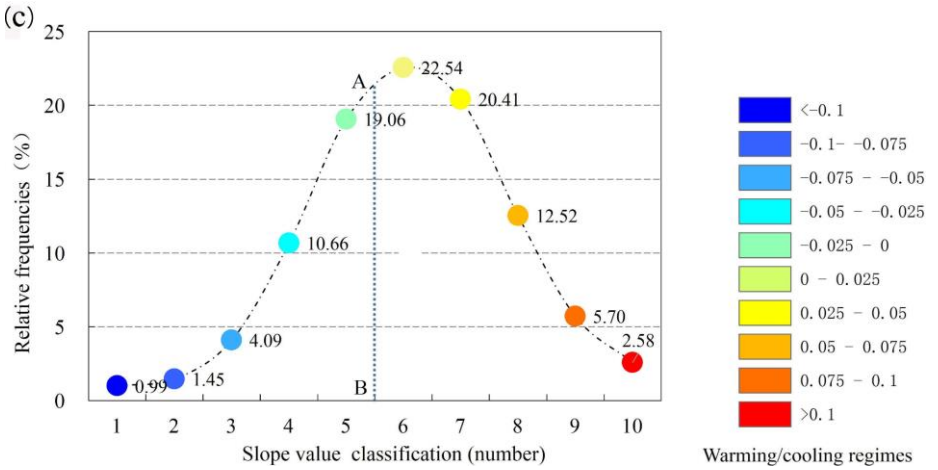
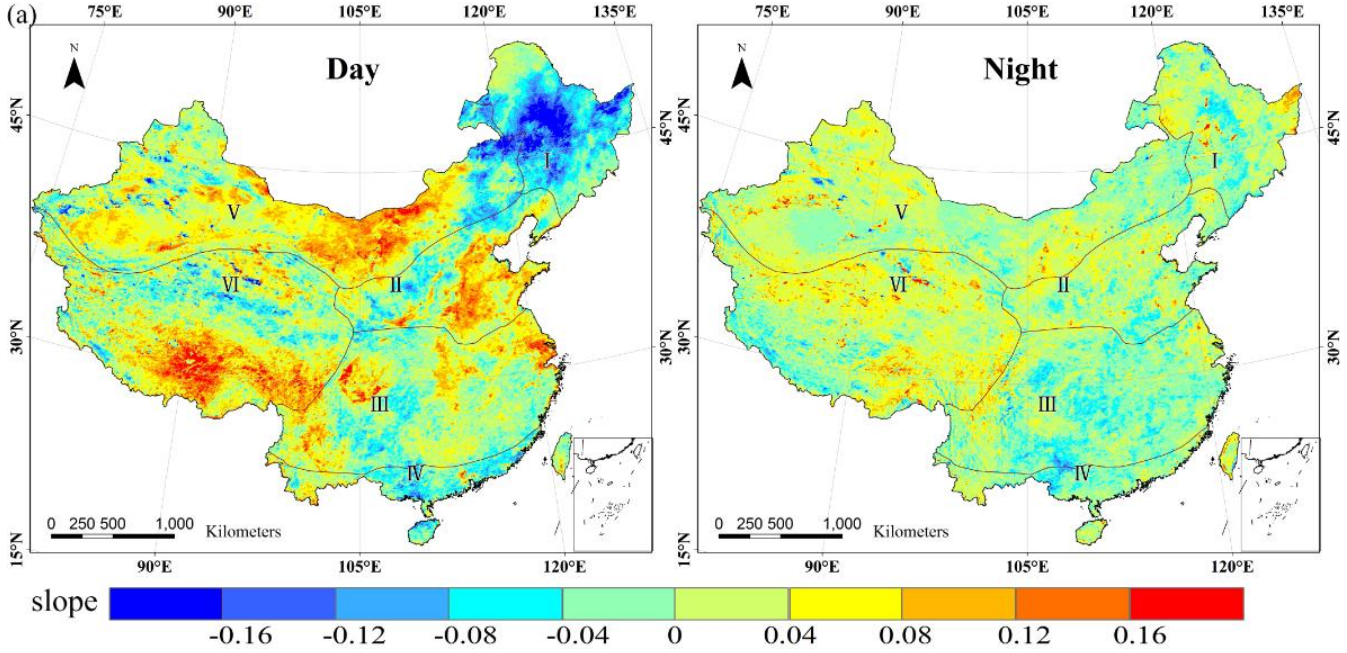
