Remote Sensing of Environment 277 (2022) 113070

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

# A two-step deep learning framework for mapping gapless all-weather land surface temperature using thermal infrared and passive microwave data

Penghai Wu[a,b], Yang Su[a], Si-bo Duan[c,*], Xinghua Li[d,*], Hui Yang[b], Chao Zeng[e], Xiaoshuang Ma[a], Yanlan Wu[a,b], Huanfeng Shen[e]

[a] *School of Resources and Environmental Engineering, Anhui University, Hefei, Anhui 230601, China*
[b] *Information Materials and Intelligent Sensing Laboratory of Anhui Province, Anhui University, Hefei, Anhui 230601, China*
[c] *Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China*
[d] *School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, Hubei 430079, China*
[e] *School of Resource and Environmental Sciences, Wuhan University, Wuhan, Hubei 430079, China*

## ARTICLE INFO

Edited by Jing M. Chen

*Keywords:*
All-weather
Gapless
Satellite remote sensing
Land surface temperature
Deep learning

## ABSTRACT

Blending data from thermal infrared (TIR) and passive microwave (PMW) measurements is a promising solution for generating the all-weather land surface temperature (LST). However, owing to swath gaps in PMW data and the resolution inconsistence between TIR and PWM data, spatial details are often incomplete or considerable losses are generated in the all-weather LST using traditional methods. This study was conducted to develop a two-step deep learning framework (TDLF) for mapping gapless all-weather LST over the China's landmass using MODIS and AMSR-E LST data. In the TDLF, a multi-temporal feature connected convolutional neural network bidirectional reconstruction model was developed to obtain the spatially complete AMSR-E LST. A multi-scale multi-temporal feature connected generative adversarial network model was then designed to blend spatially complete AMSR-E LST and cloudy-sky MODIS LST, and generate gapless all-weather LST data. Gapless all-weather LST data were evaluated using six in-situ LST data from the Tibetan Plateau (TP) and the Heihe River Basin (HRB). The root mean squared errors (RMSEs) of the gapless all-weather LST were 1.71–2.0 K with determination coefficients ($R^2$) of 0.94–0.98 under clear conditions, and RMSEs of 3.41–3.87 K and $R^2$ of 0.88–0.94 were obtained under cloudy conditions. Compared to the existing PMW-based all-weather LSTs, the validation accuracy and image quality (such as spatial detail) of the generated gapless all-weather LSTs were superior. The TDLF does not require the use of any additional data and can potentially be implemented with other satellite TIR and PWM sensors to produce long-term, gapless, all-weather MODIS LST records on a global scale. Such a capability is beneficial for generating further gapless all-weather soil moisture and evapotranspiration datasets that can all be applied in global climate change research.

## 1. Introduction

Land surface temperature (LST) is one of the most significant indexes to evaluate energy exchanges of land-atmosphere system and to reveal climate changes (Hansen et al., 2010; Tierney et al., 2008). Spaceborne remote sensing platforms provide a unique and efficient approach for acquiring multi-scale LST over large regions, particularly over inaccessible areas or those with dangerous conditions, and their results have an acceptable accuracy (Li et al., 2013). As a result, satellite-derived LST products have been widely utilized across various scientific disciplines, including meteorology, climatology, hydrology, ecology, and epidemiology, both globally and in specific areas (Guillevic et al., 2017; Liu and Weng, 2012; Trigo et al., 2008).

Satellite-based LST is mainly derived from two types of measurements: thermal infrared (TIR) measurements and passive microwave (PMW) measurements. TIR LST data has attracted significant attention due to its comparatively fine spatial resolution and retrieval accuracy (Li et al., 2013; Prata et al., 1995). Over the past two decades, various operational TIR LST products (e.g., TIRS, ASTER, MODIS, VIRR, GLASS, and SEVIRI LST) have been available free of charge and used extensively (Göttsche and Hulley, 2012; Guillevic et al., 2017; Ma et al., 2020a). However, owing to certain effects from the atmosphere and clouds, data

---

* Corresponding authors.
*E-mail addresses:* duansibo@caas.cn (S.-b. Duan), lixinghua5540@whu.edu.cn (X. Li).

https://doi.org/10.1016/j.rse.2022.113070
Received 24 October 2021; Received in revised form 23 April 2022; Accepted 29 April 2022

are missed in TIR LST products (Holmes et al., 2016; Mo et al., 2021; Wu et al., 2021). For instance, approximately 67% of MODIS LST products are covered by clouds, with higher percentages occurring in humid regions. PMW radiation can penetrate non-precipitating clouds and is influenced minimally by atmospheric water vapor absorption (Prigent et al., 2016); therefore, in the aspect of the spatial completeness, the derived PMW LST product is superior to TIR LST products (Huang et al., 2019). However, compared to TIR measurements, PMW measurements are limited to a lower retrieval accuracy and spatial resolution. Therefore, the applications of PMW and TIR LST products are strongly limited in areas where all-weather LST products at a moderate/high spatial resolution are urgently needed (Duan et al., 2017).

More and more kinds of methods have been proposed to reconstruct missing TIR LST data (Mo et al., 2021; Wu et al., 2021), which mainly include four categories: (1) spatial and/or temporal interpolation methods (Crosson et al., 2012; Kilibarda et al., 2014; Liu et al., 2019; Wang et al., 2019), (2) statistical regression methods (Li et al., 2021; Shwetha and Kumar, 2016; Zhang et al., 2020b; Zhao and Duan, 2020), (3) surface energy balance methods (Fu et al., 2019; Jia et al., 2021; Jin, 2000; Martins et al., 2019; Zeng et al., 2018), and (4) passive microwave (PMW)-based LST methods (Kou et al., 2016; Sun et al., 2019; Xu and Cheng, 2021; Xu et al., 2019). The first three methods usually link cloud-free TIR LST with additional information (such as DEM, NDVI, downward shortwave radiation, and near-surface meteorological observations) via empirical/semi-empirical functions or regression models, and these are subsequently applied to derive cloudy LST. However, when additional information can be used to discriminate between clear-sky and cloudy conditions, the estimation of LSTs under cloudy conditions is acceptable; if not, these methods only provide hypothetical LST values (Kang et al., 2018; Li et al., 2014; Li et al., 2018; Sun et al., 2017). In addition, although the additional information is usually widely available, its supportive use is not always adequate, especially over large areas, those with heterogeneous surfaces, or when extensive amounts of data are missing (Yang et al., 2019; Yu et al., 2019).

PMW LST data usually reflect the actual surface thermal statuses in cloudy sky circumstances and can be considered as a proxy for the actual LST under cloudy conditions (Xu and Cheng, 2021). The blending of TIR and PMW measurement data has thus became a potential direction for generating all-weather LST; in particular, the use of MODIS and AMSR-E aboard the Aqua satellite, as these have the same acquisition time. In recent years, several studies have blending MODIS and AMSR-E/AMSR-2 LST based on empirical and semi-empirical (Duan et al., 2017), physical decomposition (Zhang et al., 2019), nonlinear geostatistical (Kou et al., 2016; Xu and Cheng, 2021), and machine-learning models (Sun et al., 2019; Xu et al., 2021). However, although progress has been made, the PWM-based methods have three remaining issues that need to be resolved, and these are as follows: 1) the sensing depth mismatch between TIR and PMW measurement. As the PMW not only penetrates atmosphere but also the vegetation, the temperature difference between soil and vegetation vary over time (Zhao et al., 2020b). Subsurface temperatures retrieved from PMW differ from LST obtained from TIR sensors and requires conversion to surface temperature. This work is a very challenging problem to solve, and is beyond the scope of this article. 2) there is few special design method (to our knowledge) to fill swath gaps of PMW data observed from the middle and low latitude regions, and to thus provide gapless all-weather available LST (Duan et al., 2017; Wu et al., 2021). Previous studies have adopted the inverse distance weighted (IDW) (Duan et al., 2017) and the empirical orthogonal function (DINEOF) (Xu and Cheng, 2021) methods, but the performance has been shown to degrade with large widths of swath gaps. 3) The spatial resolution inconsistence between TIR and PMW LST requires an effective downscaling method to enable the retention of greater textural details. In this respect, empirical models with various auxiliary parameters (e.g., DEM, NDVI, and surface albedo) have been used, and the downscaling effect is very sensitive to these parameters. However, it is acknowledged that the distribution of LSTs is spatially and temporally inhomogeneous and non-stationary owing to the land surface heterogeneity and the surface type changes. Moreover, the scale relationship between TIR and PMW LST is also typically nonlinear and needs to be appropriately specified. Therefore, further research is necessary to provide an efficient framework that integrates the TIR and PMW LST to ensure the quality of all-weather LST.

Inspired by the powerful ability of nonlinear representation of deep learning (DL) architectures, tremendous achievements have been made in conducting DL-based environmental remote sensing studies (Yuan et al., 2020), such as those relating to LST retrieval (Wang et al., 2021), LST reconstruction (Wu et al., 2019) and LST spatiotemporal fusion (Yin et al., 2021). For the generation of gapless all-weather LST, a DL-based framework with an enhanced nonlinear representation will be valuable. Therefore, this article proposes a two-step deep learning framework (TDLF) for mapping gapless all-weather LST data. Our key contributions are as follows: 1) we developed a multi-temporal feature connected convolutional neural network (MTFC-CNN) bidirectional reconstruction model with a sample enhancement strategy to obtain spatially complete AMSR-E LST, and 2) we designed a multi-scale multi-temporal feature connected generative adversarial network model (MSTFC-GAN) to blend spatially complete AMSR-E LST and cloudy-sky MODIS LST, with the aim of generating gapless all-weather LST data.

The remainder of this article is arranged as follows. The study area and the data are given in Section 2, and Section 3 introduces the study's methodology. Results are reported in Section 4 and followed by a discussion in Section 5, and the conclusions are shown in Section 6.

## 2. Study area and data

### *2.1. Study area*

The China's mainland was our study area. LST distribution of the study area is largely varied due to topographic and climatic differences (Zhao et al., 2020a, 2020b). To validate the method, sub-areas in two geographically different locations containing a total of six ground sites were selected. The first validation sub-area is within TP region (validation range of 28.4°–33.45°N and 87.7°–95.73°E), which is sparsely covered with grassland. The second validation sub-area is located in HRB region at the border of Qinghai and Gansu provinces (validation range of 36.93°–40.01°N and 98.38°–101.46°E), where the land types are farmland, forest, and sparse grassland. The corresponding land cover type map and images of two validation sub-areas are shown in Fig. 1, and the site locations are marked with red triangles.

### *2.2. Satellite data*

The daily MYD11A1 version 6 product with 1 km spatial resolution was regarded as the TIR LST, and this is available from the following website address of National Aeronautics and Space Administration (https://ladsweb.nascom.nasa.gov/search). The MYD11A1 product was derived from Aqua MODIS, which had approximate observation times of 13:30 and 01:30 (local solar time) in the ascending and descending orbits, respectively and was produced using the generalized split-window model (Wan, 2014; Wan and Dozier, 1996). The version 6 MODIS LST product had been validated in many published studies, and its quality is superior to other versions. In addition to the LST images, the product also include other important information layers, such as observation time layers, observation angles layers, emissivities layers, quality control (QC) layer, and so on. The QC layer in the products was used to recognize LST pixels of high quality.

The selected PMW LST was obtained from the AMSR-E sensor with 25 km spatial resolution. The AMSR-E is also aboard on Aqua satellite, and therefore an excellent temporal collocation exists between AMSR-E PMW and the MODIS TIR. The AMSR-E LST data was retrieved from the AMSR-E BT data L3 product (NISDC-0302) using an empirical look-up table based method (Zhang and Cheng, 2020), and NISDC-0302 was

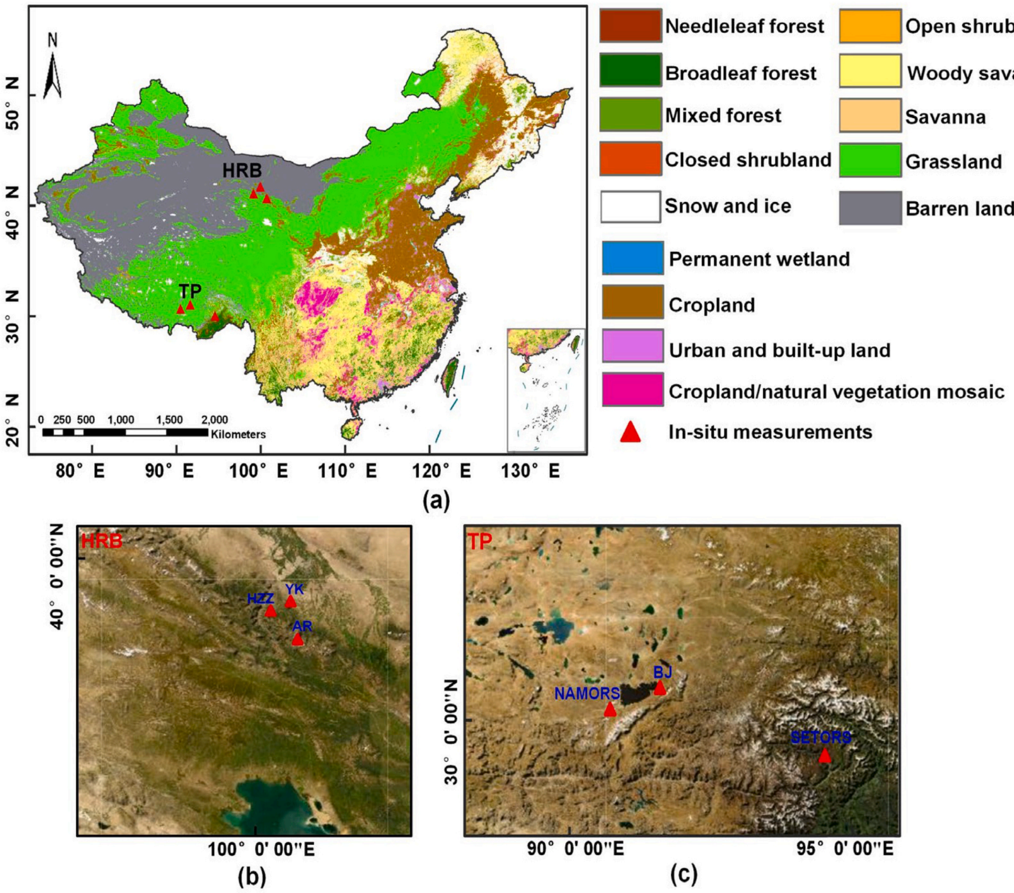


**Fig. 1.** (a) Land cover map of the China's mainland, and images of the two validation sub-areas: (b) Heihe River Basin (HRB) and (c) Tibet Plateau (TP). The land cover map is generated from the MCD12Q1 product and includes 14 land cover types. The site locations are marked with red triangles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

downloaded from the National Snow and Ice Data Center (http://nsidc.org). The retrieved AMSR-E LST in 2010 was fully validated with MODIS and ground measured data by Zhang and Cheng (2020).

