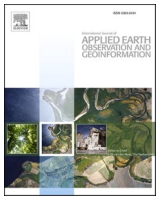
International Journal of Applied Earth Observations and Geoinformation 108 (2022) 102733

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

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



Combining ATC and 3D-CNN for reconstructing spatially and temporally continuous land surface temperature

Huyan Fu [a,b], Zhenfeng Shao [a,*], Peng Fu [c], Xiao Huang [d], Tao Cheng [a], Yewen Fan [a]

[a] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing and Collaborative Innovation Centre of Geospatial Technology, Wuhan University, Wuhan, China
[b] International Institute for Earth System Science, Nanjing University, China
[c] Institute for Sustainability, Energy, and Environment, University of Illinois at Urbana-Champaign, Urbana, IL, USA
[d] Department of Geosciences, University of Arkansas, Fayetteville, AR 72701, USA

ARTICLE INFO

Keywords:
High spatiotemporal
LSTs
Reconstruction
ATC
3D-CNN
Urban

ABSTRACT

More than half of the satellite-derived Land surface temperatures (LSTs) data are missing due to poor weather conditions (e.g., clouds, shadows, and other atmospheric conditions) and/or sensor failure, which has greatly limited the acquisition of spatially consistent and temporally regular LST data, leading to reduced utilization and accuracy of the data and hindering the understanding of the spatial-temporal patterns of surface thermal environments. Although several reconstruction methods have been developed, they are not effective in reconstructing daily LSTs for high spatial-temporal dynamics in urban regions. In this study, we developed a hybrid reconstructing method based on the annual temperature cycle (ATC) and three-dimensional convolutional neural network (3D-CNN) models. The proposed method has been validated on both Landsat 8 Thermal Infrared Sensor (TIRS) and Moderate Resolution Imaging Spectroradiometer (MODIS) Terra satellite LST datasets. The reconstructed results show that the proposed model incorporated with auxiliary data and 3D-CNN model significantly outperforms standard annual temperature cycle (ATCs) and enhanced annual temperature cycle (ATCe) models (Root Mean Square Error (RMSE): Landsat ATCs = 4.21 K, Landsat ATCe = 3.25 K, Landsat ATCe&SATs&CNN = 0.96 K, MODIS ATCs = 3.83 K, MODIS ATCe = 3.15 K, MODIS ATCs&SATs&CNN = 0.61 K). The proposed method is simpler, more efficient with higher robustness in terms of reconstructing high spatiotemporal LSTs in urban settings compared with existing reconstruction methods.

1. Introduction

Global climate warming has caused multiple impacts on the ecosystems and climate systems, e.g., the increase in both land and ocean temperatures and the number of extreme climate events (IPCC, 2018). The significant increase in global mean surface temperature, which peaked at 0.87 °C in 2006–2015, has strengthened the frequency, duration, and magnitude of these impacts (IPCC, 2018). Land surface temperature (LST) is a key indicator for understanding various important issues that include surface energy balance(Bastiaanssen et al., 1998; Su, 2002), global environmental change(Zeng et al., 2018), and urban climate (Maimaitiyiming et al., 2014; Sun et al., 2019; Tayyebi and Jenerette, 2018). It has been widely used in a variety of disciplines such as hydrology, meteorology, climate change, ecology, environmental monitoring(Anderson et al., 2012; Deng and Wu, 2013; Estoque and Murayama, 2017; Kustas and Anderson, 2009; Wang et al., 2009; Zhang and Li, 2018). Accurate and consistent temperature observations across the globe are crucial to understanding the land surface thermal anomalies, particularly over the complex terrains and urban regions.

Traditionally, LST has been recorded by radiometers at weather stations as in-situ point data. Given the influences from solar radiation, surface albedo, atmospheric conditions, soil thermal properties, and land cover, LST tends to have strong spatiotemporal heterogeneity (Fu et al., 2021). Compared with the traditional ground-based LST measurements, satellite remote sensing techniques provide notable advantages in terms of obtaining spatially continuous and temporally regular LSTs over extensive areas. Numerous LSTs products from medium to coarse spatial resolutions (Landsat 8 with 100 m resolution, MODIS with

* Corresponding author.
E-mail addresses: fuhuyan@nju.edu.cn (H. Fu), shaozhenfeng@whu.edu.cn (Z. Shao), pengfu@illinois.edu (P. Fu), xh010@uark.edu (X. Huang), taocheng@whu.edu.cn (T. Cheng), fyw@whu.edu.cn (Y. Fan).

https://doi.org/10.1016/j.jag.2022.102733
Received 26 October 2021; Received in revised form 17 February 2022; Accepted 19 February 2022
Available online 14 March 2022

1 km resolution, Meteosat Second Generation (MSG)—Spinning Enhanced Visible and Infrared Imager (SEVIRI) with 3 km resolution) have been developed from thermal infrared (TIR) data via various algorithms, e.g., Radioactive transfer equation (Sobrino et al., 2004), single-channel (Jiménez-Muñoz and Sobrino, 2003; Qin et al., 2001a) and split-window (Qin et al., 2001b; Wan, 2014). However, the availability of LSTs has been greatly limited by the existence of TIR data gaps resulting from the global prevalence of clouds, shadows, and unsatisfactory atmospheric conditions. The missing pixels are generally filled with invalid values. As a result, the incomplete spatial and temporal nature of these LSTs products pose great challenges to the spatiotemporal analysis in urban thermal environments. Therefore, the reconstruction of missing values in LSTs has been one of the hot issues for analyzing spatial and temporal LST information.

Several methods have been proposed to reconstruct LSTs data, which can be generally divided into four groups: the first group of methods usually rely on statistical relationships to estimate the missing data, including the leverage of the available LSTs and auxiliary data (e.g., vegetation index, latitude, longitude, elevation, precipitation and soil moisture) or the same/different satellites data (Crosson et al., 2012; Gallo et al., 2011; Hengl et al., 2012; Key, 2002; Key and Wong, 1999; Kustas et al., 2003). The uncertainty from the statistical-empirical LST model has an influence on the accuracy of reconstructed LST (Tan et al., 2021). The second group of methods reconstruct the missing data, taking advantage of the spatial distribution (Shuai et al., 2014; Yu et al., 2014), temporal variation (Van Nguyen et al., 2015; Xu and Shen, 2013), and spatiotemporal information (Pede and Mountrakis, 2018; Sun et al., 2017; Weiss et al., 2014) of LSTs data. Spatial information interpolation methods, such as Kriging, inverse distance weighting (IDW), aim to obtain the missing value based on the relationship between the missing data and the nearest clear-sky pixels (Neteler, 2010). Temporal approaches, such as harmonic analysis of time series and the Savitzky-Golay filter, aim to reconstruct the missing LST data by fitting the time series of clear-sky pixels (Xu and Shen, 2013; Zeng et al., 2015). However, the large areas of missing pixels limit the spatial method availability and then lead to unsatisfactory accuracy (Xiao et al., 2021). The accuracy of the temporal method depends on the time gaps between two clear-sky LSTs. Therefore, spatiotemporal interpolation methods are developed to reconstruct the missing LSTs by combining the advantages of spatial and temporal methods. The aforementioned methods can obtain the hypothetic clear-sky LSTs rather than the actual cloudy-sky LSTs. The third group of methods are proposed to estimate the LST, especially those pixels under cloudy conditions. Jin and Dickinson (2000) proposed the physical algorithm to derive the real LST based on surface energy balance (SEB). The missing data can be reconstructed with its spatially and temporally neighboring cloud-free pixels and the SEB difference. However, the related physical parameters, such as the surface incident radiation, air temperature, and wind velocity, are difficult to obtain. The satellite passive microwave (PMW) measurements renders the possibility to retrieve all-sky LSTs (Han et al., 2019; Huang et al., 2019). Regrettably, the temperature obtained from PMW remote sensing data is the subsurface temperature rather than skin temperature, and the coarse spatial resolution and orbit gaps lead to unsatisfactory accuracy (Li et al., 2021). The fourth group of methods are hybrid approaches (Li et al., 2018; Yang et al., 2019; Zeng et al., 2018), aiming to fill LST gaps by combining a series of methods so as to take advantage of their merits. In addition, machine learning (Arslan and Sekertekin, 2019; Wu et al., 2019), LST downscaling, and spatiotemporal image fusion (Gevaert and García-Haro, 2015; Weng and Fu, 2014) are becoming increasing popular in reconstructing LSTs.

However, the aforementioned methods have their advantages and limitations under different conditions. Furthermore, most existing algorithms are based on linear models, leading to great difficulty when dealing with complex scenarios and data with missing values that cover large geographical regions and long temporal spans. With the increasing demand for consistent spatiotemporal LSTs datasets, it is necessary to ensure the reconstruction of large-scale LSTs data with high efficiency and high accuracy. Data-driven machine learning approaches cannot fully replace physical modeling, but they can certainly enhance and complement them (Reichstein et al., 2019). A hybrid modeling approach that synergizes physical process and machine learning approach can lead to a better understanding of earth system science (Reichstein et al., 2019).