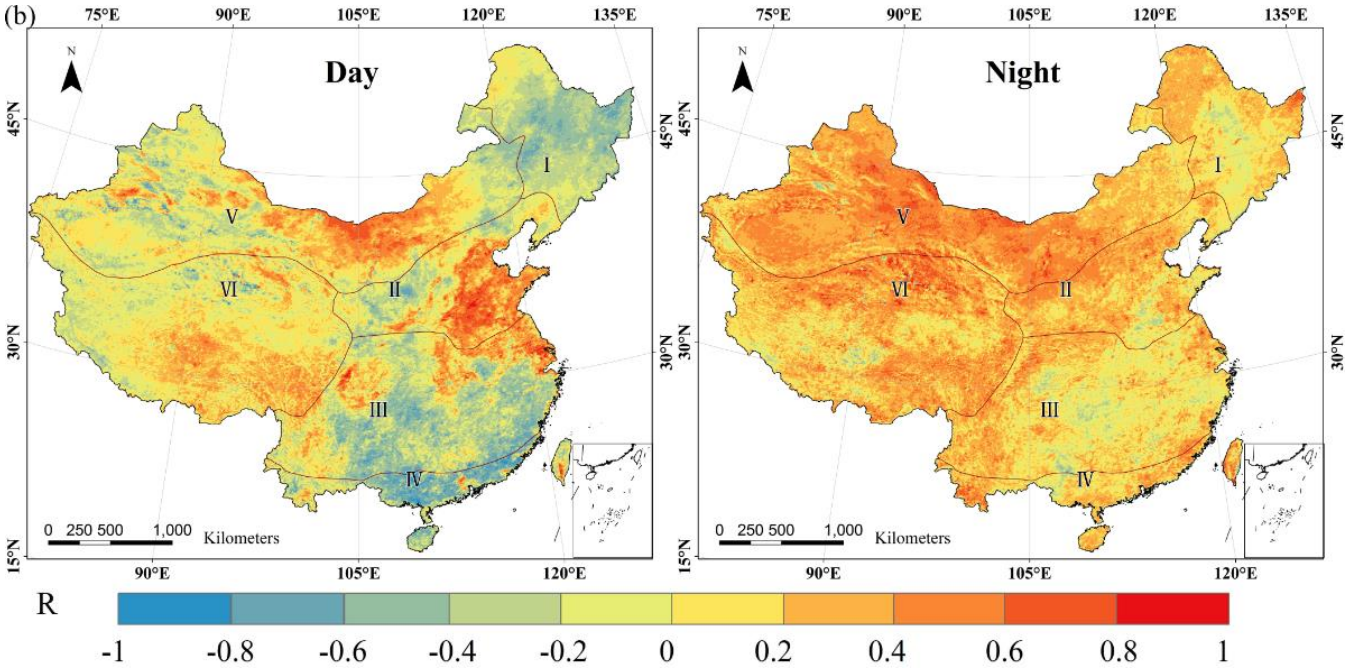