### 2.3. *Ground measured data*

To assess the generated gapless all-weather LST data, six in-situ measurements were collected within the Tibetan Plateau (TP) and the Heihe River Basin (HRB) regions in 2010. The corresponding data were from the Watershed Allied Telemetry Experimental Research (Li et al., 2009) and the Second Tibetan Plateau Scientific Expedition and Research (Ma et al., 2020b) programs at the National Tibet Plateau Data Center (http://www.tpdc.ac.cn/zh-hans/data). The Arou station (AR), Huazhaizi station (HZZ), and Yingke station (YK) in the HRB were selected, and the observation data were uniformly organized into 30-min sampling periods for storage. In the TP region, Naqu-BJ station (BJ), NAMORS station, and SETORS station were selected, and the observation data were uniformly organized into 60-min sampling periods for storage. Details on the six sites is presented in Table 1.

The in-situ LST ($T_s$) was calculated using the Stefan–Boltzmann law,

$$T_s = \left[\frac{F^{\uparrow} - (1-\varepsilon_b)F^{\downarrow}}{\sigma\varepsilon_b}\right]^{1/4} \tag{1}$$

where $F^{\uparrow}$denotes the surface upwelling longwave radiation, and $F^{\downarrow}$ denotes atmospheric downwelling longwave radiation; which are measured by the long-wave radiation meter of a four-component radiation sensor, respectively; $\sigma$ is the Stefan-Boltzmann constant ($5.67 \times 10^{-8}$ W·m$^{-2}$ K$^{-4}$), and $\varepsilon_b$ is surface emissivity, calculated with the following expression (Wang et al., 2005),

**Table 1**
Details on the six sites.

| Name | Latitude (degree) | Longitude (degree) | Elevation (meter) | Temporal Resolution (minute) | Land Cover |
|---|---|---|---|---|---|
| HZZ | 38.77 | 100.32 | 1726 | 30 | Desert steppe |
| AR | 38.05 | 100.46 | 3033 | 30 | Alpine Meadow |
| YK | 38.85 | 100.41 | 1519 | 30 | Cropland |
| SETORS | 29.76 | 94.73 | 3326 | 60 | Grassland |
| NAMORS | 30.77 | 90.98 | 4730 | 60 | Alpine meadow |
| BJ | 31.37 | 91.90 | 4450 | 60 | Cropland |

$$\varepsilon_b = 0.2122\varepsilon_{29} + 0.3859\varepsilon_{31} + 0.4029\varepsilon_{32} \tag{2}$$

where$\varepsilon_{29}$, $\varepsilon_{31}$, and $\varepsilon_{32}$ are surface emissivities of MODIS band29, band31, and band32, respectively.

## 3. Methodology

A flow chart of the proposed TDLF that includes the generation of the spatially complete PWM LST and the gapless all-weather TIR LST is shown in Fig. 2, specifically as follows: 1) For generating the spatially complete PWM LST, the MTFC-CNN model was first used to obtain spatially complete AMSR-E LST, and sample enhancement technology

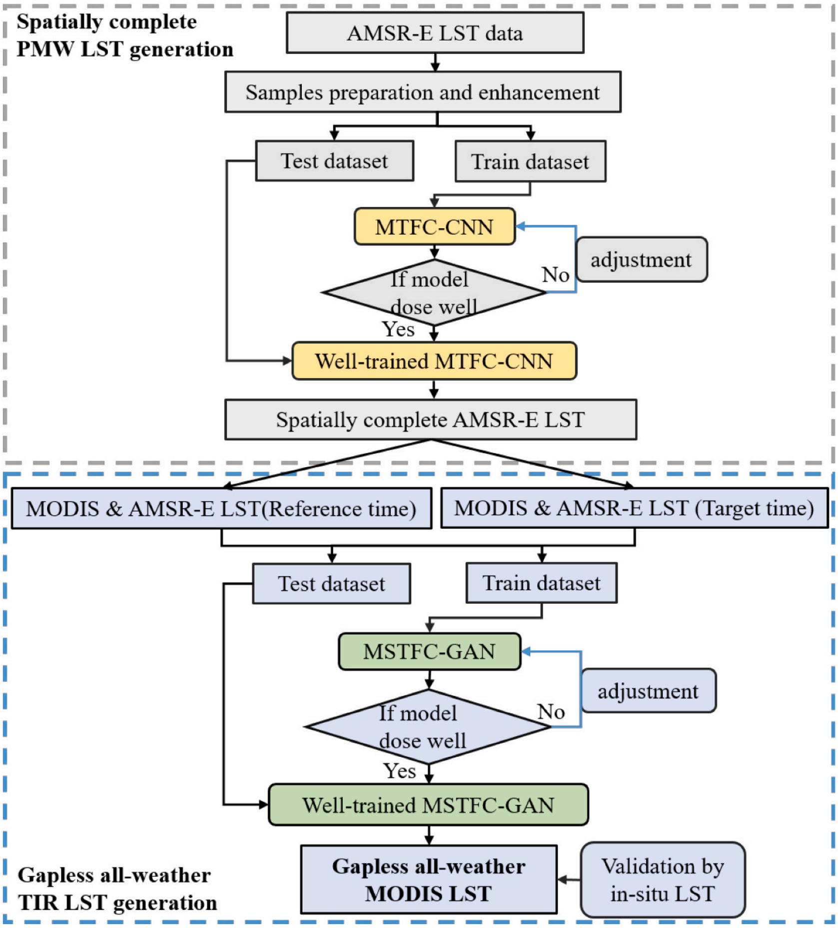


**Fig. 2.** Flow chart of the proposed two-step deep learning framework (TDLF) for mapping gapless all-weather LST, which is composed mainly of spatially complete PWM LST generation and gapless all-weather TIR LST generation.

was then designed to solve sample shortages. A training dataset was subsequently employed to optimize parameters of the MTFC-CNN model. The test dataset was put into the well-trained MTFC-CNN model to generate spatially complete PWM LST. 2) To generate gapless all-weather TIR LST, the cloudy-sky MODIS LST with missing pixels and the corresponding AMSR-E LST at the target time, and MODIS LST without missing pixels and corresponding AMSR-E LST at the reference time were used to create the training and test datasets. The former was consequently employed to optimize parameters of the MSTFC-GAN model. The latter was subsequently put into the well-trained MSTFC-GAN model to obtain the gapless all-weather MODIS LST. Furthermore, in-situ LSTs were collected to validate the gapless all-weather LST.

### 3.1. Spatially complete PWM LST generation

Considering periodical changes in the orbit swath gap and the complementary spatial information obtained from adjacent AMSR-E data, we assumed that the variation in LST on two adjacent days inside gaps would be similar to that outside gaps within a local area (Su et al., 2022). We also assumed that nonlinear variations could be learned if sufficient samples were employed. In this respect, a bidirectional reconstruction model based on MTFC-CNN with a sample enhancement strategy was designed to obtain the spatially complete AMSR-E LST.

#### 3.1.1. Sample preparation and enhancement

The distribution of LST markedly varies in each season and diurnally over the China's mainland. In this respect, the AMSR-E LST data obtained in 2010 were divided into eight data subsets (spring, summer, autumn, and winter; daytime and nighttime). In each subset, the AMSR-E LST of three consecutive days was selected as the data at times T1, T2, and T3, respectively, and uniformly cropped into 24 × 24 size patch pairs, with 12 overlapping. Patches without invalid values were selected as sample datasets; otherwise, they were treated as mask datasets. Sample datasets were normalized to a range of 0–1, which was convenient for training the MTFC-CNN network. Mask datasets were used to simulate swath gaps for a spatially complete image patch at time T2. On the consideration of convergence and generalization of the network, training and testing datasets were randomly selected from the sample datasets at a ratio of 7:1.

As locations of swath gaps vary for AMSR-E LST data in a single time cycle, it was acknowledged that sequential cropping (i.e., from top left to bottom right) could result in unequal numbers of training samples for different patches. Therefore, when the number of training samples was insufficient for one image patch, a sample enhancement strategy was employed, and other cropping methods (e.g., from top right to bottom left) with the same size (i.e., 24 × 24) and overlapping (i.e., 12) were also implemented.

#### 3.1.2. Network structure of MTFC-CNN

As shown in Fig. 3, the main structure of the MTFC-CNN has three components: combining multi-temporal data, and conducting down-sampling and up-sampling procedures. For the data combination part, the MTFC-CNN uses two kinds of inputs: the AMSR-E LST at time T2 with simulated missing pixels and auxiliary AMSR-E LSTs without missing pixels at times T1 and T3. To better utilize all inputs, auxiliary data (i.e. T1 and T3) and target data (i.e., T2) were first added. We then

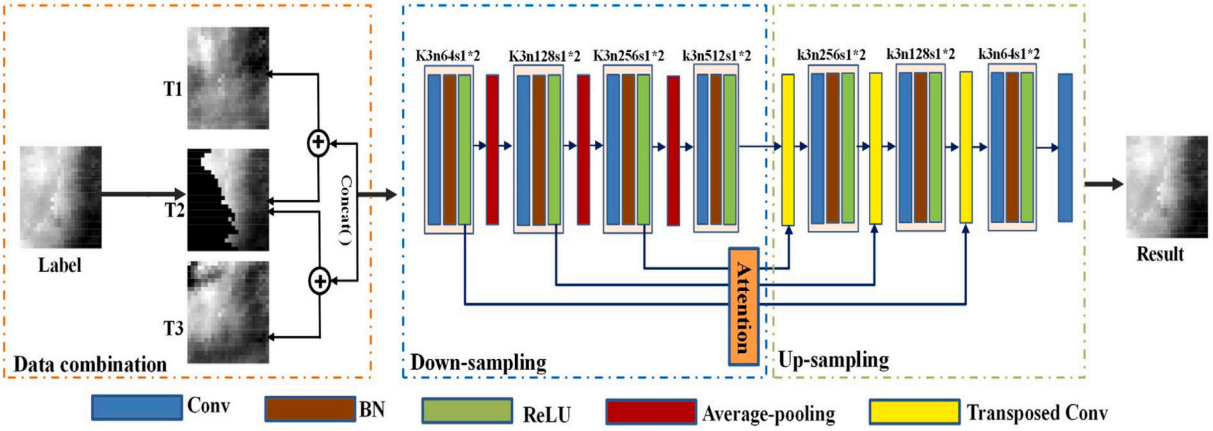


**Fig. 3.** Network architecture of multi-temporal feature connected convolutional neural network (MTFC-CNN) with kernel size (abbr. K), feature maps number (abbr. n), and stride of convolutional layer (abbr. s).

composited the three inputs and the two added results. This enabled spatiotemporal information to be utilized simultaneously in MTFC-CNN.

For the down-sampling part, a four-layer convolutional network was adopted to extract the feature information. A batch normalization (BN) layer is added between the convolutional layers to accelerate the convergence speed and parameter fitting of the model. The activation function was set as rectified linear unit (ReLU), which is used to ensure that the output of the convolutional layers is a nonlinear combination of the inputs. Additionally, to make the model suitable for filling large gaps, it is necessary to extend the perceptual field in the convolution process. The average pooling layer with a convolution kernel size of 2 × 2 was used to enlarge the perceptual field of convolution kernel for enhancing the contextual information of feature maps. For the up-sampling part, a transposed convolution was selected to recover the feature map. As same as the down-sampling process, the BN and ReLU were also adopted in the convolutional layers. During the last convolution layer, preceding feature maps were combined, and the high-level information output was the same size as that input.

It is noteworthy that spatial information of the AMSR-E LST is lost during the down-sampling process. To reuse the spatial information, the jump connection structure is generally used to directly connect low-level features to high-level features. However, overfitting occurs easily if very low-level features are involved. Therefore, a spatial attention mechanism was introduced to solve the problem. The weight of the low-level features was output by activating the valid information in high-level features. More important spatial information was extracted to gain the weighted low-level features. The different levels of features were then combined.

#### 3.1.3. *Training process*

During the training, the image patches at time T1 and T3, and the simulated image patch at time T2, were used as the input, while the original spatially complete image patch at time T2 was treated as labels. A weight initialization method by He et al. (2015) was used to output random values from a truncated normal distribution (0, 1 × $10^{-3}$). The initial learning rate and epoch were set to 1 × $10^{-4}$ and 90, respectively, and after every 30 epochs, learning rate is multiplied by a decay factor of 0.1 to reduce parameters' ranges.

As the LST values was the same of output image as that of input image outside the missing area, the residual learning strategy was used in the training process, and the backpropagation mechanism was applied to update training parameters and better determine the association between LSTs. The loss function of the MTFC-CNN can be expressed as Eq. (3),

$$\mathscr{L}' = \frac{1}{WH}\sum_{x=1}^{W}\sum_{y=1}^{H}\left((L_{T2})_{x,y} - F(L_{T1}, L_{T2\_m}, L_{T3})_{x,y}\right)^{2} \tag{3}$$

where $\mathscr{L}'$ is the loss function of training; $W$and $H$ represent dimensions of the images; $L_{T1}$, $L_{T2}$, and $L_{T3}$ are the spatially complete LST image patches at T1, T2, and T3, respectively; and $L_{T2\_m}$, $F(L_{T1}, L_{T2\_m}, L_{T3})_{x,\,y}$ are the simulated missing image patch with swath gaps and the reconstructed image patch at T2, respectively.

### 3.2. *Gapless all-weather TIR LST generation*

It is acknowledged that the image super-resolution (SR) can obtain a fine-resolution (FR) image from its corresponding coarse-resolution (CR) counterpart. Satisfactory results have been obtained for the SR using deep learning methods, particularly with the advent of generative adversarial networks, such as SR-GANs. The idea of SR-GAN provided the motivation for generating gapless all-weather LST.