Semi-physical models such as annual temperature cycle (ATC) and diurnal temperature cycle (DTC) can cope with the problem of temporally irregularly LST observations caused by poor atmospheric conditions (Fu and Weng, 2015). For example, the ATC model developed by (Bechtel, 2012) expresses temporal LST dynamics through three parameters. The synthetic approach was developed by combining the ATCs originally proposed by Bechtel with auxiliary remote sensing and in-situ data to explain the short-term LST variations owing to synoptic conditions (Zou et al., 2018). However, existing ATC models have relatively lower prediction accuracy due to the limited number of parameters. In addition, they belong to temporal models, thus lacking spatial information (Liu et al., 2019).

In recent years, deep learning (DL) has become the mainstream in artificial intelligence and machine learning, providing promising insights on solving climate change science at various spatial and temporal scales (Reichstein et al., 2019; Singh et al., 2021). The commonly DL models, e.g., long short-term memory (LSTM), recurrent neural network (RNN), two-dimensional (2D) convolutional neural networks, are capable of making classifications and predictions from spatial or temporal perspectives, such as seasonal forecasting, extreme events, and land use land cover change(Kumar et al., 2021; Qiu et al., 2020; Singh et al., 2021; Yu et al., 2021). However, the majority of these methods are constructed based on two-dimensional features, failing to capture the spatial and temporal features in both dimensions (Qi et al., 2019; Xu et al., 2018). The three-dimensional (3D) kernel of 3D-CNN is able to slide across both spatial and temporal dimensions, thus simultaneously extracting spatiotemporal features from time-series remote sensing images in an automatic manner (Ji et al., 2013). Therefore, it is capable of capturing more complicated spatial and temporal information.

The objective of this study is to develop a hybrid method by combining the ATC model and the 3D-CNN model for reconstructing high spatiotemporal LSTs, especially in complex urban areas.

2. Study area and data processing

2.1. Study area

The study area covers the urban areas of Beijing, China, containing 14 urban and suburban districts and two rural counties (Fig. 1), with a coordinate range between 115.7°~117.4°E, 39.4°~41.6°N. As the capital of China, Beijing has experienced rapid urbanization development with rapid population growth and economic development since 1949. Major land cover types in the study area include farmland, forest, grassland, bushes, wetlands, waters, impervious surfaces, and bare land.

Beijing owns a typical temperate monsoon climate characterized by cold and dry winters, hot and rainy summers, and short spring and autumn. The UHI (Urban Heat Island) effect has become increasingly severer in Beijing in recent years, and many scholars have investigated this issue in an extensive manner (Meng et al., 2018; Qiao et al., 2013; Quan et al., 2014). However, the lack of high spatiotemporal resolution temperature datasets in urban systems greatly limits our understanding of the spatiotemporal characteristics and driving factors of SUHI (Surface Urban Heat Island) and their environmental impacts. A seamless LSTs dataset with consistent spatial and temporal information largely benefits the investigation of spatial and temporal changes of UHI in Beijing, providing essential references for understanding the urban thermal environment and for the decision-making in public health and sustainable development.

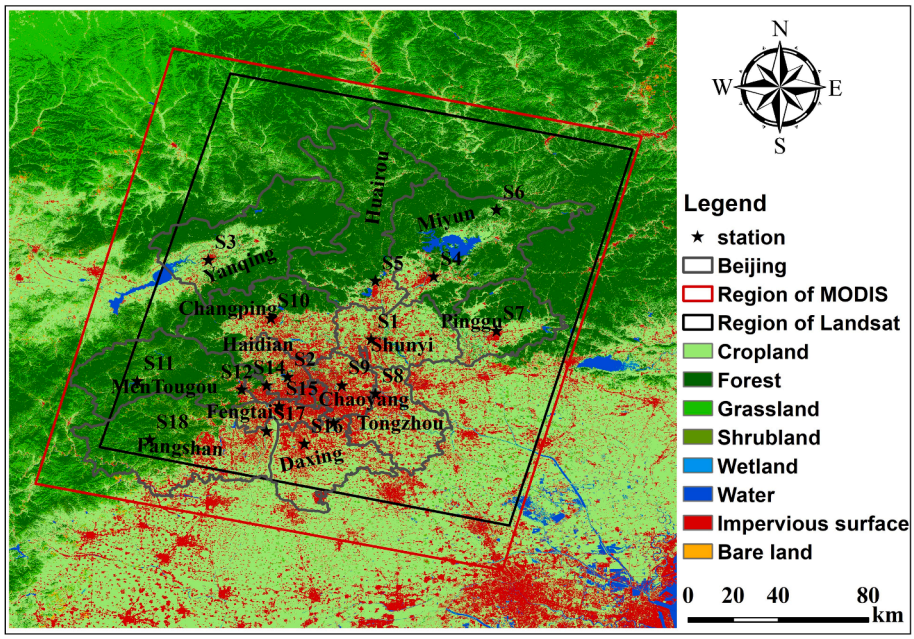


Fig. 1. The study area (Beijing, China).

2.2. Data processing

Detailed information regarding the datasets used in this study are shown in Table 1. This study employed Landsat 8 and MODIS satellite data for the year 2017. The MODIS satellite data derived from the NASA Earth Science Data (https://ladsweb.modaps.eosdis.nasa.gov/) include MODIS daily LSTs product MOD11A1 (onboard Terra satellite, daytime (~10:30 am)) and 16-day normalized difference vegetation index (NDVI) product (MOD13A2). The MODIS data were resampled to 1000 m and re-projected to the Universal Transverse Mercator (UTM) projection system with the Datum of World Geodetic System 1984 (WGS84) and cropped to the study area using the Beijing administration boundary. A total of 365 scenes of MODIS LSTs images were collected, with 121 scenes having the clear sky pixels ratio of 0–30%, 69 scenes the ratio of 30–60%, and 175 scenes the ratio of 60–100%.

In this study, Level-2 Landsat-8 Surface Reflectance products and raw TIR images between January and December 2017 were downloaded from the USGS website (https://espa.cr.usgs.gov/). The Landsat (30 m) images were resampled to 100 m. A total of 22 Landsat 8 TIRS images and 365 MODIS LSTs images were obtained in 2017. For Landsat 8 TIRS images, 11 scenes were all clear sky pixels, accounting for 50% of the total images. There were five scenes with strong coverage of clouds and shadows, among which four scenes occurred in summer. The unequal distribution of available scenes throughout the year, i.e., more clear scenes during the autumn and winter, fewer scenes during the winter and summer, again highlights the necessity of developing a robust LST reconstruction model.

Hourly surface air temperatures (SATs) from 18 meteorological stations in 2017 were collected from National Meteorological Information Center (http://data.cma.cn/). The SATs were resampled into raster images with a resolution of 100 m and 1 km, using the IDW method to respectively match the Landsat and MODIS LST data.

The global land cover (GLC) product 2017 (Gong et al., 2019) was employed to group modeling results by land cover types (http://data.ess.tsinghua.edu.cn/).

Table 1
Detailed data information.

variable	products	Spatial resolution	Temporal resolution	Data source
MODIS LSTs	MOD11A1	1000 m	Daily	NASA
NDVI	MOD13A2	1000 m	16-Day	NASA
Lansat 8 OLI/ TIRS	Surface Reflectance products	100 m	16-Day	USGS
SATs	Hourly SATs	/	houly	National Meteorological Information Center
Global land cover product	GLC_10	10 m	Yearly	Tsinghua university
Administration boundary	Beijing administration boundary	/	/	

3. Method

The proposed method includes four major components (Fig. 2). As LSTs can be affected by weather conditions such as cloud cover, wind speed, air humidity, and land cover types, we first add SATs, as the auxiliary data, to the model. Secondly, the missing LST data are reconstructed using ATCs/ATCe models in a coarse manner. Then, a 3D-CNN model is constructed to reconstruct missing LSTs considering the SATs data in a fine manner. Finally, the reconstructed results are evaluated using the observed LSTs from satellite images.

3.1. Annual temperature cycle modeling

The standard annual temperature cycle (ATCs) model, proposed by Bechtel (Bechtel, 2012, 2011), uses a single sinusoidal function to describe the dynamics of annual LST. The ATCs model can be defined as follows (Eq. (1)):

$$LST(d) = \varphi(MAST, YAST, \theta, d) = MAST + YAST^{*}\sin\left(\frac{2d\pi}{365} + \theta\right) \quad (1)$$

where d is the day relative to the spring equinox, $MAST$ is the mean annual value of LST, $YAST$ is the ATCs yearly amplitude, and θ is the phase shift relative to the spring equinox.