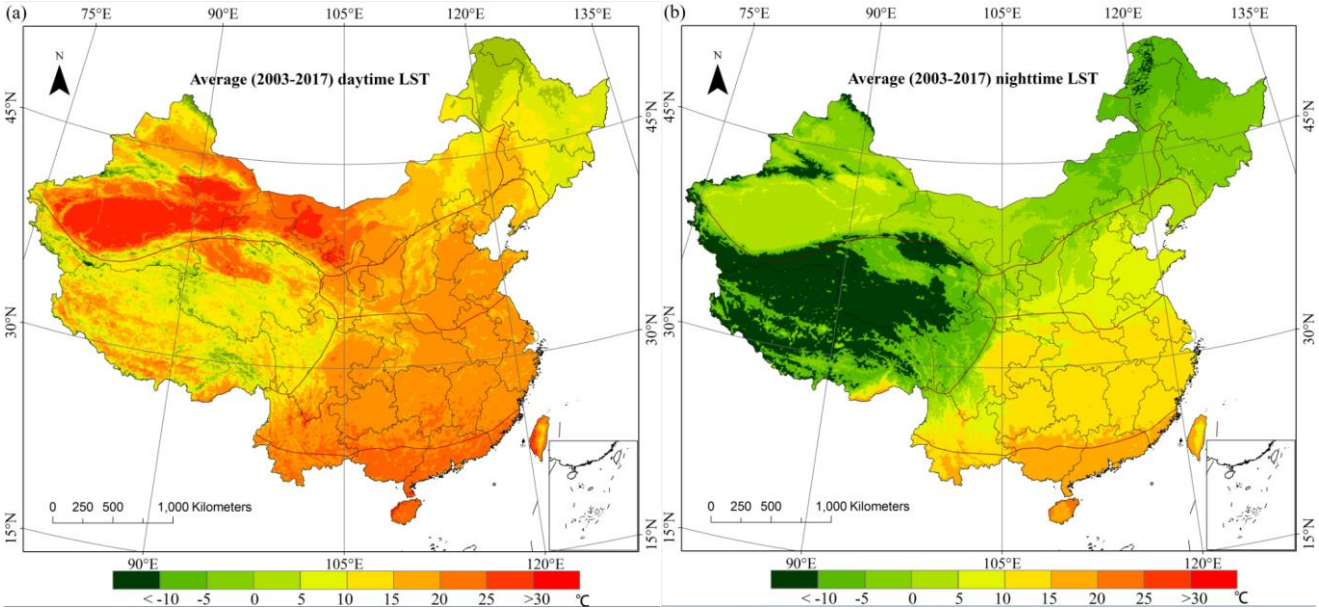
Figure 6: Spatial dynamics of interannual change trends in LST from the slope (a) computed by Eq. (8), the correlation coefficient (b) computed by Eq. (9) and frequency distribution of the slope (c) during 2003-2017. In panel c, the different temperature trends (slope) are divided into 10 subinterval ranges corresponding to the ranges in panel a. The area to the left of the line AB represents the proportion of the area that experienced cooling, and the area to the right represents the proportion that experienced warming.