If we treat the spatially complete AMSR-E and MODIS LST as the CR image and its corresponding FR image, the relationship between the FR and CR images can be modeled with GAN from a large number of sample pairs. Based on this relationship, it should then be possible to estimate the cloudy-sky MODIS LST at the target time using the corresponding spatially complete AMSR-E LST. However, the obvious scale difference between MODIS and AMSR-E observations causes a considerable loss of spatial detail for the reconstructed MODIS LST. Fortunately, some widely used spatiotemporal LST fusion methods have enabled spatial details to be obtained from fine-resolution LSTs at a reference time (Wu et al., 2015; Wu et al., 2021). Inspired by spatiotemporal fusion, a new relationship can be modeled using a large number of sample pairs from both the target and reference times. Therefore, we developed and proposed the MSTFC-GAN to generate the gapless all-weather MODIS LST.

#### 3.2.1. *Network structure of MSTFC-GAN*

As shown in Fig. 4, the main structure of the MSTFC-GAN includes a generative network and a discriminant network, and a detailed description of each network is provided below.

1) Generative network: We designed a multi-scale multi-temporal feature-connected CNN as the generative network. As shown in Fig. 4 (a), except for the inputs and the number of convolutional layers, the generator network is similar to the MTFC-CNN. For the inputs, the MODIS LST and spatially complete AMSRE LST data at reference time (T1) and target time (T0) were subtracted to obtain referenced cloudless data ($Sub_{T1}$) and targeted cloudless data ($Sub_{T0}$),as training labels, respectively. $Sub_{T0_C}$represents the simulated cloudy-sky data

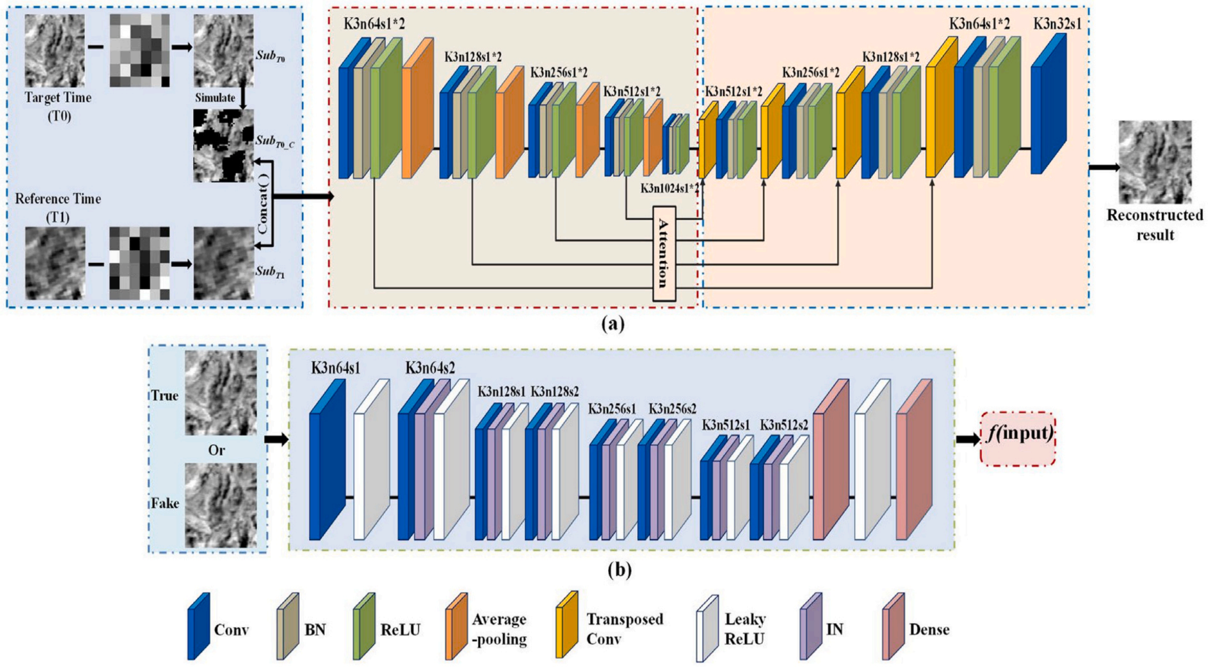


**Fig. 4.** Network architecture of MSTFC-GAN: (a) generator network, (b) discriminator network, with the kernel size (abbr. K), number of feature maps (abbr. n), and stride of convolutional layer (abbr. s).

of $Sub_{T0}$. $Sub_{T1}$ and $Sub_{T0\_C}$ were consequently combined to obtain more dynamic relations from multi-scale multi-temporal LST data. In addition, a five-layer convolutional network was used to extract features of the combined multi-scale multi-temporal data, while a four-layer convolutional network was used in the MTFC-CNN.

2) Discriminant network: A discriminant network was used to discriminate actual images (True) from reconstructed images (Fake). As the Fig. 4(b) shows, the discriminator involves eight convolutional layers where the filter kernels are gradually enlarged from 64 to 512. These layers could extract the LST features, improved the discriminant accuracy, and generated reconstructed images with greater realism. The final feature images were followed by two dense layers to determine the probability of sample classification. Unlike the generator, a leaky rectified linear unit activation function is used in the discriminator, and the gradient during the back-propagation process was calculated when the input was less than zero. An instance normalization (IN) layer was selected to compute a single sample. The Wasserstein GAN (WGAN) was adopted to guide the training process with a loss value and get the optimum generator (Arjovsky et al., 2017).

### 3.2.2. Dataset preparation and training process

The spatially complete AMSR-E LST images were first resampled to match MODIS LST data. All resampled AMSR-E and MODIS LST data in 2010 for MSTFC-GAN were also divided into eight data subsets. In each subset, the AMSR-E and the corresponding MODIS LST were uniformly cropped into 64 × 64 size patch pairs, with 32 overlapping. If the MODIS data were spatially complete in the data pairs, they were treated as sample datasets; otherwise, as mask datasets. Mask datasets were used to simulate cloudy-sky MODIS LST images. The training and testing datasets were normalized, and randomly selected from the sample datasets at a ratio of 6:1.

For optimization, the initial learning rate and epoch were set to $1 \times 10^{-4}$ and 60, respectively. To reduce the search range of parameters, learning rate was multiplied by a decay factor of 0.1 after every 20 epochs. To avoid network overfitting, $L_2$ regularization with a weight decay of $1 \times 10^{-3}$ was applied. Although WGAN is able to progress the stable training of GANs, it still has difficulty in convergence (Gulrajani et al., 2017). Therefore, the WGAN with a gradient penalty term (WGAN-GP) was used as a more stable algorithm for training GANs.

### 3.2.3. Loss function of MSTFC-GAN

The loss function of MSTFC-GAN contains three aspects: mean square error (MSE) loss, visual geometry group (VGG) loss and critic loss. The MSE loss ca*W*n be expressed as:

$$\mathscr{L}_{MSE} = \frac{1}{WH}\sum_{x=1}^{W}\sum_{y=1}^{H}\left((Sub_{T0})_{x,y} - G(Sub_{T0\_C}, Sub_{T1})_{x,y}\right)^2 \tag{4}$$

where and represent the dimensions of the actual and reconstructed images, respectively, and $G(Sub_{T0\_C}, Sub_{T1})$ is the reconstructed result at T0 and can be calculated by a non-linear linkage between $Sub_{T0\_C}$ and $Sub_{T1}$.

However, the MSE loss was inadequate for reconstructing texture details. Instead, the VGG loss was closer to perceptual similarity and provided solutions that had a higher structural similarity, and can be expressed as:

$$\mathscr{L}_{VGG/i,j} = \frac{1}{W_{i,j}H_{i,j}}\sum_{x}^{W_{i,j}}\sum_{y}^{W_{i,j}}\left(\varphi_{i,j}(Sub_{T0})_{x,y} - \varphi_{i,j}(G(Sub_{T0\_C}, Sub_{T1}))_{x,y}\right)^2 \tag{5}$$

where $\varphi_{i,\,j}$ represents the feature map extracted from the *j*th convolution layer and the *i*th pooling layer, and is usually obtained from the encapsulated function within the well-trained VGG network; and $W_{i,\,j}$ $H_{i,\,j}$ are the dimensions of the corresponding feature maps $\varphi_{i,\,j}$The content loss unites the advantages of MSE loss and VGG loss functions to recover both high-and low-frequency information, and it enables the better reconstruction of images. The critic loss is defined as:

$$\mathscr{L}_{Critic} = \frac{1}{N}\sum_{x=1}^{N} D(G(Sub_{T0\_C}, Sub_{T1})) - \frac{1}{N}\sum_{x=1}^{N} D(Sub_{T0}) + \lambda \underbrace{\frac{1}{N}\sum_{x=1}^{N}\left(\left(\|\Delta_{\widehat{\phi}} D(\widehat{\phi})\|_2 - 1\right)^2\right)}_{Gradient\ penalty} \tag{6}$$

$$\widehat{\phi} = \epsilon Sub_{T0} + (1-\epsilon)G(Sub_{T0\_C}, Sub_{T1}) \tag{7}$$

where $N$ denotes the number of training samples; $D$ is the output result of the discriminator; $\lambda$ is set to 10; $\widehat{\phi}$ is calculated in Eq. (7), which weighs the actual and generated samples; $\Delta_{\widehat{\phi}}$ is the gradient operator of $\widehat{\phi}$, and is a random number $\epsilon \sim U$ [0,1].

The generator ($G$) and discriminator ($D$) were trained for the min-max optimization problem, where $G$ minimized the difference between the generated images and actual images, and $D$ maximized the probability of treating a fake result as an actual sample. The $\alpha$ and $\mu$ in Eq. (8) are balanced parameters that were empirically set to $2 \times 10^{-7}$ and $1 \times 10^{-3}$, respectively,

$$\underset{G}{Min}\underset{D}{Max}\mathscr{L} = \mathscr{L}_{MSE} + \alpha\mathscr{L}_{VGG} + \mu\mathscr{L}_{Critic} \tag{8}$$

## 4. Results

### 4.1. Effectiveness of using the MTFC-CNN

The effectiveness of the MTFC-CNN was validated by simulation and real experiments using AMSR-E LST data from 2010. The scatterplots of actual AMSR-E LST during daytime and nighttime against that reconstructed with MTFC-CNN in the simulation experiments are shown in Fig. 5. During the daytime, the averages of the bias, RMSE, and $R^2$ were 0.04 K, 1.21 K and 0.96, respectively; while those values during the nighttime were − 0.03 K, 0.99 K and 0.93, respectively. It is should be noted that these statistics are calculated by only the data points over the gaps. These results indicate the MTFC-CNN can effectively reconstruct the simulated swath gaps of AMSR-E LST data, although the RMSEs add additional uncertainty.

A comparison between the spatial distribution patterns of original AMSR-E LST and reconstructed AMSR-E LST of real experiments during the daytime is shown in Fig. 6. Compared with the original AMSR-E LST, the reconstructed AMSR-E LST seems to reflect the natural spatiotemporal variations in LST, indicating that the MTFC-CNN method could be used to successfully reconstruct large swath gaps of AMSR-E LST in real experiments.

In consideration of the coarse spatial resolution of AMSR-E LST, corresponding MYD11A1 images in 2010 were selected for evaluation during the daytime and nighttime in real experiments. MYD11A1 images excluded low-quality pixels relating to the QC layer and the spatial resolution were resampled to 25 km. The processed MYD11A1 images were adopted to evaluate the reconstructed AMSR-E LST and the observed AMSR-E LST (outside the gaps), and the scatterplots of observed AMSR-E LST and reconstructed AMSR-E LST against MYD11A1 during daytime and nighttime are presented in Fig. 7. Compared with the MYD11A1, the daytime (nighttime) average RMSE, $R^2$ and Bias were 3.44 K (2.70 K), 0.94 (0.92) and 0.21 K (0.10 K) for the observed AMSR-E LST, and the corresponding values were 4.39 K (3.24 K), 0.88 (0.90) and 0.29 K (0.26 K) for the reconstructed AMSR-E LST. The difference between the two RMSEs for the daytime and nighttime were all less than 1 K, and the $R^2$ was approximately 0.9, which again proved the effectiveness of using the MTFC-CNN method in real experiments.