The physical foundation of the ATCs model requires that the annual temperature dynamics are mainly controlled by the solar radiation flux

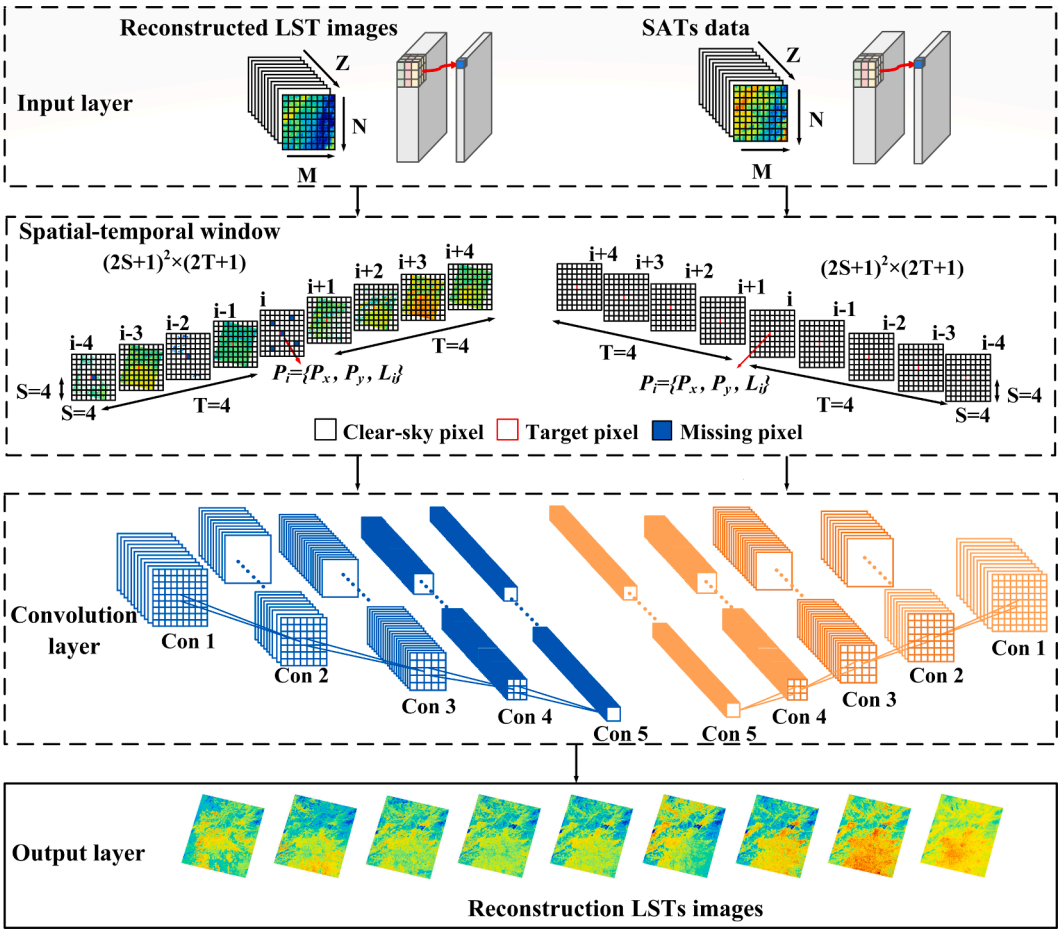


Fig. 2. The workflow of the proposed model.

at the top of the atmosphere, and the land surface properties remain unchanged during the annual cycle. However, LSTs' variation owns a harmonic nature, which is affected by land surface alterations, weather fluctuations, and human interference. Therefore, auxiliary data (i.e., NDVI, soil moisture, and SATs) related to LSTs variations were incorporated to establish an ATC model that is sufficiently stable and can represent short-term LST fluctuations. Zou et al. (2018; 2017) proposed an enhanced annual temperature cycle model (ATCe) by considering synoptic conditions and vegetation phenology to estimate LSTs fluctuations using SATs and NDVI as auxiliary data (Eq. (2)):

$$LST(d) = MAST + YAST^{*}\sin\left(\frac{2d\pi}{365} + \theta\right) + \Delta T_{air}d^{*}\gamma(d) \tag{2}$$

where $\Delta T_{air}d$ and $\gamma(d)$ can be expressed by Eq. (3) and Eq. (4).

$$\Delta T_{air}d = T_{air}d - MAST_{air} + YAST_{air}{}^{*}sin\left(\frac{2d\pi}{365} + \theta_{air}\right) \tag{3}$$

$$\gamma(d) = \lambda{\cdot}(NDVI_{max} - NDVI_{min})/[NDVI(d) - NDVI_{min} + 1] \tag{4}$$

where $T_{air}d$ is the daily mean SATs $MAST_{air}, YAST_{air}, \theta_{air}$ are the annual mean, amplitude, and the corresponding phase shift of the SATs, respectively; λ is a multiplier; $NDVI_{max}$ and $NDVI_{min}$ are the maximum and minimum vegetation index values, respectively; $NDVI(d)$ represents the vegetation index at a specific day.

3.2. Spatiotemporal reconstruction of LSTs based on 3D-CNN model

3.2.1. CNN models

Convolutional neural networks (CNNs) facilitate the automatic learning of internal feature relationships between different data sources and are advantageous in solving highly nonlinear and dynamic problems, which are mainly used in two-dimensional images. In a 2D-CNN, the convolutional layer performs 2D convolution to extract the local neighborhood features of the feature map of the previous layer from the spatial dimension only. Then, the additive bias is applied, and the result is passed through an activation function. However, the above processes are not applicable for spatiotemporal LST data. For spatiotemporal problems, it is necessary to capture spatial and temporal information in time-series LSTs images. Compared to the 2D convolution, the 3D convolution operation is applied to a 3D cube generated by stacking multiple contiguous frames. To effectively incorporate the spatial and temporal information in LST reconstruction, we employ 3D convolution layers of CNNs so that discriminative features in both spatial and temporal dimensions are captured (Eq. (5)).

$$v_{ij}^{xyz} = \mathrm{g}\left(b_{ij} + \sum_{m}\sum_{p=0}^{P_i}\sum_{q=0}^{Q_i-1}\sum_{r=0}^{R_i-1}\omega_{ijm}^{pqr}v_{(i-1)m}^{(x+p)(y+q)(z+r)}\right) \tag{5}$$

where v is the output variable in the feature map.P, Q, and R are the size of the 3D kernel along spectral, spatial, temporal dimensions, respectively.p, q, r are the indexes of the kernel. x, y, z are the indexes of the feature map. i, j, m are the indexes of the input layer, output layer

and feature map, respectively. ω_{ijm}^{pqr} is the pqr-th value of the kernel connected to the m-th feature map. b_{ij} is the bias term for feature map.

3.2.2. Spatiotemporal reconstruction of LSTs

In this study, the 3D-CNNs aim is to reconstruct spatiotemporal in a fine manner. It contains the following three steps.

Step 1: Extracting training, test and validation samples. The 5% of the clear sky pixels are randomly selected as the validation samples when the clear sky pixels are more than 30% of the total number of images on a specific day. After partitioning out the validation set, the remaining data will be used as the training set and test set of the model in the ratio of 7:3.

Step 2: Constructing the 3D-CNN model (Fig. 3). Given an LST image set $L = \{L_1, L_2, \cdots\cdots, L_{n-1}, L_n\}$ that corresponds to different acquisition times, for a given pixel P, the position of the pixel through time by indexing through L can be observed as $P_i = \{P_x, P_y, L_i\}$. The spatial and temporal radius of a missing pixel is defined as S and T, respectively. Thus, the size of the spatial and temporal window are respectively $[2S+1, 2S+1]$ and $[i-T, i+T]$. The depth of the temporal window is $2T+1$ and the size of image cubes is $[2S+1, 2S+1, 2T+1]$. In this study, S and T are set to 4, the size of the spatiotemporal window is $9 \times 9 \times 9$. Further, the sample with cubes from the image to be reconstructed and auxiliary data (SATs) are employed as the input layer, followed by 3D convolutions with 3D convolution kernels. Mean Absolute Error (MAE) is used as the Loss function (Eq. (6)):

$$MAE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{6}$$

where y_i is the observed LSTs value, $\hat{y}_i$ represents the reconstruction LSTs value, and N denotes the number of pixels.

To reduce the instability caused by the Loss function oscillation, we adopted the dynamic learning rate in this study:

$$decayed_{learning_{rate}} = learning_{rate} \times decay_{rate} \wedge \left(global_{step}/decay_{steps}\right) \tag{7}$$

where $\text{learning}_{\text{rate}}$ is the initial learning, $\text{global}_{\text{step}}$ is the number of iterations, $\text{decay}_{\text{steps}}$ is decay period, $\text{decay}_{\text{rate}}$ represents decay-rate. In this study, we set $\text{learning}_{\text{rate}} = 1e-3$, $\text{decay}_{\text{rate}} = 0.99$, $\text{decay}_{\text{steps}} = 1$, $\text{global}_{\text{step}} = 3200$, Batchsize = 32, and iteration = 3200. To avoid overfitting, we save the model on every ten iterations. The model with the least validation (Val) Loss is selected as the optimal model for LSTs reconstruction.