795

Figure 7: Spatial dynamics of (a) day and (b) night LST change trends based on slope (a) and correlation coefficient (b).



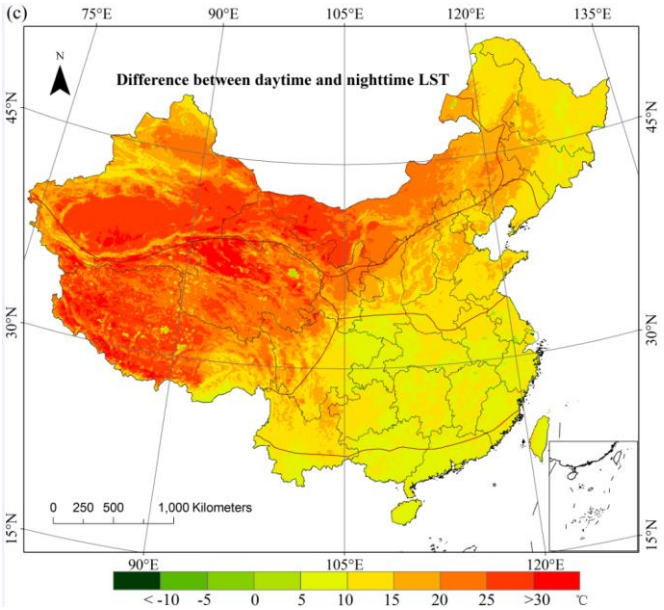
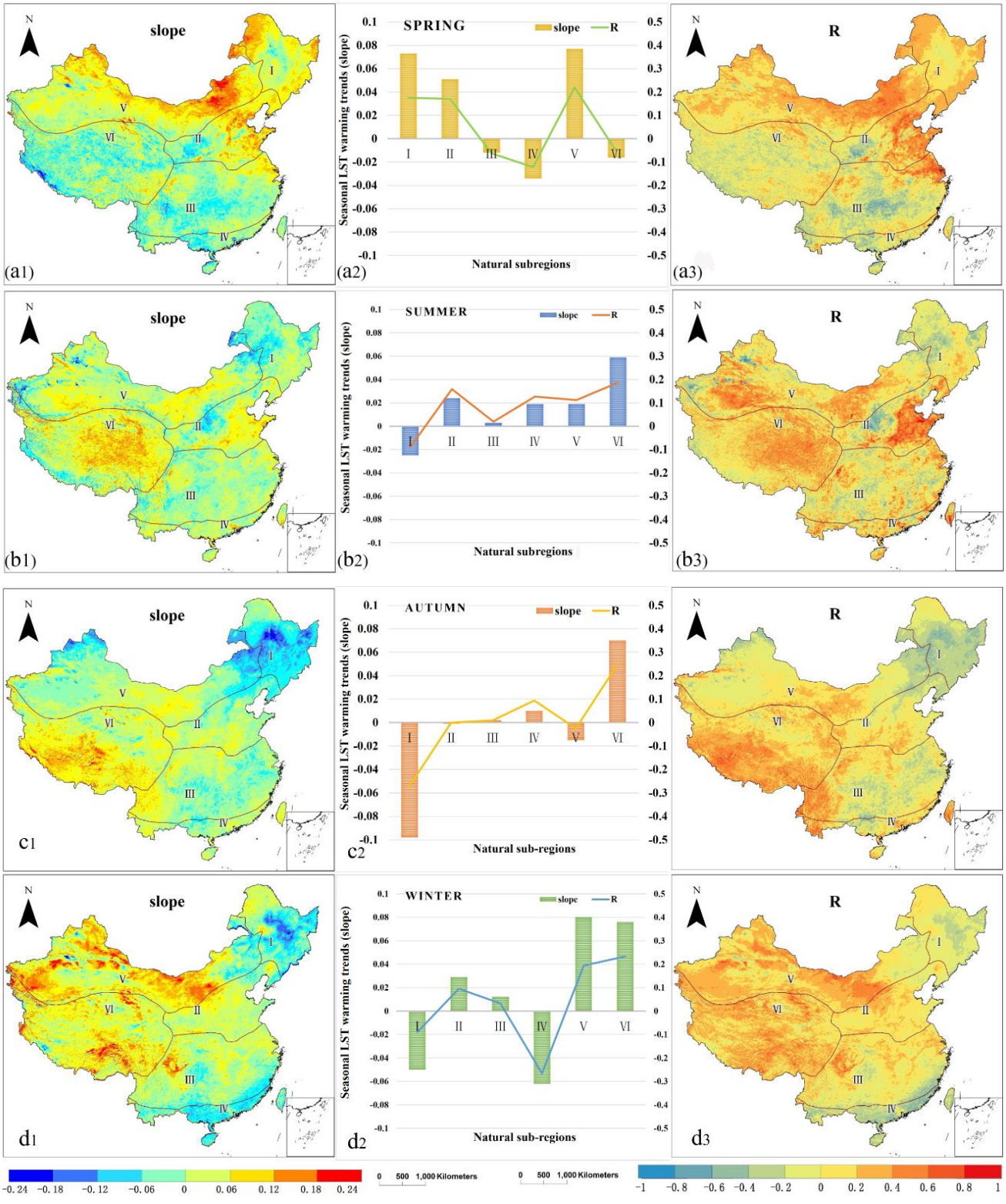


Figure 8: Spatial dynamics of average daytime LST (a) and average nighttime LST (b) from 2003 to 2017.


slope
R
SPRING
SUMMER
AUTUMN
WINTER
Seasonal LST warming trends (slope)
Natural subregions
Natural sub-regions
(a1)
(a2)
(a3)
(b1)
(b2)
(b3)
c1
c2
d1
d2
d3
0 500 1,000 Kilometers

800

Figure 9: The interseasonal variability rates (slope) and correlation coefficients (R) of LST in spring (a), summer (b), autumn (c) and winter (d) from 2003 to 2017: a1, b1, c1 and d1 are the spatial distributions of the slopes in the four seasons; a2, b2, c2 and d2 are histograms of the slopes in the four seasons; and a3, b3, c3 and d3 are the spatial distributions of the correlation coefficients (R) in the four seasons.