### 4.2. Simulation experiments of the MSTFC-GAN

#### 4.2.1. Visualization results of simulation experiments

This section presents results of the simulation experiments using the aforementioned eight sample datasets, which were conducted to evaluate the reliability of the MSTFC-GAN model. The sample images in each dataset were simulated as cloud-covered images with the mask datasets introduced in Section 3.2.2. To effectively set missing rates in the simulation experiments, the percentages of the different missing rate ranges in all MODIS LST over China’s mainland in 2010 were investigated, as shown in Fig. 8. Almost all of the MODIS LSTs had missing rates ranging between 20% and 80%, and most of the missing rates ranged from 50% to 60%. The percentages of high missing rates were greater during the daytime than those during nighttime, while those of low missing rates during the nighttime were less than those during the daytime. Therefore, the missing rates of masks in the simulation

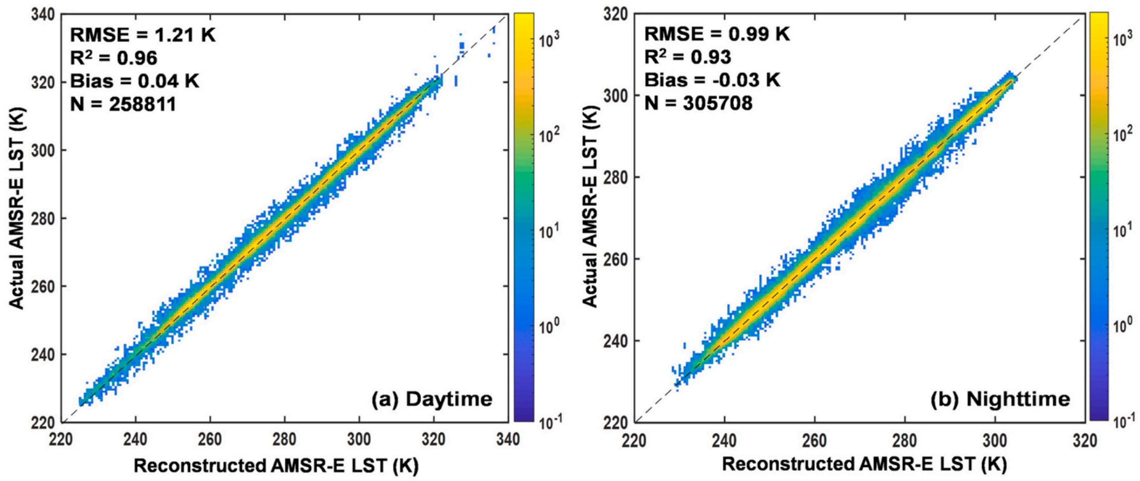
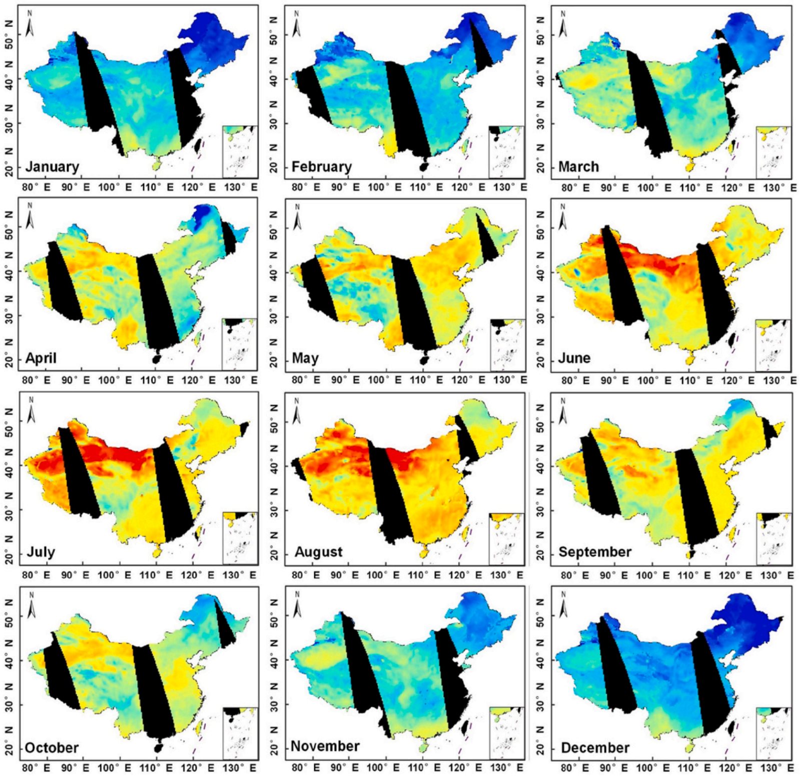
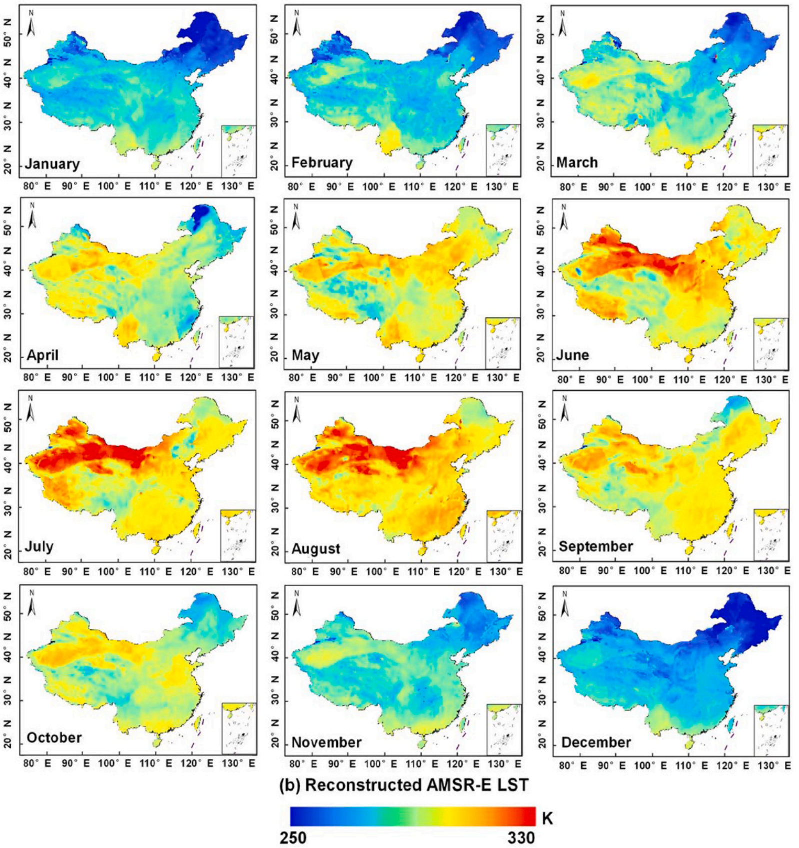


**Fig. 5.** Scatterplots of actual AMSR-E land surface temperature (LST) against AMSR-E LST reconstructed in simulation experiments with multi-temporal feature connected convolutional neural network (MTFC-CNN) during (a) daytime and (b) nighttime.



(a) Original AMSR-E LST



(b) Reconstructed AMSR-E LST

**Fig. 6.** Comparison between the spatiotemporal distributions of daytime original and reconstructed AMSR-E LST on the 15th day of every month in 2010: (a) Original AMSR-E LST, (b) reconstructed AMSR-E LST.

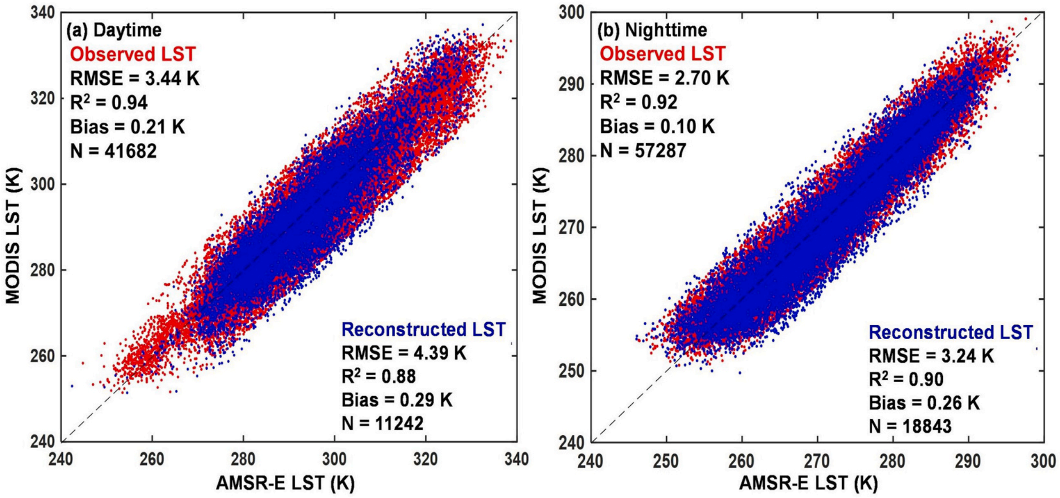


**Fig. 7.** Scatterplots of observed AMSR-E LST and reconstructed AMSR-E LST against MYD11A1 (MYD11A1): (a) Daytime, (b) Nighttime. Red samples are the observed results while blue samples are those reconstructed in this study. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

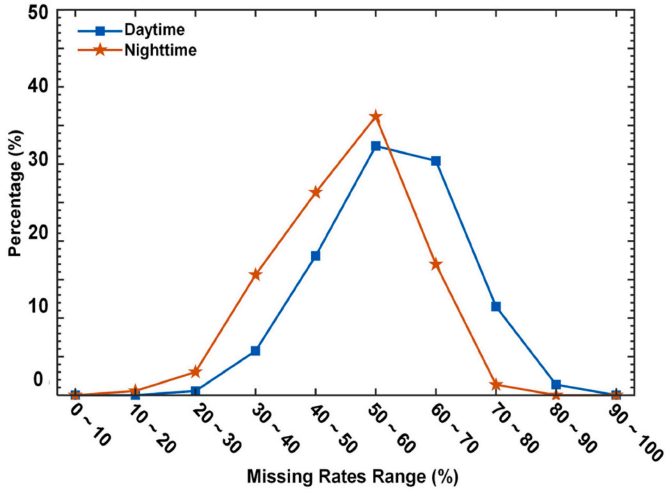


**Fig. 8.** Percentage of missing LST pixels during daytime and nighttime in 2010.

experiments were set as 25%, 30%, 35%, 40%, 55%, 65%, and 75%, respectively.

To investigate the efficacy of the MSTFC-GAN, the AMSR-E and the corresponding MODIS during daytime (Mar 17, June 18, September 7, and December 11, 2010) and nighttime (March 7, June 10, September 6, and December 1, 2010) were selected to conduct experiments from sample datasets. The simulated cloudy-sky MODIS LST images were generated using four missing rates (40%, 55%, 65%, and 75%) from the mask datasets. The proposed MSTFC-GAN was then used for the simulated cloudy-sky MODIS LST images. In addition, the proposed MSTFC-GAN without reference data was also used for simulated cloudy-sky MODIS LST, to provide a comparison with these results. Visualization results for AMSR-E LST, actual MODIS LST, simulated cloudy-sky MODIS LST using mask datasets, all-weather MODIS LST from MSFCR-GAN without reference data, and all-weather MODIS LST from MSTFC-GAN, are given in Fig. 9 for daytime and Fig. 10 for nighttime, respectively. Compared with spatial patterns of the all-weather MODIS LST from MSTFC-GAN without reference data (Fig. 9 (d) and Fig. 10 (d)), those of the all-weather MODIS LST from MSTFC-GAN (Fig. 9 (e) and Fig. 10 (e)) and the actual MODIS LST images were evidently more consistent. These results demonstrate that the spatial detail can be enhanced when using reference data and the spatiotemporal fusion concept, which suggests that the proposed MSTFC-GAN performs well in learning large-scale differences between the AMSR-E and MODIS LST.

### 4.2.2. *Evaluation using different missing rates*

The aforementioned eight sample datasets were used to evaluate MSTFC-GAN for each of the seven percentage missing rates (i.e., 25%, 30%, 35%, 40%, 55%, 65%, and 75%). The RMSE between the actual MODIS LST and the all-weather MODIS LST for the eight sample datasets were then calculated. The percentages of the average RMSE and its range distributions between all-weather MODIS LST and actual MODIS LST under different missing rates are shown in Fig. 11, where the four colors denote the RMSE ranges (from 0 to 4 K). Similarly, these values were calculated using only the data over the fake cloud mask gaps. With an increase in the missing rate, the average RMSE generally showed an upward trend, and a proportionate increase in the percentage of pixels with high RMSE values was observed. When the missing rates were lower than 40%, more than 60% of pixels had RMSE ranges of 0–1 K, and more than 95% of pixels had RMSE ranges of 0–2 K. RMSE ranges of 1–2 K were dominant for high missing rates (for example, for 55% and 65%), while ranges of 2–3 K were dominant with the highest missing rate of 75%. In summary, the results indicate that the MSTFC-GAN is able to reconstruct cloudy-sky MODIS LST for different missing rates when employing the corresponding AMSR-E LST data.

### 4.2.3. *Evaluation of different training sample sizes*

The training sample size (TSS) is an important factor that influences the outcomes of DL-based methods. To evaluate the influence of different training sample sizes on the efficacy of the MSTFC-GAN, sample training datasets during daytime and nighttime in spring were divided into different percentages of 20%, 40%, 60%, 80%, and 100%, respectively. The daytime and nighttime average RMSEs for the different percentages of sample training datasets under seven missing rates are shown in Fig. 12. The results show that when a lower percentage of training datasets was used, the RMSEs obtained were generally higher under the seven missing rates for both daytime and nighttime. However, the RMSEs did not always increase when the missing rate improved during low missing rate ranges (such as those less than 40%), particularly for daytime LST with a high dynamic variation. The missing rate distribution features (i.e., concentrated or scattered) also strongly influenced the RMSEs, which is in agreement with our previous study (Liu et al., 2017; Wu et al., 2019). In addition, the average RMSE for each missing rate during daytime was higher than that during nighttime.

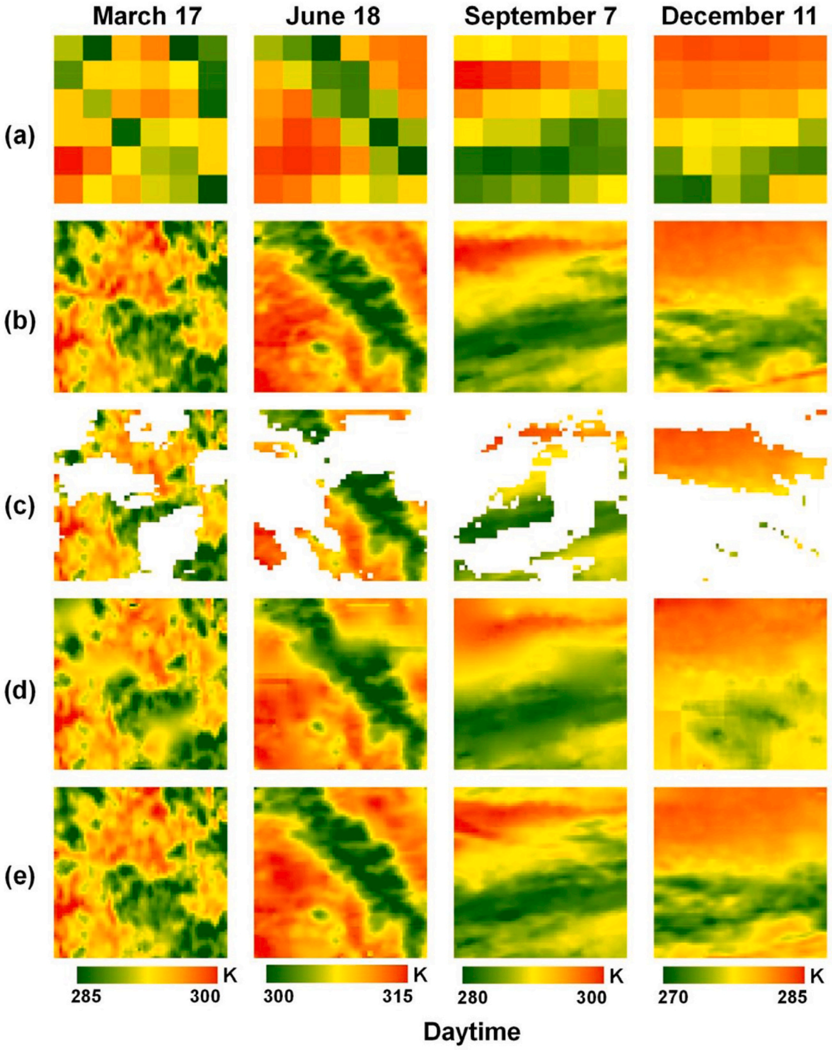


**Fig. 9.** Results of simulated experiment using LST data during daytime: (a) AMSR-E LST; (b) actual MODIS LST; (c) simulated cloudy-sky MODIS LST using mask datasets; (d) all-weather MODIS LST from multi-scale multi-temporal feature connected generative adversarial network model (MSTFC-GAN) without reference images; and (e) all-weather MODIS LST from MSTFC-GAN.