We construct experiments that include a total of 12 groups. Details are documented in Table 2.

Step 3: Assessing accuracy. The models' performance is evaluated using the validation dataset. The determination coefficient R^2 and Root Mean Squared Error (RMSE) are used to assess the performance.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{8}$$

$$R^2 = \frac{SS_{regression}}{SS_{total}} = 1 - \frac{SS_{residual}}{SS_{total}} = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y_i})^2} \tag{9}$$

where y_i is the observed LSTs value, $\hat{y}_i$ represents the reconstruction LSTs value, $\overline{y_i}$ means the average of LSTs, N denotes the number of pixels, SS_{total} represents the sum of squared deviations, $SS_{regression}$ represents the regression sum of squares, and $SS_{residual}$ represents the residual sum of a square.

4. Results

4.1. Performance of 3D-CNN model

At the beginning of training, the MAE value with the 3D-CNN model reaches the maximum, followed by a rapid decrease. It further reaches a stable state with slight fluctuations when the iterations reach 500 for both Landsat 8 LSTs and MODIS LSTs (Fig. 4 and Fig. 5). The minimum validation (Val) Loss values are 0.27 K (iterations = 2540), 0.26 K

Table 2
Experimental groups.

Reconstruction models	ATCs	ATCe	SATs	3D-CNN
Landsat ATCs	√			
Landsat ATCe		√		
Landsat ATCs & CNN	√			√
Landsat ATCe & CNN		√		√
Landsat ATCs & CNN & SATs	√		√	√
Landsat ATCe & CNN & SATs		√	√	√
MODIS ATCs	√			
MODIS ATCe		√		
MODIS ATCs & CNN	√			√
MODIS ATCe & CNN		√		√
MODIS ATCs & CNN & SATs	√		√	√
MODIS ATCe & CNN & SATs		√	√	√

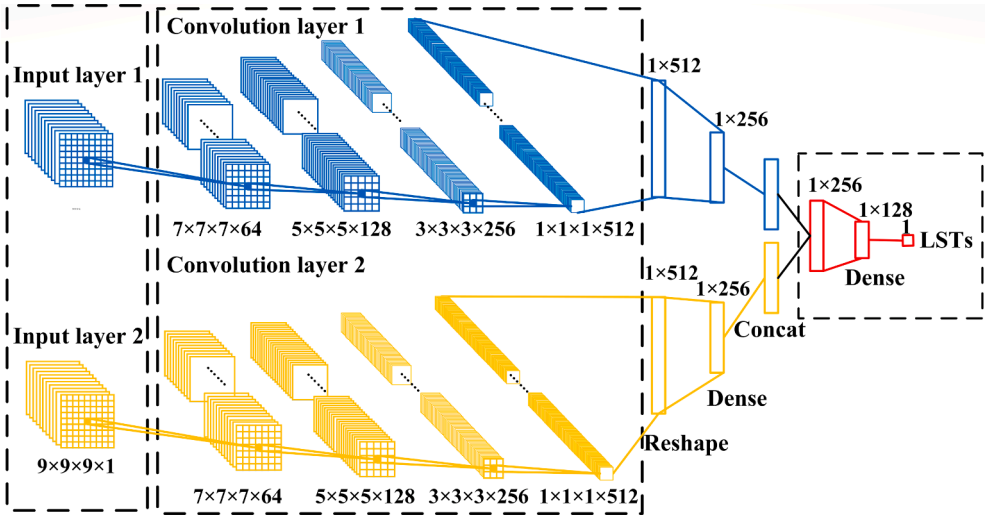


Fig. 3. the proposed 3D-CNN model for the fine spatiotemporal reconstruction of LSTs.

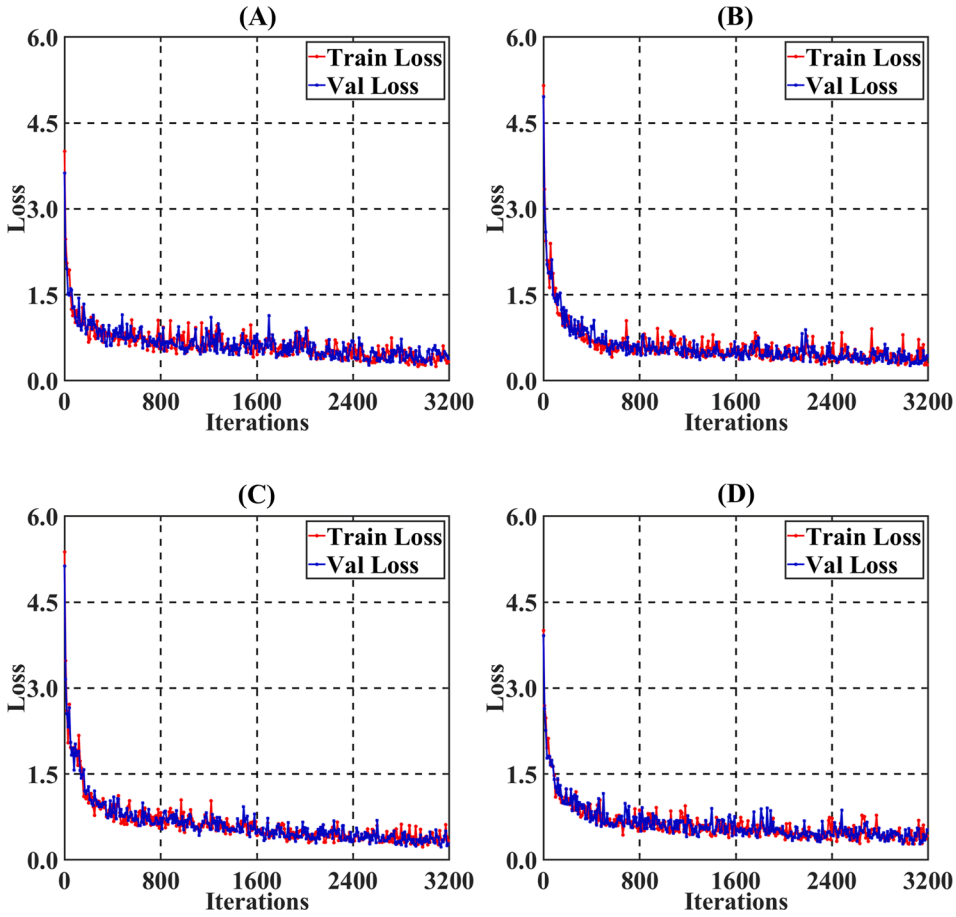


Fig. 4. Train Loss and Validation Loss from Landsat LSTs (A-D represent for Landsat ATCs&CNN, Landsat ATCe&CNN, Landsat ATCs&CNN&SATs, and Landsat ATCe&CNN&SATs, respectively).

(iterations = 3060), 0.23 K (iterations = 2540), 0.28 K (iterations = 3080) for Landsat ATCs&CNN, Landsat ATCe&CNN, Landsat ATCs&CNN&SATs and Landsat ATCe&CNN&SATs respectively. The corresponding train Loss values are 0.31 K, 0.37 K, 0.36 K, 0.29 K, respectively.

The minimum Val Loss values are 0.46 K (iterations = 3120), 0.48 K (iterations = 2300), 0.34 K (iterations = 2820), and 0.39 K (iterations = 3080) for MODIS ATCs&CNN, MODIS ATCe&CNN, MODIS ATCs&CNN&SATs, and MODIS ATCe&CNN&SATs respectively. Their corresponding train Loss values are 0.9 K, 0.51 K, 0.62 K, and 0.70 K, respectively. We notice that ATCs&CNN&SATs owns the lowest Val Loss in both Landsat 8 LSTs and MODIS LSTs. Compared with Landsat LSTs (0.3–1.0 K), the fluctuation of Loss values in MODIS LSTs (0.5–3.0 K) is more intense than that in Landsat LSTs (0.3–1.0 K).

4.2. *Spatiotemporal pattern of reconstruction LSTs*

The reconstruction results from different models are displayed in supplementary materials (see Section1-Reconstructed Landsat LSTs and Section2-MODIS LSTs), where columns 1–8 respectively represent the original LSTs data, LSTs to be reconstructed, LSTs reconstructed by ATCs, ATCe ATCs&CNN, ATCe&CNN, ATCs&CNN&SATs, and ATCe&CNN&SATs. Fig. 6 and Fig. 7 only present the spatiotemporal patterns of reconstructed Landsat LSTs and MODIS LSTs on different seasons with cloud cover percentages of more than 60%.