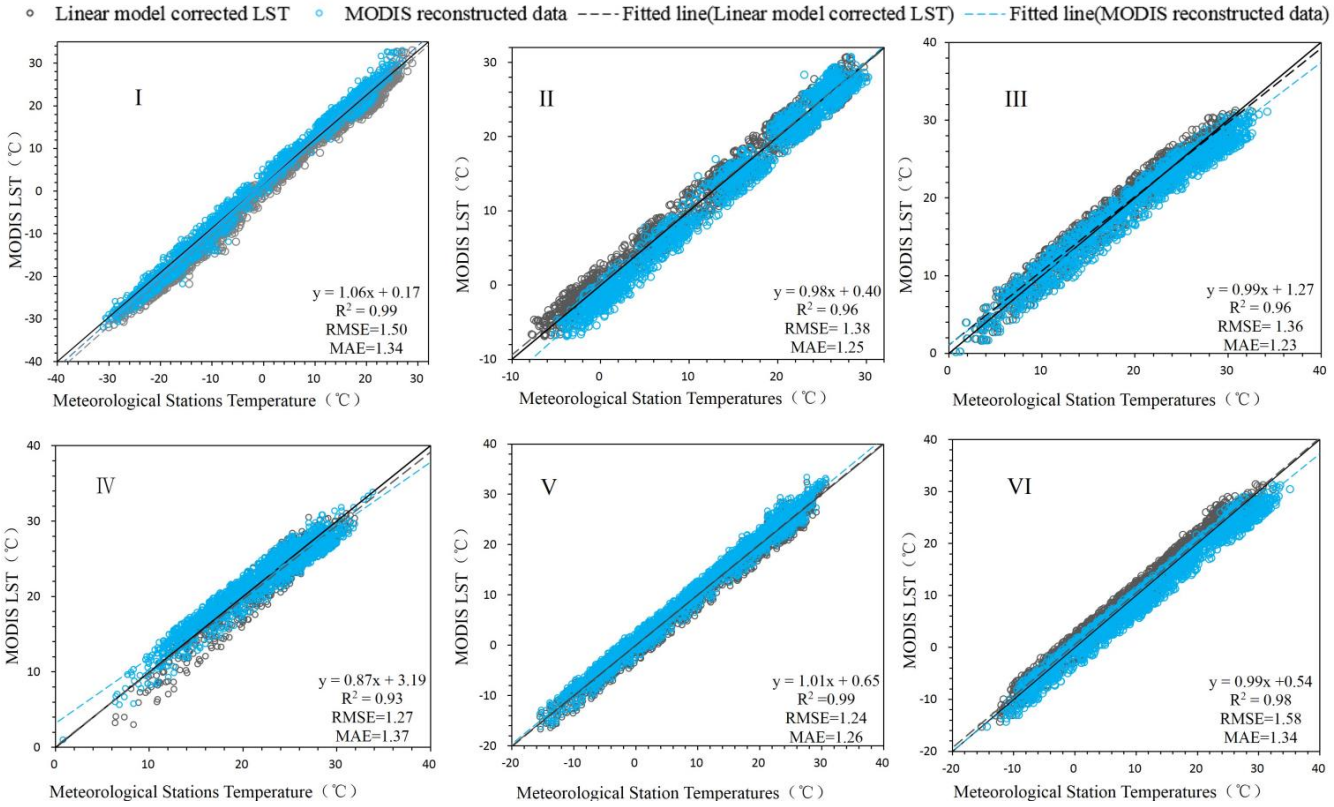


Figure 10: The scatter diagrams in six natural subregions (I, II, III, IV, V, and VI) between the ground station data and the monthly MODIS LST data. The blue points represent the verification results of the reconstructed MODIS LST, and the statistical accuracy measures (R^2, RMSE, and MAE) are also indicated. The results of the corrected linear model are indicated in gray.

Table 1 Basic information for some of the meteorological stations in key zones

Region	Key zone	ID	Latitude (°)	Longitude (°)	Elevation (m)
I Northeast Region	a	50758	47.10	125.54	249
I Northeast Region	a	50658	48.03	125.53	237
I Northeast Region	a	50756	47.26	126.58	239
I Northeast Region	a	50656	48.17	126.31	278
I Northeast Region	a	50548	49.05	123.53	282
II North China Region	b	54525	117.28	39.73	5
II North China Region	b	54527	117.05	39.08	3