There were only slight differences between the RMSEs of different training dataset percentages under low missing rates, such as 25% and 30%. When the missing rate was 35%, 20% of the training datasets resulted in a higher RMSE than with the other training dataset percentages. With an increase in the missing rate, particularly for the missing rate of 65%, the daytime RMSEs of training dataset percentages lower than 60%, 80%, and 100% were approximately 3.5, 2.0 K, and 1.5 K, respectively. Thus, the training sample size is an important factor for MSTFC-GAN, and sufficient training samples were required when the LST data contained high missing rates.

### 4.3. Real experiments using the MSTFC-GAN

#### 4.3.1. Visualization results of real experiments

To evaluate the visual effects of the MSTFC-GAN when conducting real experiments, LST data were investigated both before and after using the MSTFC-GAN. The spatial distribution pattern of the original actual MODIS LST during the daytime (a) and nighttime (b) from Jan. 15 to Dec. 15, 2010 is shown in Fig. 13. Because of the presence of clouds, many regions (particularly in Southern China) of missing data are evident during both the daytime and nighttime. Although the original MODIS LST values during the daytime are consistently higher than those during nighttime, reflecting the spatial variations in LST during both daytime and nighttime is difficult.

All-weather LST can be generated using the proposed MSTFC-GAN and employing spatially complete AMSR-E LST (as described in Section 4.1) and original MODIS LST (see Fig. 13). The spatial distribution pattern of gapless all-weather LSTs during (a) daytime and (b) nighttime from Jan. 15 to Dec. 15, 2010, are shown in Fig. 14. Compared with the original MODIS LST, the all-weather LST more proficiently indicates the spatial variations of LST during both daytime and nighttime. Relatively evenly distributed LST occurs in winter, especially during nighttime, whereas LST varies greatly in space in summer, especially during the daytime. Relatively high LST value can be observed in northern China (particularly northwest region) during the daytime in summer, where arid and semi-arid regions are widespread. Furthermore, a greater number of clear-sky days can be found in northern China in summer than those in southern China. Conversely, the all-weather LSTs in southern China during nighttime are widely higher than those in northern China, owing to the greater amount of cloud covering in southern China, and the increased downward longwave radiation. The LSTs of Hainan Island show a smaller diurnal LST difference, mainly because the specific heat capacity of seawater is smaller than that of sand. When compared to the other regions at the same latitude, lower LSTs during both daytime and nighttime are found in the TP region, owing to its high altitude. These finds are generally in keeping with

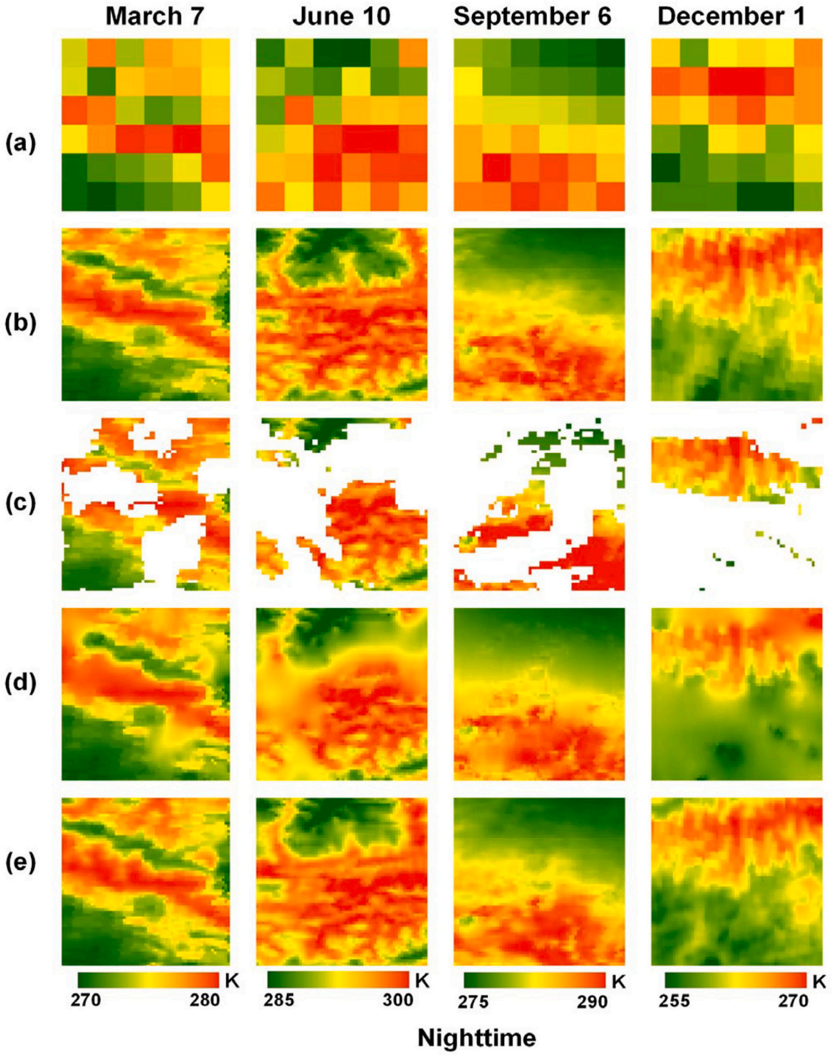


**Fig. 10.** Results of simulated experiment using LST data during nighttime. (a) AMSR-E LST; (b) actual MODIS LST; (c) simulated cloudy-sky MODIS LST using mask datasets; (d) all-weather MODIS LST from multi-scale multi-temporal feature connected generative adversarial network model (MSTFC-GAN) without reference image; and (e) all-weather MODIS LST from MSTFC-GAN.

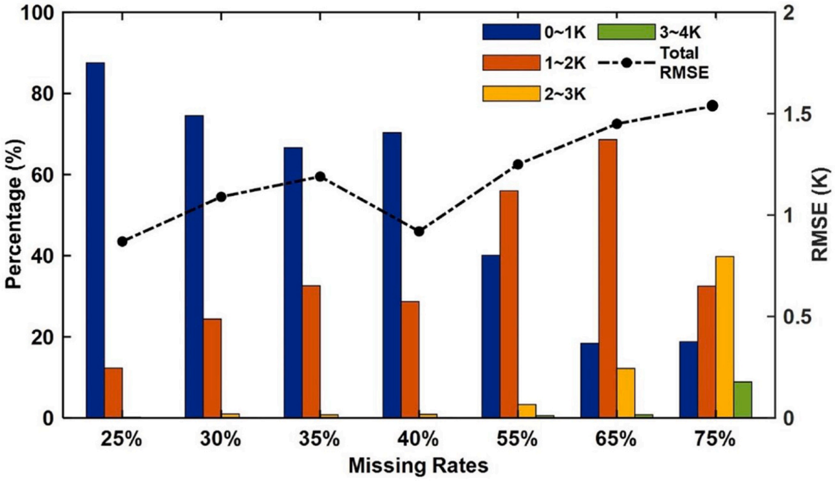


**Fig. 11.** Percentages of average RMSE and its range distributions between all-weather MODIS LST and actual MODIS LST under different missing rates.

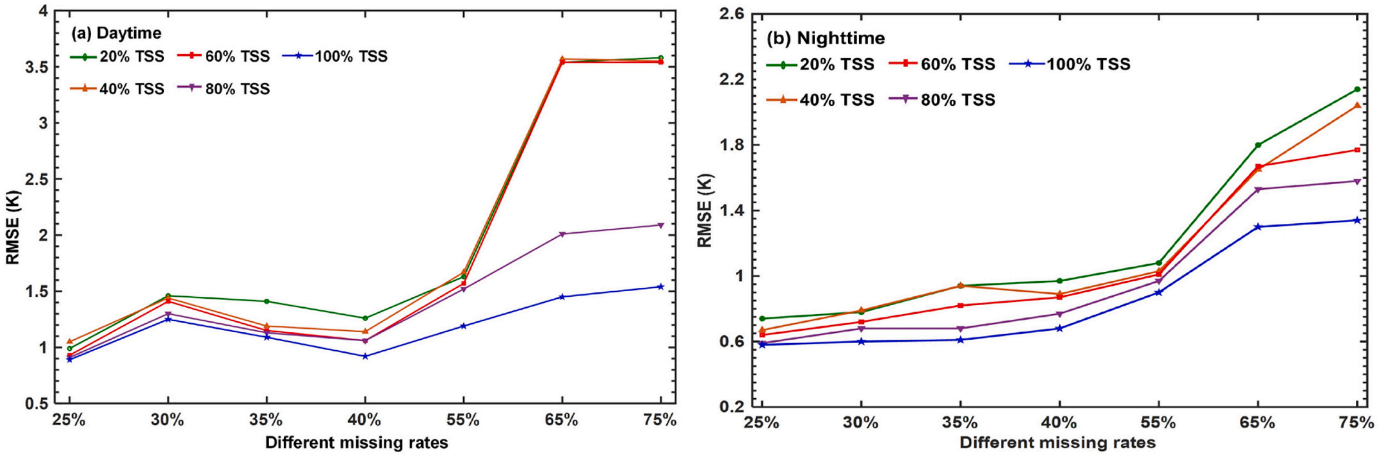


**Fig. 12.** Average RMSE for different sample training dataset percentages under seven missing rates: (a) daytime, (b) nighttime.

those of previous studies of Duan et al. (2017) and Xu and Cheng (2021). Furthermore, compared with the reconstructed 25 km spatial resolution AMSR-E LSTs (Fig. 6), LST variation is enhanced by the all-weather LSTs on a 1-km scale, which makes this method more suitable for meeting the requirements of LST studies on a regional scale or a finer scale.

### 4.3.2. Validation against in-situ measured LST

LST of the six sites (AR, HZZ, and YK in the HRB region, and BJ, SETORS, and NAMORS in the TP region) were used to assess the gapless all-weather LST. Scatterplots of the generated all-weather LST against in-situ LST from the six sites under clear and cloudy skies are shown in Fig. 15. In general, remarkable fluctuation in the average bias at all sites under both clear and cloudy skies were observed, while the average RMSE was relatively stable at all sites under both clear and cloudy skies. Specifically, the average bias of the clear LST varied from −0.88 K at the NAMORS site to 0.65 K at the AR site, and the average RMSEs of the clear LST were about 2 K at six sites. The accuracy of clear LST was much better than that of cloudy LST. The average bias of cloudy LST varied from −1.64 K at the HZZ to 1.51 K at the NAMORS. Under cloudy skies, the lowest RMSE of 3.41 K was measured at the YK, whereas the highest RMSE of 3.87 K was measured at the NAMORS.

To further evaluate the gapless all-weather LST, the average RMSE, bias, and $R^2$ of the all-weather LST against the in-situ LST were calculated during daytime and nighttime in different seasons, and the results are listed in Table 2. The performance of MSTFC-GAN was generally affected by seasons, and the influence during the nighttime was slightly greater than that during the daytime. During both daytime and nighttime, the smallest RMSEs of 3.50 K and 2.95 K were obtained in the winter, whereas the largest RMSEs of 3.96 K and 3.83 K were obtained in summer. The smallest $R^2$ of 0.74 was obtained during daytime in summer, whereas the largest $R^2$ of 0.84 was observed during nighttime in autumn. A negative bias occurred in winter during both daytime and nighttime, and also occurred in spring during nighttime, while a positive bias occurred at other times.