For Landsat LSTs, Except for the selected 5% as the validation set, the rest of the data are all clear-sky pixels for the day of year (DOY) 31, DOY 63, DOY 127, DOY 143, DOY 191, DOY 255, DOY 271, DOY 303, DOY 319, DOY 335, and DOY 351 (see Supplementary materials: Section1-Reconstructed Landsat LSTs). The simulated LSTs from Landsat ATCe&CNN, Landsat ATCe, and Landsat ATCe&SATs&CNN are similar to the observed LSTs for DOY 31, DOY 335 and DOY 351 images (all in winter), respectively. For DOY 31, we notice that the reconstructed LSTs from Landsat ATCs, Landsat ATCe, Landsat ATCs&CNN, and Landsat ATCe &CNN are higher than observed LSTs, especially in urban centers with high impervious surface coverage. The reconstructed LSTs for Landsat ATCe&SATs&CNN are lower than observed LSTs. For DOY 335, the estimated LSTs from Landsat ATCs, Landsat ATCs&CNN, Landsat ATCe&CNN, Landsat ATCs&SATs&CNN, Landsat ATCe&SATs&CNN are all higher than the actual LSTs. For DOY 351, the reconstructed LSTs for Landsat ATCs, Landsat ATCe Landsat ATCs&CNN, Landsat ATCe&CNN, and Landsat ATCs&SATs&CNN are all higher than the actual satellite data LSTs. In spring, the reconstructed LSTs from Landsat ATCs and Landsat ATCe are lower than the observed LSTs on DOY 63 and DOY 127 (Fig. 6), especially in urban cores. However, the opposite patterns can be observed on DOY 143. The reconstructed LSTs from Landsat ATCs&CNN, Landsat ATCe&CNN, Landsat ATCs&CNN&SATs, Landsat ATCe&CNN&SATs are similar to the actual observed LSTs on DOY 63 and DOY 127. In summer, the LSTs reconstructed by Landsat ATCs and Landsat ATCe are generally higher than the observed LSTs (Fig. 6). In autumn, the reconstructed LSTs are close to the original LSTs on DOY 255 and DOY 271, except for the LSTs derived from Landsat ATCs. On DOY 303, the reconstructed LSTs fromLandsat ATCs, Landsat ATCe and Landsat ATCs&CNN&SATs are all lower than the observed LSTs. Such a pattern occurs in both urban and rural areas. On DOY 319, the reconstructed LSTs from Landsat ATCs&CNN and Landsat ATCe are lower

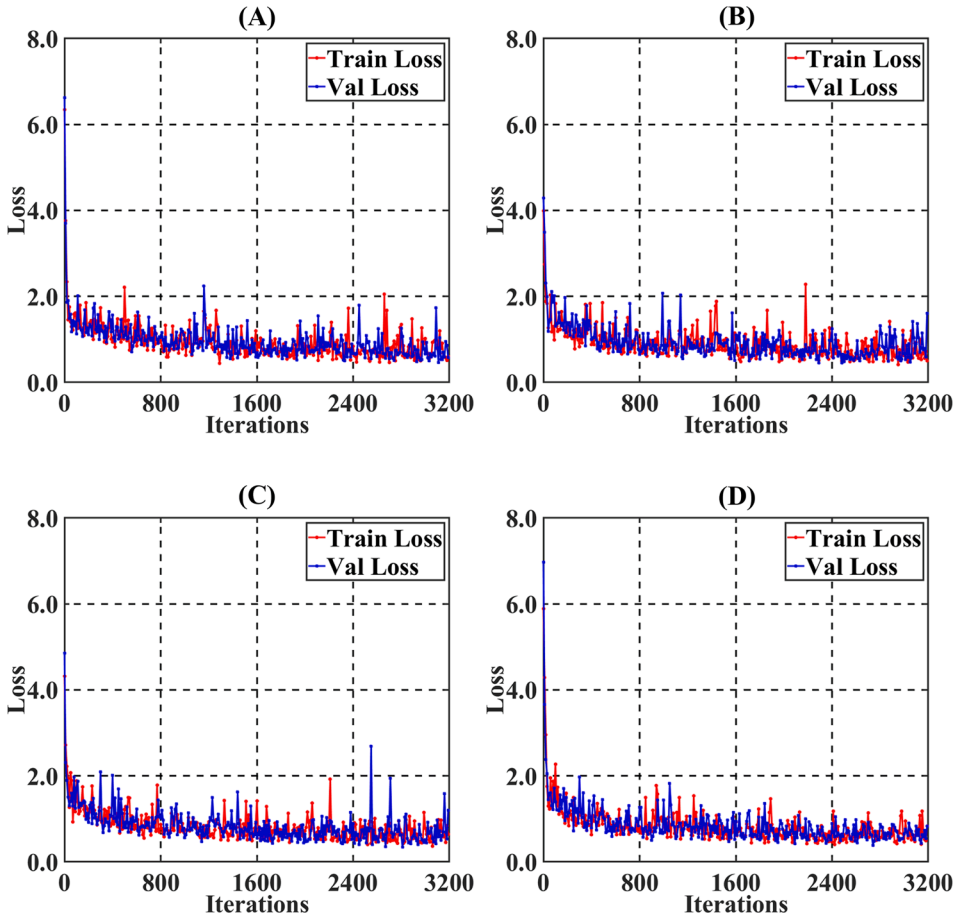


Fig. 5. Train Loss and Validation Loss for MODIS LSTs (A-D represent MODIS ATCs&CNN, MODIS ATCe&CNN, MODIS ATCs&CNN&SATs, and MODIS ATCe&CNN&SATs, respectively)

than the original satellite LSTs, while the LSTs from Landsat ATCe&CNN are similar to the actual LSTs.

For images with a clear sky pixel ratio of about 60%, images on DOY 95 and DOY 111 were both obtained in spring. The missing pixels of DOY 95 and DOY 111 are mainly in the east and southeast and in the mountainous areas in the northwest, respectively. In general, in these two images, the reconstructed LSTs from all models in the urban cores and the southern suburbs are close to the observed LSTs, except for the LSTs from Landsat ATCs and Landsat ATCe.

The reconstructed LSTs from MODIS ATCs and ATCe are higher than observed LSTs on DOY 127 (spring), DOY 191 (summer), and DOY 351 (winter), but lower than actual LSTs on DOY 303 (autumn). The results of MODIS LSTs are generally consistent with the results of Landsat LSTs. MODIS LSTs from different models are shown in supplementary materials (see Section2-Reconstructed MODIS LSTs).

4.3. Accuracy assessment

4.3.1. Different clear-sky pixel percent

To understand the impacts of clouds (or the number of samples available on a per-pixel basis) on the LST reconstruction accuracy, we validate the reconstructed results given the proportion of clear sky pixels: 0–30%, 30–60%, 60–100%, and 30–100%. The use of a large interval (30%) rather than a small one (e.g., 5%, 20%) is to provide variation in the number of images (or samples) for understanding such impacts on the reconstruction accuracy. The scatter density plot between Landsat/MODIS LSTs observed values and the reconstructed from different models under different clear-sky pixel percent are shown in Fig. S1 (Landsat LSTs: 30–60%), Fig. S2 (MODIS LSTs: 30–60%), Fig. S3 (Landsat LSTs: 60–100%), Fig. S4 (MODIS LSTs: 60–100%), Fig. 8 (Landsat LSTs:30–100%) and Fig. 9 (MODIS LSTs: 30–100%) (see Supplementary materials: Section3-Appendix). Sub-figures A-F respectively show the reconstructed results observations from Landsat ATCs, Landsat ATCe, Landsat ATCs&CNN, Landsat ATCe&CNN, Landsat ATCs&CNN&SATs, and Landsat ATCes&CNN&SATs.