II North China Region	b	54518	116.39	39.17	8
II North China Region	b	54511	116.19	39.57	52
II North China Region	b	54624	117.21	38.22	7
II North China Region	b	54623	117.43	38.59	6
IV South China	c	59431	22.63	108.22	122
IV South China	c	59242	23.45	109.08	85
IV South China	c	59037	23.93	108.10	170
IV South China	c	59228	23.32	107.58	108
IV South China	c	59446	22.42	109.30	66
V Northwest Region	d	53336	41.40	108.48	1275
V Northwest Region	d	53446	40.34	109.50	1044
V Northwest Region	d	53602	38.52	105.34	1561
V Northwest Region	d	53513	40.48	107.30	1039
V Northwest Region	e	51730	40.33	81.19	1012
V Northwest Region	e	51716	39.48	78.34	1117
V Northwest Region	e	51810	38.56	77.40	1178
V Northwest Region	e	51811	38.26	77.16	1231
VI Qinghai-Tibet Plateau Region	f	55279	31.48	89.40	4700
VI Qinghai-Tibet Plateau Region	f	55591	29.42	91.08	3648
VI Qinghai-Tibet Plateau Region	f	55598	29.15	91.47	3560
VI Qinghai-Tibet Plateau Region	f	56106	31.53	93.48	4022

Table 2. RMSEs of seasonal LST between monthly LST data (including the original LST data and reconstructed LST data) and ground-based LST data (Orig. indicates original LST values at the ground stations. Recon. indicates the reconstructed LST values at the ground stations)

Region	Key zone	ID	Spring		Summer		Autumn		Winter	
			Orig.	Recon.	Orig.	Recon.	Orig.	Recon.	Orig.	Recon.
I Northeast Region	a	50758	2.11	1.48	1.36	1.23	1.16	0.61	3.80	3.81
I Northeast Region	a	50658	2.33	1.03	1.61	0.63	0.29	0.27	4.32	3.20
I Northeast Region	a	50756	3.51	0.23	1.03	0.43	0.51	0.26	3.91	3.52
I Northeast Region	a	50656	0.65	0.65	0.90	0.92	0.42	0.04	3.63	3.67
I Northeast Region	a	50548	0.82	0.89	1.09	0.61	0.51	0.40	0.15	0.15
II North China Region	b	54525	3.11	2.26	3.30	2.23	2.11	1.51	2.11	0.94
II North China Region	b	54527	1.30	1.11	1.24	1.25	0.93	0.54	2.36	0.14
II North China Region	b	54518	3.64	1.64	0.52	0.51	0.45	0.15	0.71	0.04

II North China Region	b	54511	1.06	1.26	0.33	0.32	0.50	0.66	1.07	1.27
II North China Region	b	54624	1.99	1.55	1.15	0.49	0.84	0.33	0.40	0.46
II North China Region	b	54623	0.13	0.06	0.48	0.17	1.31	1.06	2.65	2.02
IV South China	c	59431	1.71	2.73	0.12	0.06	1.05	1.02	2.91	2.91
IV South China	c	59242	2.0	1.08	2.52	1.86	0.03	0.09	2.91	2.59
IV South China	c	59037	1.08	0.73	1.26	0.94	0.78	0.78	1.00	1.01
IV South China	c	59228	0.92	0.38	1.99	1.75	1.61	0.84	0.75	0.28
IV South China	c	59446	2.01	1.30	0.97	0.78	0.49	0.49	2.40	2.39
V Northwest Region	d	53336	3.88	3.88	3.04	3.04	3.53	2.81	1.90	1.82
V Northwest Region	d	53446	2.00	2.01	3.78	3.18	1.96	1.65	0.35	0.35
V Northwest Region	d	53602	4.48	4.28	3.91	3.75	3.97	3.47	1.65	1.65
V Northwest Region	d	53513	1.55	1.48	5.33	5.15	5.01	4.93	2.04	2.24
V Northwest Region	e	51730	3.01	2.97	4.09	5.08	1.48	1.06	2.63	2.10
V Northwest Region	e	51716	0.80	0.75	0.47	0.15	0.74	0.09	0.66	0.32
V Northwest Region	e	51810	2.33	1.29	1.20	0.76	0.33	0.32	1.24	0.28
V Northwest Region	e	51811	0.57	0.57	0.52	0.90	0.62	0.36	1.34	0.39
VI Qinghai-Tibet Plateau Region	f	55279	3.63	3.44	1.37	1.74	1.83	1.45	0.99	0.99
VI Qinghai-Tibet Plateau Region	f	55591	1.76	1.79	5.56	4.08	2.99	2.59	1.95	0.41
VI Qinghai-Tibet Plateau Region	f	55598	0.85	0.85	4.37	4.62	2.95	2.91	0.63	0.69
VI Qinghai-Tibet Plateau Region	f	56106	0.52	0.58	1.44	1.44	0.88	0.68	2.11	1.99
Average			1.92	1.51	1.96	1.72	1.40	1.12	1.88	1.49