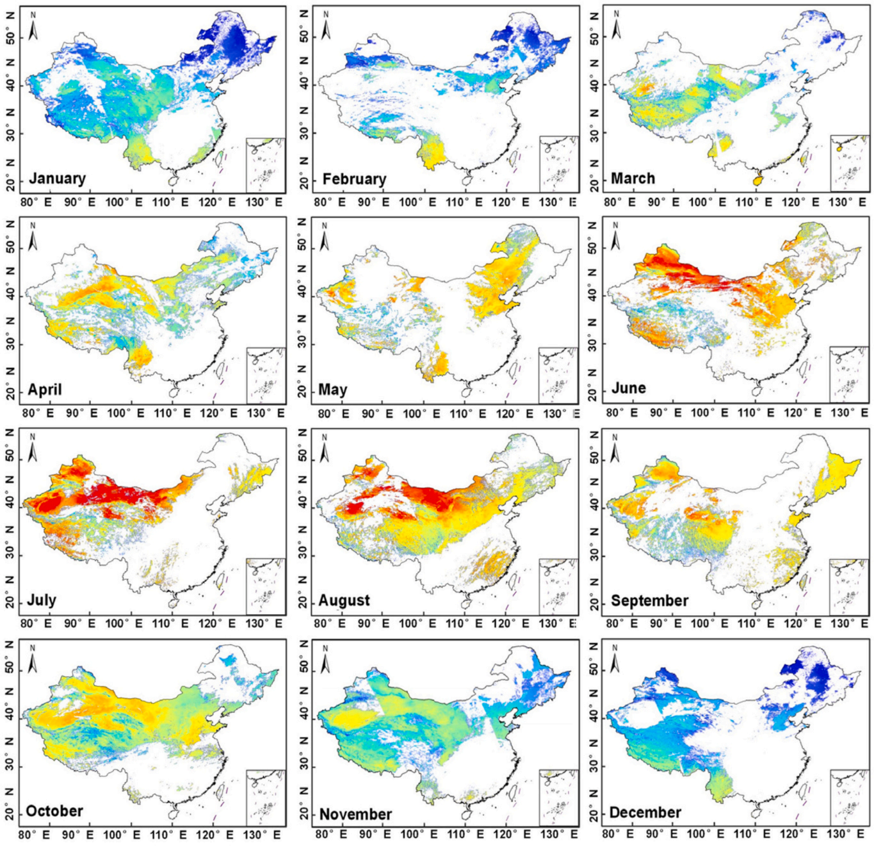
## 5. Discussion

### 5.1. Advantages and limitations of the TDLF

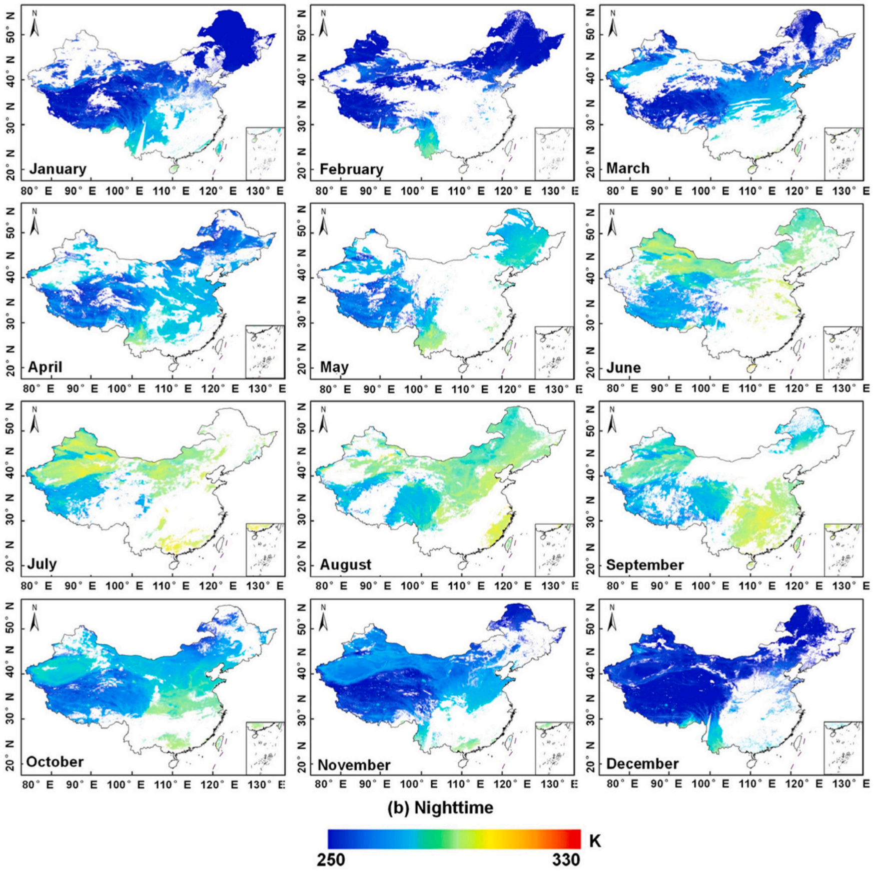
In this study, the TDLF was proposed for mapping gapless all-weather LST with TIR and passive microwave observations, and the MTFC-CNN filled large swath gaps in AMSR-E LST data. The MSTFC-GAN was subsequently designed to blend the spatially complete AMSR-E LST and original MODIS LST. The effectiveness of MTFC-CNN was subsequently validated against ground measurements for different missing rates and sample sizes, and was also compared with other methods.

As for the other methods, the enhanced performance of TDLF relates primarily to its ability to accommodate nonlinear relationships between the LST data with multi-temporal and multi-scale. Specifically, the three main advantages of developing the TDLF are as follows: first, although the IDW and DINEOF methods were adopted in previous studies (Duan et al., 2017; Xu and Cheng, 2021; Xu et al., 2021), their assumption of linear or orthogonal function relations between input and output LST data can result in uncertainty and noise pixels. Therefore, we designed the bidirectional reconstruction model based on the MTFC-CNN in consideration of the large and periodic orbit swath gap, the non-stationarity in AMSR-E LST data, and the ability to use complementary spatial information from adjacent data. Second, studies have shown that GANs have great potential for exploiting high-level information for multi-scale data (Ma et al., 2019; Zhang et al., 2020a), such as MODIS and AMSR-E. However, owing to the large scale inconsistence between MODIS and AMSR-E LST, direct learning of their relationship at a single observation time can result in the reconstructed MODIS LST losing a considerable amount of spatial detail, as shown in Fig. 9(d) and Fig. 10(d). A key component of the MSTFC-GAN is the incorporation of the spatiotemporal fusion concept. The basic premise behind spatiotemporal LST fusion is to generate LST with high resolution at a predicted time while simultaneously using low resolution LST and a pair of spatially complete high and low resolution LSTs observed at a reference time (Wu et al., 2021). Based on this idea, the reconstructed LST value and its spatial detail can be learned from the reference pairs of MODIS and AMSR-E LST, as shown in Fig. 9(e) and Fig. 10(e). The third advantage of the TDLF is that it requires no other additional data, whereas additional data (such as NDVI and DEM) are essential in the methods presented in previous studies. However, it should be noted that these additional data may also be very useful for the TDLF. How to bring additional data into TDLF is interesting, and should be investigated further in the future.

However, the TDLF has several potential limitations, which are as follows: (1) as LST varied strongly both spatially and temporally, a total of eight datasets obtained in 2010 were trained for the MTFC-CNN and MSTFC-GAN, respectively. In this respect, DL-based methods generally have higher time cost than traditional methods. (2) To enhance the spatial detail, the MSTFC-GAN required a spatially complete MODIS LST observed at the reference time. Experimental results showed no significant effect on the time interval between the predicted and referenced MODIS LST during a season. However, the performance of MSTFC-GAN might degrade when the spatially complete MODIS LST was unavailable during a given period of time, which might limit the applicability of the model. It should be noted that the difference of view angles between the predicted and referenced MODIS LST may further increase the

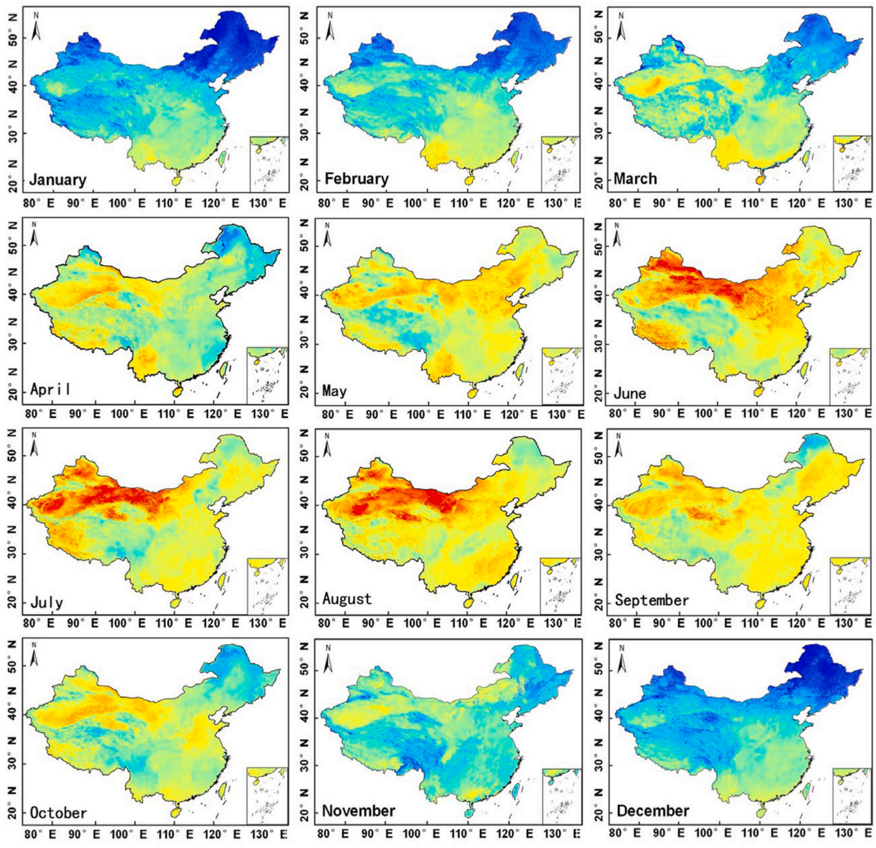


(a) Daytime

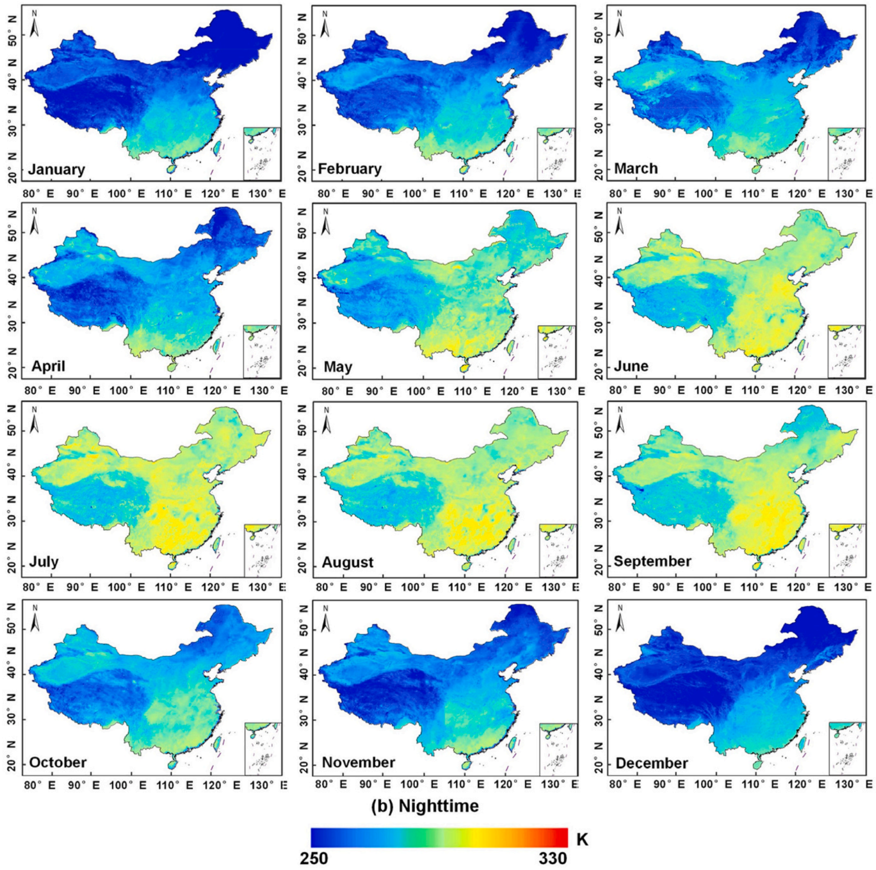


(b) Nighttime

**Fig. 13.** Spatial distribution pattern of the original MODIS LST on the 15th day of every month in 2010: (a) daytime and (b) nighttime.



(a) Daytime



(b) Nighttime

**Fig. 14.** Spatial distribution pattern of the gapless all-weather LST images on the 15th day of every month in 2010: (a) daytime and (b) nighttime.

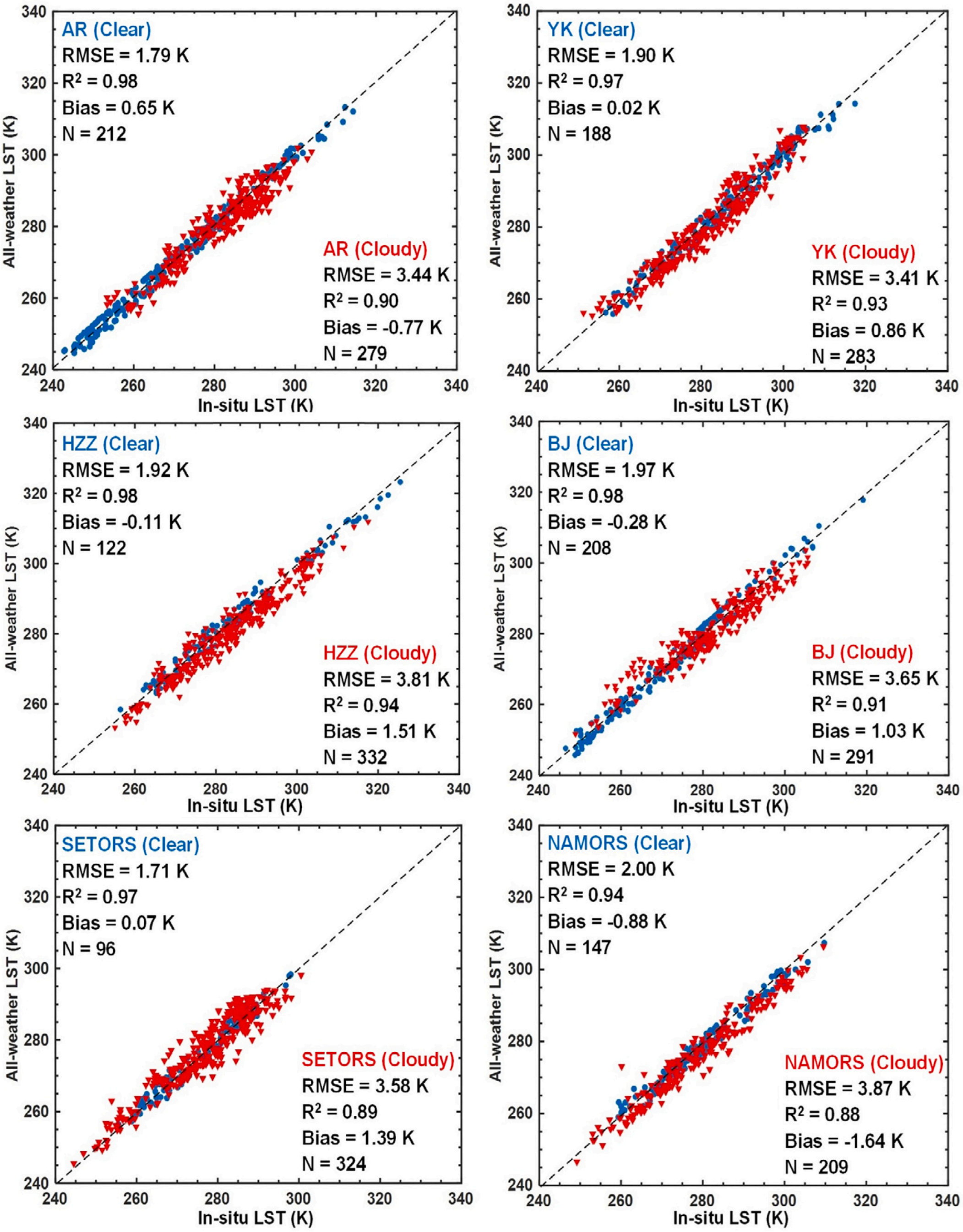


**Fig. 15.** Scatterplots of all-weather LST against six in-situ LST at Arou station (AR), Huazhaizi station (HZZ), Yingke station (YK), Naqu-BJ station (BJ), SETORS, and NAMORS sites under clear and cloudy skies.