Compared with 60–100% (Fig. S2) and 30–100% (Fig. 8), the RMSE of 30–60% (Fig. S1) is the lowest in all models, except for Landsat ATCs ($R^2 = 0.55$, RMSE = 3.23 K) and Landsat ATCe (R2 = 0.58, RMSE = 2.66 K). In general, the reconstruction accuracies are similar under clear-sky pixel percentages of 60–100% and 30–100%. Note that there are only two scenes with clear sky pixels between 30% and 60% (DOY 95 and DOY 111). Such a small amount of validation data is responsible for the low RMSEs. Fig. 8 (Fig. 9) presents the RMSE and R^2 between the observed LSTs from Landsat (MODIS) images and the fitted LSTs. The R^2 values of Landsat ATCs and Landsat ATCe are 0.88 and 0.93, respectively, while the RMSEs are 4.21 K and 3.25 K, respectively. The auxiliary data (NDVI and SATs) lead to the improved reconstruction of LSTs compared with ATCs. The reconstruction accuracy of the 3D-CNN model is about 2–3 times better than the ones of the ATCs and ATCe model. The R^2 of Landsat ATCs&CNN, Landsat ATCe&CNN, Landsat ATCs&CNN&SATs, Landsat ATCe&CNN&SATs are 0.97, 0.99, 0.98, 0.99, respectively, and the RMSEs are all within 2 K (1.85 K, 1.13 K, 1.56 K, and 0.96 K) and evenly distributed around the 1:1 line, indicating no observable biases in their reconstruction results. The RMSE of

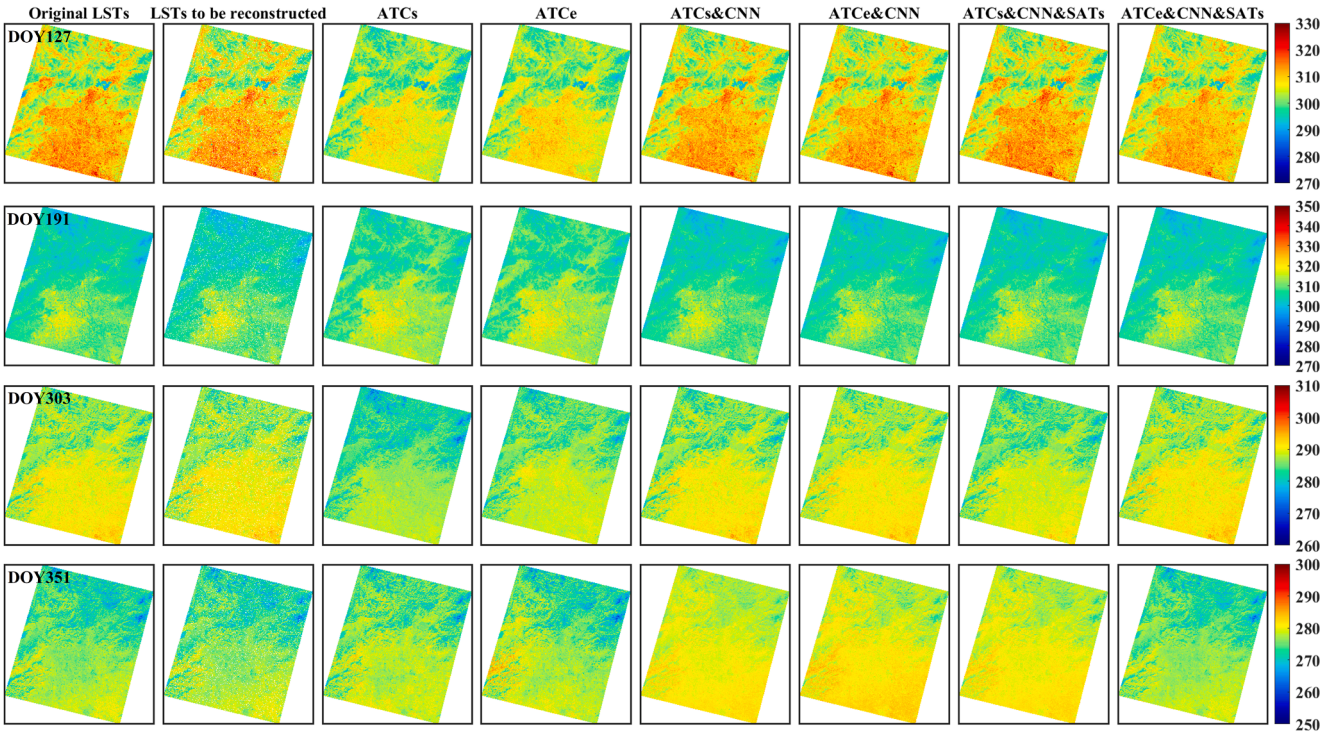


Fig. 6. The reconstructed LSTs for Landsat in DOY 127 (spring), DOY191 (summer), DOY303 (autumn), DOY351 (winter). Images in columns 1–8 respectively represent the original Landsat LSTs data, LSTs to be reconstructed, LSTs reconstructed from Landsat ATCs, Landsat ATCe, Landsat ATCs&CNN, Landsat ATCe&CNN, Landsat ATCs&SATs&CNN, and Landsat ATCe&SATs&CNN.

Landsat ATCe&CNN is smaller than Landsat ATCs&CNN. We notice that Landsat ATCe&CNN outperforms Landsat ATCs&CNN&SATs. Note that the RMSE from Landsat ATCe&CNN&SATs (RMSE = 0.96 K) is one-fourth of the one from Landsat ATCs (RMSE = 4.21 K) and half of the one from Landsat ATCs&CNN (RMSE = 1.85 K), indicating that the involvement of SATs as auxiliary data during the LST reconstruction

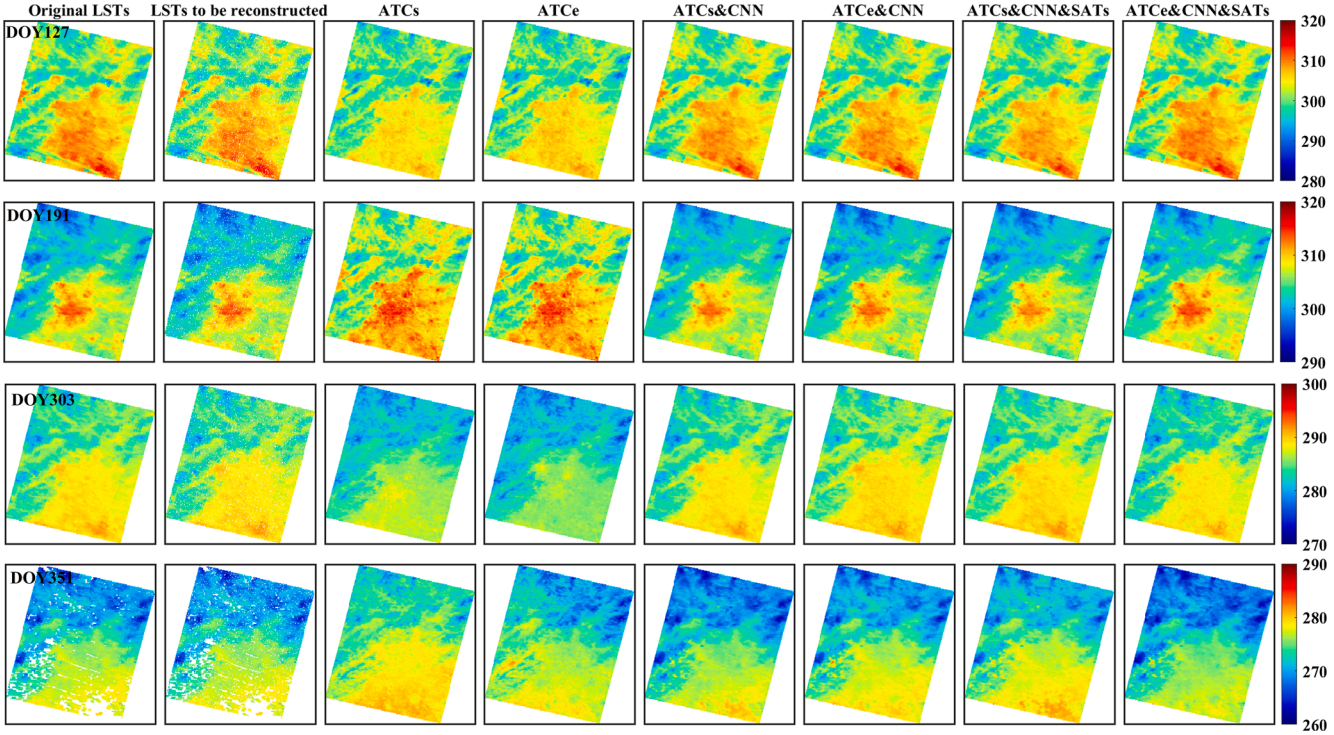


Fig. 7. The reconstructed LSTs for MODIS in DOY 127 (spring), DOY191 (summer), DOY303 (autumn), DOY351 (winter). Images in columns 1–8 respectively represent the original MODIS LSTs data, LSTs to be reconstructed, reconstructed LSTs from MODIS ATCs, MODIS ATCe, MODIS ATCs&CNN, MODIS ATCe&CNN, MODIS ATCs&SATs&CNN, and MODIS ATCe&SATs&CNN.