Appendix A

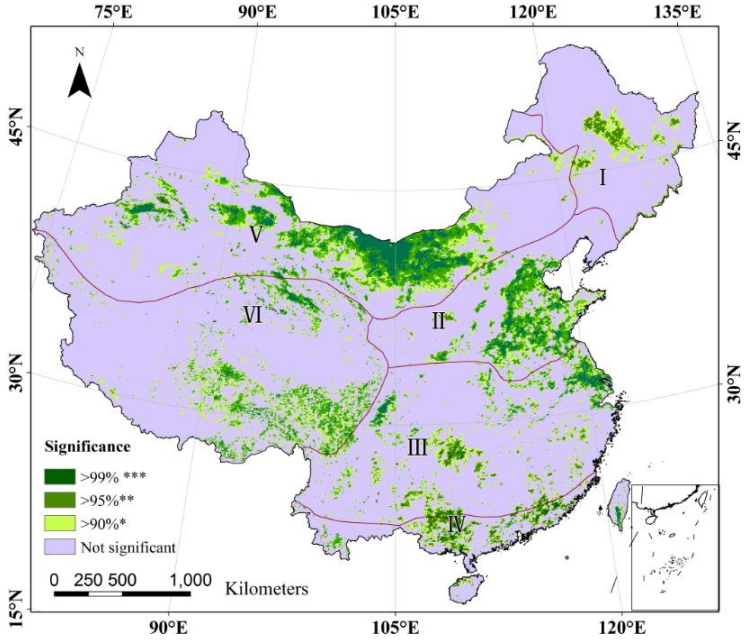


Figure A1: Significance of the spatial distribution of annual average LST trends based on an independent-samples t-test in China from 2003 to 2017. Note that the symbols *, **, and * explain that there are increasing/ decreasing tendencies averaged over China at the 99 %, 95 % and 90 % confidence level, respectively (same as below).**

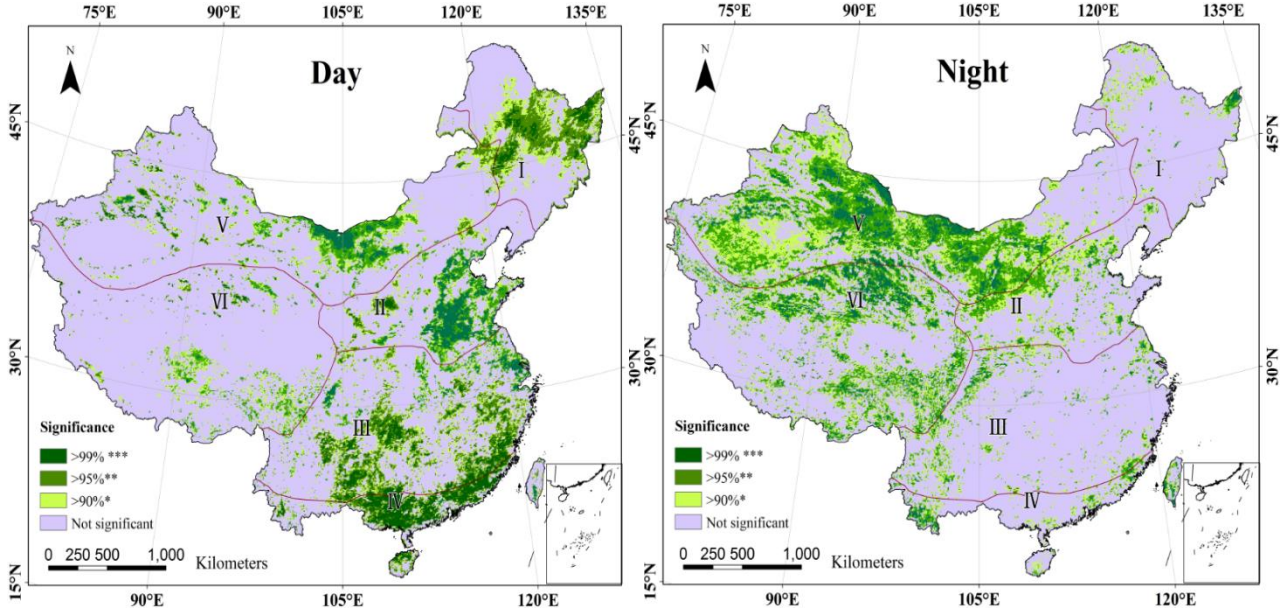


Figure A2: Distribution of diurnal LST change trends significance during 2003 - 2017