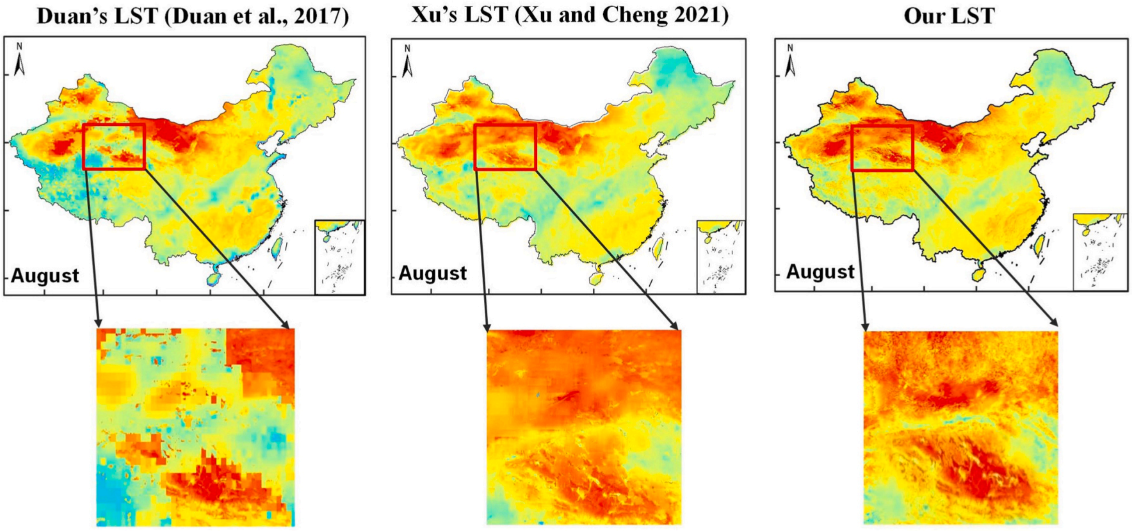
**Table 2**
Comparison between all-weather LST and in-situ LST in different seasons during daytime and nighttime, respectively.

| Time | Season | RMSE (K) | Bias (K) | $R^2$ |
|---|---|---|---|---|
| Daytime | spring | 3.74 | 1.08 | 0.80 |
| | summer | 3.96 | 1.20 | 0.74 |
| | autumn | 3.94 | 1.54 | 0.81 |
| | winter | 3.50 | −1.55 | 0.72 |
| Nighttime | spring | 3.40 | −0.79 | 0.83 |
| | summer | 3.83 | 0.98 | 0.80 |
| | autumn | 3.15 | 0.82 | 0.84 |
| | winter | 2.95 | −0.08 | 0.78 |

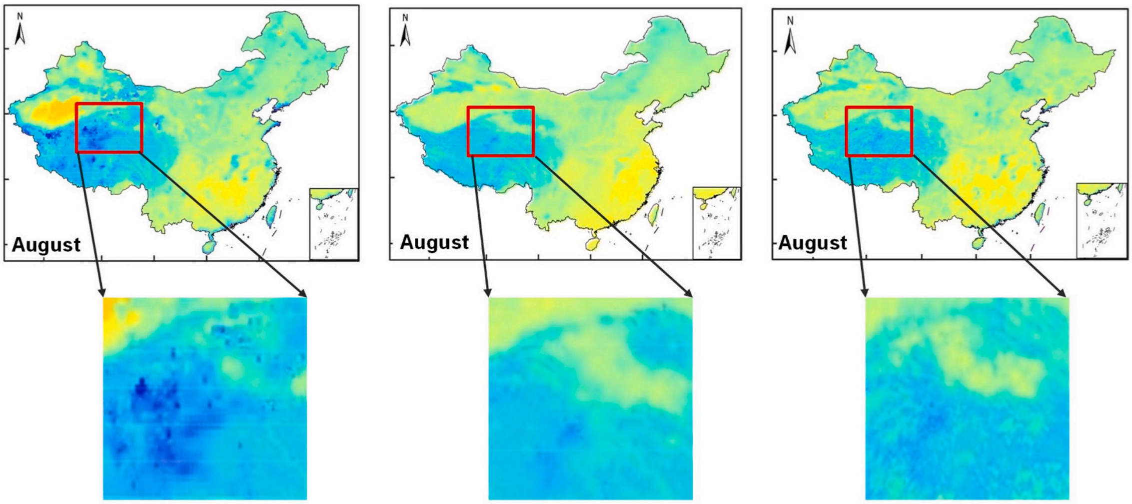
uncertainty of the LST estimates. The thermal radiation directionality is a complicated process (Cao et al., 2019), and requires further comprehensive study. (3) Considering the transit time of PWM, the proposed method was designed only for MYD11A1, but not for MOD11A1. Once temporal normalization methods are sufficiently mature, other LST data, such as reanalyzed LST products, may be promising alternatives for generating all-weather LST at a required time (Long et al., 2020; Zhang et al., 2021).

### 5.2. *Comparisons with previous all-weather LST*

Although a number of studies have produced all-weather LST, we compare the TDLF with the method by Duan et al. (2017) and the fusion


Duan's LST (Duan et al., 2017)
Xu's LST (Xu and Cheng 2021)
Our LST

(a) Daytime



(b) Nighttime

**Fig. 16.** Detailed comparison between Duan's LST (Duan et al., 2017), Xu's LST (Xu and Cheng, 2021), and the gapless all-weather LST generated from the TDLF on August 15th during (a) daytime and (b) nighttime.

strategy proposed by Xu and Cheng (2021). One of the main reasons for using these methods as comparisons was that MODIS and AMSR-E LST were adopted as the input data in both the TDLF and in the two other studies. It was also convenient to make such a comparison because the two methods were proposed to produce spatially complete all-weather LST products over the China's mainland from 2002 to 2011 (referred to as Duan's LST) and 2002 to 2020 (referred to as Xu's LST), respectively. The products were released free of charge by the National Center for Earth System Science Data (http: //www. geodata. cn) and National Tibetan Plateau Data Center (http://data.tpdc.ac.cn/) (Cheng et al., 2021), respectively. However, it should be noted that the AMSR-E LST products used in 2010 were provided by Zhang and Cheng (2020), based on their published work, and we thus also generated gapless all-weather LST in 2010 and compared it with the downloaded Duan's LST and Xu's LST products in 2010. A comparison between the gapless all-weather LST generated from the TDLF and the downloaded LST products (Duan's LST and Xu's LST) on August 15th is shown in Fig. 16 during daytime (a) and nighttime (b). On a national scale, comparing the quality of the three all-weather LST images over the Chinese mainland was difficult; therefore, certain detailed regions were selected to conduct a visual assessment, as shown by the red rectangles. It is evident that the Duan's LST appear more 'hazy' than the two other LST images, and our LST enhanced the spatial details when compared with the Xu's LST.

To quantitatively compare the accuracy of the TDLF with the method of Duan et al. (2017), and the fusion strategy proposed by Xu and Cheng (2021), the in situ measurements of AR, HZZ, YK, BJ, SETORS, and NAMORS were used to validate the gapless all-weather LST generated from the two methods. Notably, the AR, HZZ, and YK sites were also utilized for the method of Duan et al. (2017) and the strategy of Xu and Cheng (2021). Fig. 17 gives the validation results of the six sites during daytime and nighttime under cloudy conditions. Although the bias of our all-weather LST was basically consistent with Duan's and Xu's LST for all six sites, the RMSEs of our all-weather LST were lower than those of both Duan's and Xu's LST for all six sites. Therefore, the proposed TDLF is generally superior to the method of Duan et al. (2017) and the strategy of Xu and Cheng (2021) in terms of the visual effect of the all-weather LST images and the results of the validation accuracy.

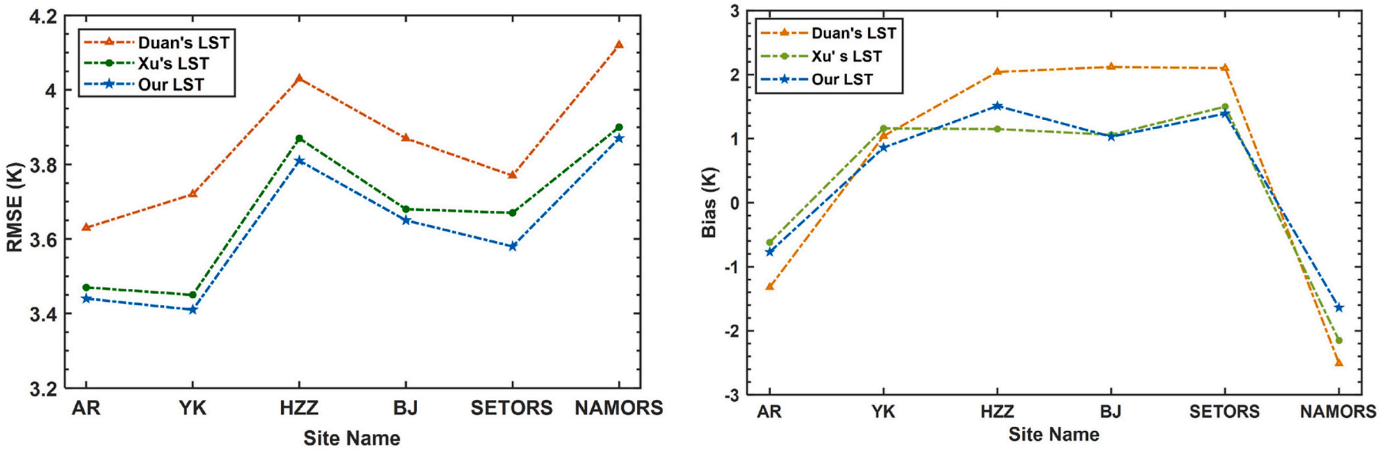


**Fig. 17.** Comparison of RMSE and bias between the all-weather LST from TDLF and LST obtained using the method of Duan et al. (2017) and the strategy of Xu and Cheng (2021) with in-situ LST under cloudy-skies at the AR, HZZ, YK, BJ, SETORS and NAMORS sites.

## 6. Conclusions

Blending data from TIR and PMW measurements is a common solution for generating all-weather LSTs. This paper propose the TDLF for mapping gapless all-weather LST over the China's landmass using MODIS LST and AMSR-E LST. A bidirectional reconstruction model based on MTFC-CNN was designed to obtain spatially complete AMSR-E LST. The MSFCR-GAN was subsequently designed using a spatiotemporal fusion concept to blend the spatially complete AMSR-E LST and cloudy-sky MODIS LST and ultimately generate gapless all-weather LST data.

The effectiveness of the MTFC-CNN was first evaluated by conducting simulations and real experiments. The average RMSEs between the reconstructed AMSR-E LST and the actual AMSR-E LST were 1.21 K for daytime and 0.99 K for nighttime, respectively. The performance of the MSTFC-GAN was also tested and evaluated using different missing rates, training sample sizes, and seasons. Experiments showed that the MSFCR-GAN provided satisfactory results, for example with an RMSE <2 K at a missing rate of 55%. The model had no significant effect for different percentages of training datasets when missing rates are less than 55%. When validated against six in-situ LSTs of HRB and TP regions, the generated gapless all-weather LSTs denoted an accuracy (average RMSE) of 1.71 K to 2.0 K in clear conditions and 3.41 K to 3.87 K in cloudy conditions. Uncertainties of the generated gapless all-weather LSTs may come from the retrieval of AMSR-E and MODIS LST, the implementation of MTFC-CNN, inconsistencies in observation time between AMSR-E and MODIS data over the high latitude region, and the implementation of MSTFC-GAN. Although the accuracy was affected by the uncertainties and varied with seasons, the method still demonstrated a reliable performance under all-weather conditions. When compared with the existing published PMW-based all-weather LSTs, the generated gapless all-weather LSTs were more satisfying at the aspect of image quality (e.g., spatial detail) and validation accuracy.

Overall, the TDLF fully utilizes the powerful nonlinear representation ability of CNNs and GANs, and it can learn nonstationary and high dynamic variations in multiple spatiotemporal scale LSTs from a large number of training samples. Although the TDLF was tested with MODIS and AMSR-E LST in 2010 and validated on the China's mainland, the capability of producing long-term, gapless, all-weather LST records on a global scale with other TIR LSTs (e.g., FY-3) and PWM LST (e.g., AMSR2 and MWRI) has been advanced. Such a capability will be beneficial for generating further spatially complete soil moisture and evapotranspiration datasets, which can be used in global climate change research.

## CRediT authorship contribution statement

**Penghai Wu:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization, Project administration, Funding acquisition. **Yang Su:** Methodology, Software, Formal analysis, Visualization, Writing – original draft. **Si-bo Duan:** Conceptualization, Methodology, Resources, Writing – review & editing, Funding acquisition. **Xinghua Li:** Conceptualization, Methodology, Formal analysis, Writing – review & editing. **Hui Yang:** Formal analysis, Investigation, Writing – review & editing, Visualization. **Chao Zeng:** Software, Formal analysis, Investigation, Writing – review & editing. **Xiaoshuang Ma:** Software, Formal analysis, Investigation, Writing – review & editing. **Yanlan Wu:** Resources, Methodology, Writing – review & editing. **Huanfeng Shen:** Conceptualization, Methodology, Resources, Writing – review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

We thank the three anonymous reviewers for their valuable comments and constructive suggestions that substantially improved the paper. Special thanks to Prof. Jie Cheng at the Beijing Normal University for providing the AMSR-E LST data. We also would like to thank the data support from NASA and National Tibet Plateau Data Center. The research was funded by the National Natural Science Foundation of China under Grants 41501376 and 41871275, Engineering Center for Geographic Information of Anhui Province, and Anhui Province Key Laboratory of Wetland Ecosystem Protection and Restoration.