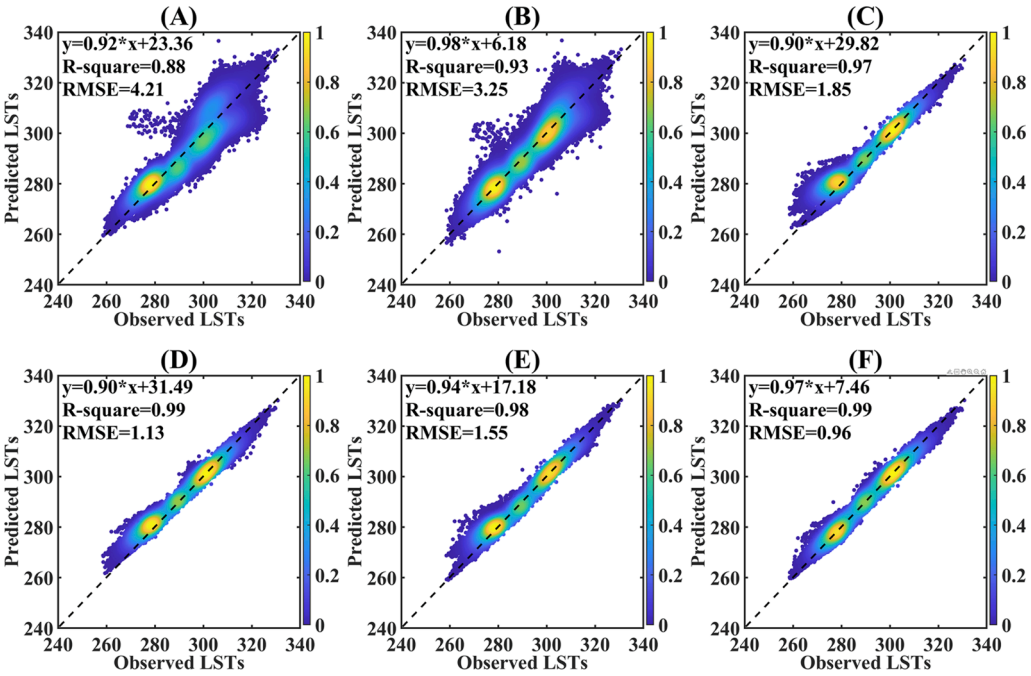


Fig. 8. Accuracy assessment for Landsat LSTs with clear-sky pixel percentages of 30–100%. (A-F present Landsat ATCs, Landsat ATCe, Landsat ATCs&CNN, Landsat ATCe&CNN, Landsat ATCs&CNN&SATs, and Landsat ATCe&CNN&SATs, respectively.

process improve the reconstruction accuracy.

For MODIS LSTs, the reconstructed results are very similar compared with Landsat LSTs: the best results come from clear sky pixel percentage of 60–100% (Fig. S4), followed by 30–100% (Fig. 9), and 30-60% (Fig. S3). Taking the 30-100% as an example, the reconstructed result from MODIS ATCe (R^2 = 0.93, RMSE = 3.15 K) is better than the one from MODIS ATCs (R^2 = 0.90, RMSE = 3.83 K). The RMSEs of MODIS ATCs&CNN (0.78 K), MODIS ATCe&CNN (0.71 K), MODIS ATCs &CNN&SATs (0.61 K), MODIS ATCe &CNN&SATs (0.75 K) are all within 1.0 K and the R^2 all reach 0.99. The reconstruction accuracy of

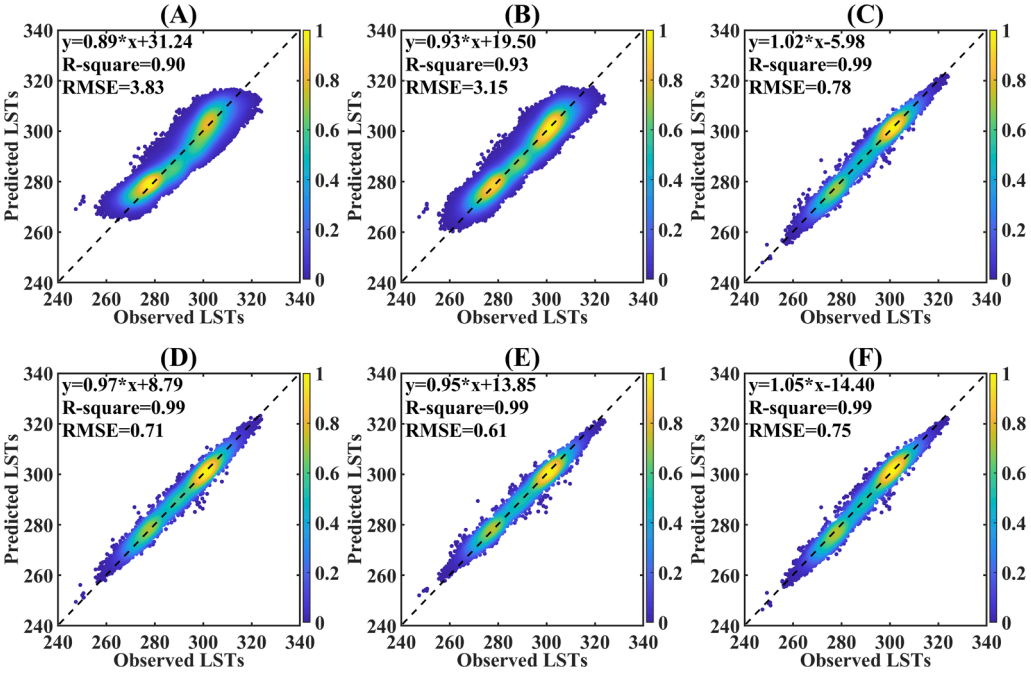


Fig. 9. Accuracy assessment for MODIS LSTs with clear-sky pixel percentages of 30–100%. (A-F 1–6 present Landsat ATCs, MODIS ATCe, MODIS ATCs&CNN, MODIS ATCe&CNN, MODIS ATCs&CNN&SATs, and MODIS ATCe&CNN&SATs, respectively.

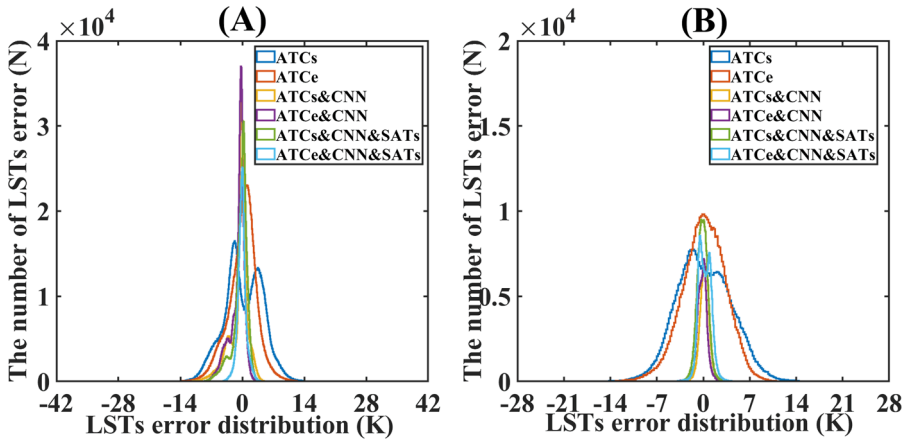


Fig. 10. The histogram of the error distribution between observed LSTs and reconstructed LSTs from selected models. (A) Landsat LSTs (B) MODIS LSTs.

MODIS ATCe&CNN is similar to that of MODIS ATCs&CNN with an RMSE difference of only 0.07 K. MODIS ATCs&CNN&SATs outperforms MODIS ATCs&CNN, evidenced by its reduced RMSE (from 0.78 K to 0.61 K). RMSEs from MODIS ATCe&CNN&SATs and MODIS ATCe&CNN are relatively close, with a difference of only 0.04 K. Therefore, we can conclude that involving the 3D-CNN model results in improved reconstruction performance, while the addition of SATs as auxiliary data fails to further improve the reconstruction results.

Fig. 10 represents the error distribution of reconstructed LSTs and observed LSTs from Landsat and MODIS. For Landsat LSTs and MODIS LSTs, ATCs&CNN, ATCe&CNN, ATCs&CNN&SATs, ATCe&CNN&SATs tend to have a relatively concentrated error distribution. Table 3 shows the percentage of the error distribution in different ranges. For Landsat LSTs, the proportions of errors within 1 K are 13% for ATCs, 27% for ATCe, 57% for ATCs&CNN, 60% for ATCe&CNN, 61% for ATCs&CNN&SATs, and 75% for ATCe&CNN&SATs, respectively. We observe that the involvement of the 3D-CNN model leads to a higher proportion of errors within 1 K (increased from 13% to 57%). The proportion of errors within 1 K, 2 K, and 4 K for ATCe&CNN&SATs have reached 75%, 93.72%, 99.78%, respectively. The proportion of errors within 4 K are 63% for ATC, 80.64% for ATCe, 90.26% for ATCs&CNN, 91.48% for ATCe&CNN, 94.60% for ATCs&CNN&SATs, and 99.77% for ATCe&CNN&SATs, respectively.

In general, the error distribution of MODIS LSTs is similar to the one of Landsat LSTs. ATCs&CNN presents the highest error distribution within 1 K (83.08%), followed by ATCe&CNN (83.87%), ATCs&CNN&SATs (79.20%), and ATCe&CNN&SATs (71.01%). The ATCs and ATCe models have the lowest error distribution within 1 K, i.e., 17.07% and 24.65%, respectively. Except for ATCs (35.05%) and ATCe (46.65%), errors within 2 K account for above 96%.