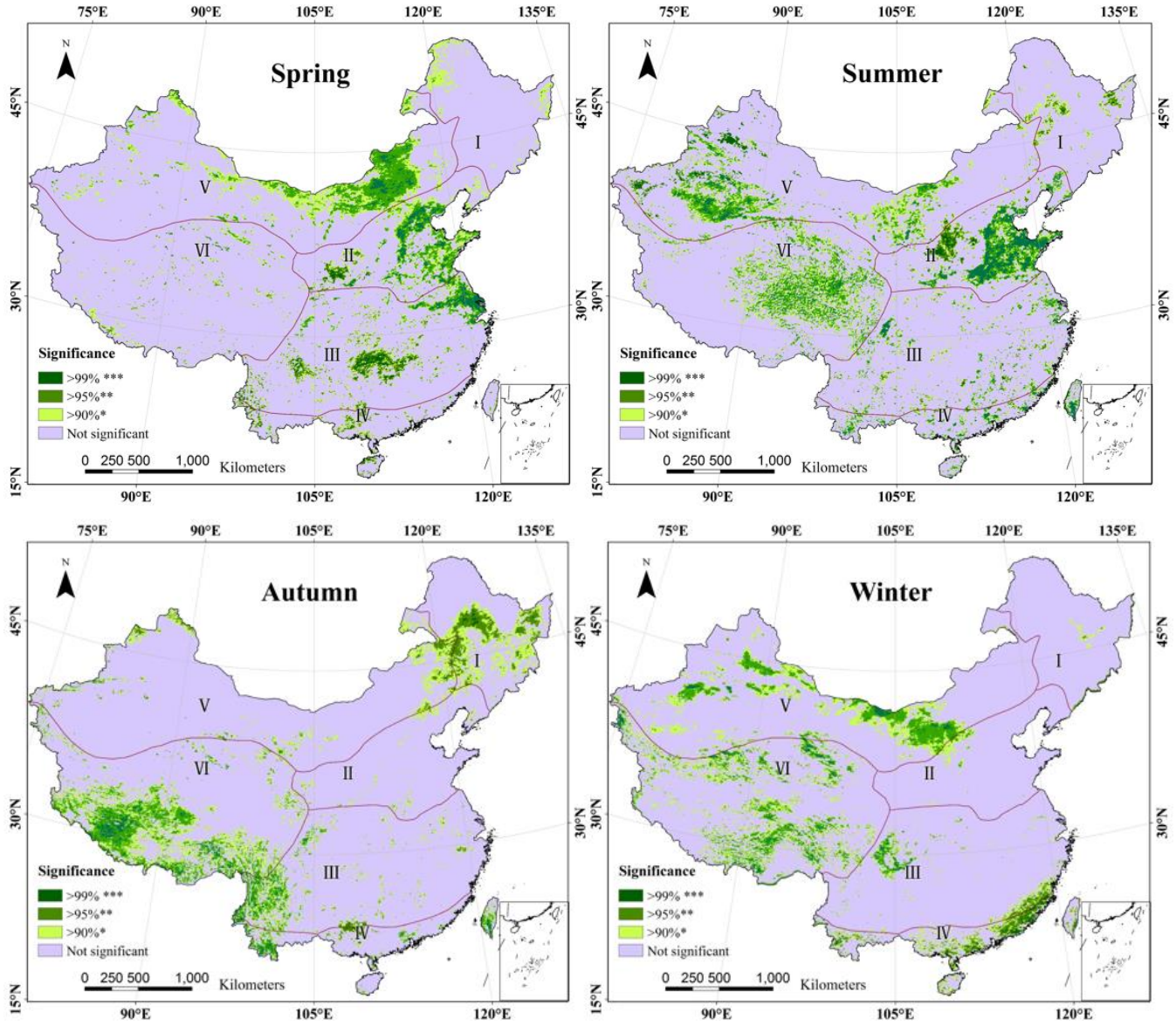


Figure A3: Distribution of seasonal LST trends significance during 2003 - 2017.