## References

Arjovsky, M., Chintala, S., Bottou, L., 2017. Wasserstein generative adversarial networks. In: Proceedings of the 34nd International Conference on Machine Learning. PMLR, pp. 214–223.

Cao, B., Liu, Q., Du, Y., Roujean, J.L., Gastellu-Etchegorry, J.P., Trigo, I.F., Zhan, W., Yu, Y., Cheng, J., Jacob, F., Lagourde, J.P., Bian, Z., Li, H., Hu, T., Xiao, Q., 2019. A review of earth surface thermal radiation directionality observing and modeling: historical development, current status and perspectives. Remote Sens. Environ. 232, 111304.

Cheng, J., Dong, S., Shi, J., 2021. In: C. National Tibetan Plateau Data (Ed.), 1km Seamless Land Surface Temperature Dataset of China (2002–2020). National Tibetan Plateau Data Center.

Crosson, W., Al-Hamdan, M., Hemmings, S., Wade, G., 2012. A daily merged MODIS Aqua–Terra land surface temperature data set for the conterminous United States. Remote Sens. Environ. 119, 315–324.

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

Fu, P., Xie, Y., Weng, Q., Myint, S., Meacham-Hensold, K., Bernacchi, C., 2019. A physical model-based method for retrieving urban land surface temperatures under cloudy conditions. Remote Sens. Environ. 230, 111191.

Göttsche, F.-M., Hulley, G.C., 2012. Validation of six satellite-retrieved land surface emissivity products over two land cover types in a hyper-arid region. Remote Sens. Environ. 124, 149–158.

Guillevic, P., Göttsche, F., Nickeson, J., Hulley, G., Ghent, D., Yu, Y., Trigo, I., Hook, S., Sobrino, J., Remedios, J., 2017. Land surface temperature product validation best practice protocol. Version 1.0. In: Best Practice for Satellite-Derived Land Product Validation, p. 60.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A., 2017. Improved training of wasserstein gans. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 5769–5779.

Hansen, J., Ruedy, R., Sato, M., Lo, K., 2010. Global surface temperature change. Rev. Geophys. 48, RG4004.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1026–1034.

Holmes, T., Hain, C., Anderson, M., Crow, W., 2016. Cloud tolerance of remote-sensing technologies to measure land surface temperature. Hydrol. Earth Syst. Sci. 20, 3263–3275.

Huang, C., Duan, S.-B., Jiang, X.-G., Han, X.-J., Leng, P., Gao, M.-F., Li, Z.-L., 2019. A physically based algorithm for retrieving land surface temperature under cloudy conditions from AMSR2 passive microwave measurements. Int. J. Remote Sens. 40, 1828–1843.

Jia, A., Ma, H., Liang, S., Wang, D., 2021. Cloudy-sky land surface temperature from VIIRS and MODIS satellite data using a surface energy balance-based method. Remote Sens. Environ. 263, 112566.

Jin, M., 2000. Interpolation of surface radiative temperature measured from polar orbiting satellites to a diurnal cycle: 2. Cloudy-pixel treatment. J. Geophys. Res.-Atmos. 105, 4061–4076.

Kang, J., Tan, J., Jin, R., Li, X., Zhang, Y., 2018. Reconstruction of MODIS land surface temperature products based on multi-temporal information. Remote Sens. 10, 1112.

Kilibarda, M., Hengl, T., Heuvelink, G.B., Gräler, B., Pebesma, E., Perčec Tadić, M., Bajat, B., 2014. Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. J. Geophys. Res.-Atmos. 119, 2294–2313.

Kou, X., Jiang, L., Bo, Y., Yan, S., Chai, L., 2016. Estimation of land surface temperature through blending MODIS and AMSR-E data with the Bayesian maximum entropy method. Remote Sens. 8, 105.

Li, X., Li, X., Li, Z., Ma, M., Wang, J., Xiao, Q., Liu, Q., Che, T., Chen, E., Yan, G., 2009. Watershed allied telemetry experimental research. J. Geophys. Res.-Atmos. 114.

Li, Z., Tang, B., Wu, H., Ren, H., Yan, G., Wan, Z., Trigo, I., Sobrino, J., 2013. Satellite-derived land surface temperature: current status and perspectives. Remote Sens. Environ. 131, 14–37.

Li, X., Shen, H., Zhang, L., Zhang, H., Yuan, Q., Yang, G., 2014. Recovering quantitative remote sensing products contaminated by thick clouds and shadows using multitemporal dictionary learning. IEEE Trans. Geosci. Remote Sens. 52, 7086–7098.

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

Li, B., Liang, S., Liu, X., Ma, H., Chen, Y., Liang, T., He, T., 2021. Estimation of all-sky 1 km land surface temperature over the conterminous United States. Remote Sens. Environ. 266, 112707.

Liu, H., Weng, Q., 2012. Enhancing temporal resolution of satellite imagery for public health studies: a case study of West Nile virus outbreak in Los Angeles in 2007. Remote Sens. Environ. 117, 57–71.

Liu, Z., Wu, P., Duan, S., Zhan, W., Ma, X., Wu, Y., 2017. Spatiotemporal reconstruction of land surface temperature derived from fengyun geostationary satellite data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10, 4531–4543.

Liu, Z., Zhan, W., Lai, J., Hong, F., Quan, J., Bechtel, B., Huang, F., Zou, Z., 2019. Balancing prediction accuracy and generalization ability: a hybrid framework for modelling the annual dynamics of satellite-derived land surface temperatures. ISPRS J. Photogramm. Remote Sens. 151, 189–206.

Long, D., Yan, L., Bai, L.L., Zhang, C.J., Li, X.Y., Lei, H.M., Yang, H.B., Tian, F.Q., Zeng, C., Meng, X.Y., Shi, C.X., 2020. Generation of MODIS-like land surface temperatures under all-weather conditions based on a data fusion approach. Remote Sens. Environ. 246.

Ma, J., Yu, W., Liang, P., Li, C., Jiang, J., 2019. FusionGAN: a generative adversarial network for infrared and visible image fusion. Information Fusion 48, 11–26.

Ma, J., Zhou, J., Göttsche, F.-M., Liang, S., Wang, S., Li, M., 2020a. A global long-term (1981–2000) land surface temperature product for NOAA AVHRR. Earth Syst. Sci. Data 12, 3247–3268.

Ma, Y., Hu, Z., Xie, Z., Ma, W., Wang, B., Chen, X., Li, M., Zhong, L., Sun, F., Gu, L., 2020b. A long-term (2005–2016) dataset of hourly integrated land–atmosphere interaction observations on the Tibetan Plateau. Earth Syst. Sci. Data 12, 2937–2957.

Martins, J., Trigo, I.F., Ghilain, N., Jimenez, C., Göttsche, F.-M., Ermida, S.L., Olesen, F.-S., Gellens-Meulenberghs, F., Arboleda, A., 2019. An all-weather land surface temperature product based on MSG/SEVIRI observations. Remote Sens. 11, 3044.

Mo, Y., Xu, Y., Chen, H., Zhu, S., 2021. A review of reconstructing remotely sensed land surface temperature under cloudy conditions. Remote Sens. 13, 2838.

Prata, A.J., Caselles, V., Coll, C., Sobrino, J.A., Ottlé, C., 1995. Thermal remote sensing of land surface temperature from satellites: current status and future prospects. Remote Sens. Rev. 12, 175–224.

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

Shwetha, H., Kumar, D., 2016. Prediction of high spatio-temporal resolution land surface temperature under cloudy conditions using microwave vegetation index and ANN. ISPRS J. Photogramm. Remote Sens. 117, 40–55.

Su, Y., Wu, P., Cheng, J., Yin, Z., Ma, X., Yang, H., 2022. Research on deep learning methods for AMSR-E land surface temperature data reconstruction. Nat. Remote Sens. Bull. https://doi.org/10.11834/jrs.20210426.

Sun, L., Chen, Z., Gao, F., Anderson, M., Song, L., Wang, L., Hu, B., Yang, Y., 2017. Reconstructing daily clear-sky land surface temperature for cloudy regions from MODIS data. Comput. Geosci. 105, 10–20.

Sun, D., Li, Y., Zhan, X., Houser, P., Yang, C., Chiu, L., Yang, R., 2019. Land surface temperature derivation under all sky conditions through integrating AMSR-E/AMSR-2 and MODIS/GOES observations. Remote Sens. 11, 1704.

Tierney, J., Russell, J., Huang, Y., Damste, J., Hopmans, E., Cohen, A., 2008. Northern hemisphere controls on tropical southeast African climate during the past 60,000 years. Science 322, 252–255.

Trigo, I., Monteiro, I., Olesen, F., Kabsch, E., 2008. An assessment of remotely sensed land surface temperature. J. Geophys. Res.-Atmos. 113.

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

Wan, Z., Dozier, J., 1996. A generalized split-window algorithm for retrieving land-surface temperature from space. IEEE Trans. Geosci. Remote Sens. 34, 892–905.

Wang, K., Wan, Z., Wang, P., Sparrow, M., Liu, J., Zhou, X., Haginoya, S., 2005. Estimation of surface long wave radiation and broadband emissivity using Moderate Resolution Imaging Spectroradiometer (MODIS) land surface temperature/emissivity products. J. Geophys. Res. 110, D11109.

Wang, T., Shi, J., Ma, Y., Husi, L., Edward, C., Ji, D., Zhao, T., Xiong, C., 2019. Recovering land surface temperature under cloudy skies considering the solar-cloud-satellite geometry: application to MODIS and Landsat-8 data. J. Geophys. Res.-Atmos. 124, 3401–3416.

Wang, H., Mao, K., Yuan, Z., Shi, J., Cao, M., Qin, Z., Duan, S., Tang, B., 2021. A method for land surface temperature retrieval based on model-data-knowledge-driven and deep learning. Remote Sens. Environ. 265, 112665.

Wu, P., Shen, H., Zhang, L., Göttsche, F., 2015. Integrated fusion of multi-scale polar-orbiting and geostationary satellite observations for the mapping of high spatial and temporal resolution land surface temperature. Remote Sens. Environ. 156, 169–181.

Wu, P., Yin, Z., Yang, H., Wu, Y., Ma, X., 2019. Reconstructing geostationary satellite land surface temperature imagery based on a multiscale feature connected convolutional neural network. Remote Sens. 11, 300.

Wu, P., Yin, Z., Zeng, C., Duan, S.-B., Gottsche, F.-M., Li, X., Ma, X., Yang, H., Shen, H., 2021. Spatially continuous and high-resolution land surface temperature product generation: a review of reconstruction and spatiotemporal fusion techniques. IEEE Geosci. Remote Sens. Mag. 9, 112–137.

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

Xu, S., Cheng, J., Zhang, Q., 2019. Reconstructing all-weather land surface temperature using the Bayesian maximum entropy method over the Tibetan Plateau and Heihe River Basin. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12, 3307–3316.

Xu, S., Cheng, J., Zhang, Q., 2021. A random forest-based data fusion method for obtaining all-weather land surface temperature with high spatial resolution. Remote Sens. 13, 2211.

Yang, G., Sun, W., Shen, H., Meng, X., Li, J., 2019. An integrated method for reconstructing daily MODIS land surface temperature data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12, 1026–1040.

Yin, Z., Wu, P., Foody, G.M., Wu, Y., Liu, Z., Du, Y., Ling, F., 2021. Spatiotemporal fusion of land surface temperature based on a convolutional neural network. IEEE Trans. Geosci. Remote Sens. 59, 15.

Yu, W., Tan, J., Ma, M., Li, X., She, X., Song, Z., 2019. An effective similar-pixel reconstruction of the high-frequency cloud-covered areas of Southwest China. Remote Sens. 11, 336.

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., Zhang, L., 2020. Deep learning in environmental remote sensing: achievements and challenges. Remote Sens. Environ. 241, 111716.

Zeng, C., Long, D., Shen, H., Wu, P., Cui, Y., Hong, Y., 2018. A two-step framework for reconstructing remotely sensed land surface temperatures contaminated by cloud. ISPRS J. Photogramm. Remote Sens. 141, 30–45.

Zhang, Q., Cheng, J., 2020. An empirical algorithm for retrieving land surface temperature from AMSR-E data considering the comprehensive effects of environmental variables. Earth Space Sci. 7 e2019EA001006.

Zhang, X., Zhou, J., Göttsche, F.-M., Zhan, W., Liu, S., Cao, R., 2019. A method based on temporal component decomposition for estimating 1-km all-weather land surface temperature by merging satellite thermal infrared and passive microwave observations. IEEE Trans. Geosci. Remote Sens. 57, 4670–4691.

Zhang, H., Song, Y., Han, C., Zhang, L., 2020a. Remote sensing image spatiotemporal fusion using a generative adversarial network. IEEE Trans. Geosci. Remote Sens. 59, 4273–4286.

Zhang, Y., Chen, Y., Li, J., Chen, X., 2020b. A simple method for converting 1-km resolution daily clear-sky LST into real LST. Remote Sens. 12, 1641.

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

Zhao, W., Duan, S.B., 2020. Reconstruction of daytime land surface temperatures under cloud-covered conditions using integrated MODIS/Terra land products and MSG geostationary satellite data. Remote Sens. Environ. 247, 111931.

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

Zhao, T., Shi, J., Lv, L., Xu, H., Chen, D., Cui, Q., Jackson, T., Yan, Guang, Jia, Li, Chen, Liang, Zhao, K., Zheng, X., Zhao, L., Zheng, C., Ji, D., Xiong, C., Wang, T., Li, R., Pan, J., Wen, J., Yu, C., Zheng, Y., Jiang, L., Chai, L., Lu, H., Yao, P., Ma, J., Lv, H., Wu, J., Zhao, W., Yang, N., Guo, P., Li, Y., Hu, L., Geng, D., Zhang, Z., 2020b. Soil moisture experiment in the Luan River supporting new satellite mission opportunities. Remote Sens. Environ. 240, 111680.