4.3.2. Validation of LSTs reconstruction results under different land cover types

In this session, we validate the reconstruction results under different land cover types and establish the regression function by comparing them against the satellite-observed LSTs in each land cover type (see Supplementary materials: Section3-Appendix Table S1 and Table S2). Fig. 11 shows the RMSE of the different models for different land cover types from Landsat and MODIS.

For Landsat LSTs (see Supplementary materials: Section3-Appendix Table S1 and Fig. 11.A), the reconstruction accuracy of different land cover types basically shows the following pattern: ATCe&CNN&SATs > ATCe&CNN > ATCs&CNN&SATs > ATCs&CNN > ATCe > ATCs. The RMSE of the ATCs model is the highest in Grassland (4.48 K) and the lowest in Water bodies (2.7 K), with an R^2 of 0.86 and 0.94. The RMSEs of Cropland, Forest, shrubland, and bareland are all above 4 K. However, the RMSEs of the ATCe model are improved around 0.7 to 1.4 K for different land covers, especially in vegetation (up to 1.1 K). The results are comparable with a study by Zou et al. (2018), who stated that the NDVI data could improve the performance of the ATC in vegetation areas, as phenology plays a vital role in the LST fluctuations. The RMSEs of ATCs&CNN and ATCe&CNN display a sharp decrease, ranging from 1.0 K/1.6 K(ATCs&CNN/ ATCe&CNN) in water bodies to 3.0 K/3.5 K (ATCs&CNN/ATCe&CNN) in vegetation area. The results illustrate that 3D-CNNs models can effectively complement and enrich the performance of the ATC model. However, the accuracy of ATCs&CNN& SATs and ATCe&CNN& SATs show a gentle increase around 0.1–0.3 K compared with the ATCs&CNN/ATCe&CNN, presumably relating to the location of stations, the interpolation method, and the time difference between the SATs and satellite observation. The best accuracy comes from ATCe&CNN&SATs, with RMSEs all close to 1 K (0.84–1.10 K).

The mean value of the reconstructed LSTs is closer to the one of the observed LSTs, fluctuating around 1 K. However, great differences can be observed between the maximum and minimum values. For example, the difference between the maximum value of the ATCe model (348.61 K) and the maximum value obtained by satellite observation (327.00 K) in farmland is 21.61 K, and the difference between the minimum value is nearly 10 K (253.21 K/262.09 K).

Table 3
The percentage of the error distribution between observed LSTs and reconstructed LSTs.

Landsat/MODIS						
Percents	ATCs	ATCe	ATCs&CNN	ATCe&CNN	ATCs&CNN&SATs	ATCe&CNN&SATs
<1 K	13.41/17.07	27.49/24.65	57.38/83.08	60.03/83.87	60.71/79.2	74.68/71.01
<2 K	16.81/17.99	24.05/22	14.58/14.31	16.93/13.54	23.24/18.32	19.04/25.26
<4 K	32.74/31.77	29.1/31.96	18.3/2.37	14.51/2.38	10.65/2.3	6.05/3.47
<6 K	21/19.82	12.06/14.99	6.28/0.19	7.06/0.17	4.15/0.14	0.19/0.22
>6 K	16.04/13.35	7.29/6.41	3.46/0.04	1.46/0.04	1.25/0.03	0.03/0.04
MIN	−37.09/−23.54	−40.83/−20.67	−21.55/−16.82	−16.18/−15.32	−20.28/−17.76	−13.63/−15.25
MAX	24.2/16.31	28.23/13.76	9.34/12.53	10.42/13.47	8.36/14.91	7.22/13.63

Note: all values are in percentages.

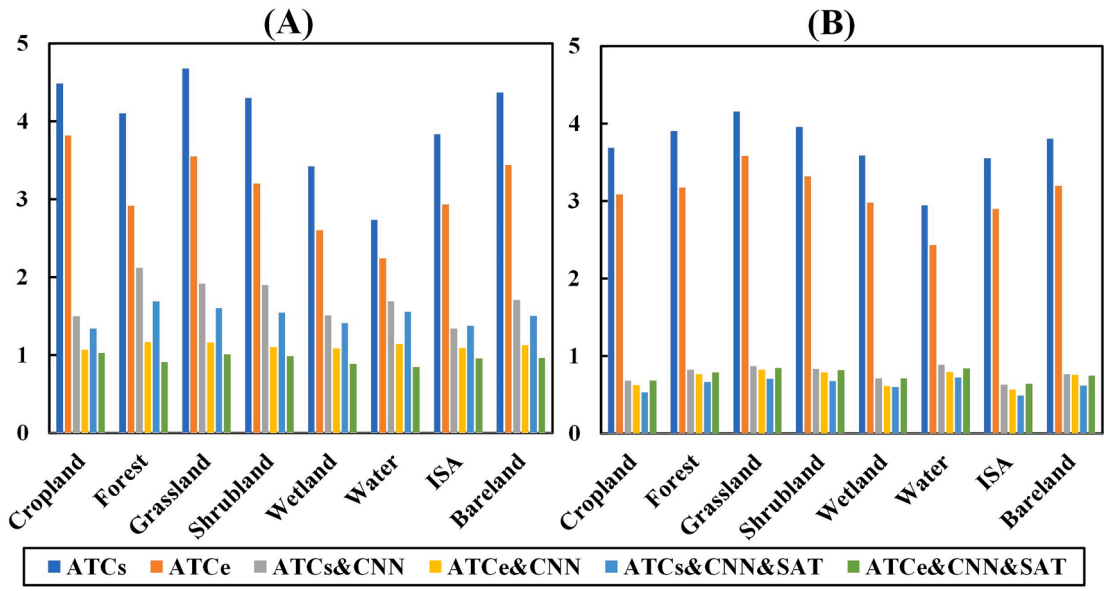


Fig. 11. RMSE of the different models for different land cover types from Landsat (A) and MODIS (B).

The reconstruction results of MODIS LSTs are slightly different from Landsat LSTs(see Supplementary materials: Section3-Appendix Table S2). In general, ATCs&CNN&SATs shows the best performance while ATCs shows the worst. Except for bare land, models' performances show the following pattern:ATCs&CNN&SATs > ATCe&CNN > ATCe&CNN&SATs > ATCs&CNN > ATCe > ATCs (see Supplementary materials: Section3-Appendix Table S2). The reconstructed MODIS LSTs from the ATCs and ATCe models are similar to Landsat, with higher RMSEs in cropland, forest, grassland, shrubland, and bareland. ATCs&CNN&SATs presents the best performance among all the land cover types. Water shows the largest RMSE (0.72 K), while the impervious surface shows the smallest RMSE (0.49 K). Meanwhile, the RMSEs from 3D-CNN and models with SATs as auxiliary data are less than 1 K in all land cover types(Fig. 11 B).

5. Discussion

5.1. Comparison of different methods and data sources

5.1.1. Accuracy of different methods

From the experimental results, we observe that the ATCe model for both Landsat and MODIS data outperforms ATCs. This result is mainly due to the fact that the ATCe model considers the LST short-term fluctuations due to the synoptic conditions (SATs) and surface properties (Zou et al., 2018). The accuracies for ATCs&CNN, ATCe & CNN, ATCs&CNN&SATs, and ATCe&CNN&SATs are improved by incorporating spatial and temporal information. ATCs and ATCe belong to the temporal framework, thus lacking the capability to capture spatial information. 3D-CNNs is able to complement and enrich the ATCs and ATCe models by embedding spatiotemporal information of LSTs.

5.1.2. Uncertainty induced by different data sources

By comparing the model performances for reconstructing Landsat and MODIS LSTs, we find that these models commonly perform better in MODIS than Landsat LSTs. This phenomenon may be attributed to the fact that the yearly data volume greatly differs between Landsat (22 scenes, 16 days) and MODIS (365 scenes, daily). MODIS LSTs can capture more temporal details compared with Landsat LSTs. We further observe that the addition of SATs has a gentle greater impact on Landsat LSTs and MODIS LSTs in the construction 3D-CNN model. This finding can be explained from the following aspects. First, the differences between the satellite overpass times and instantaneous SATs introduce a certain level of uncertainties to the model. Meanwhile, the influence of the IDW interpolation algorithm that leads to spatially continuous temperature distribution is also one of the factors that should be considered. Finally, the SATs are represented by the pixel-wise LST values. However, the land cover types contained in Landsat and MODIS pixels can be different, given their different spatial resolutions, leading to disparities in received solar radiation and the influence of surrounding ground objects (i.e., different surface-specific emissivity).

5.2. Limitations and future works

This study proposes a hybrid framework by combining ATC and 3D-CNN for reconstructing spatially and temporally continuous LSTs. The reconstruction framework shows better performance and flexibility compared with different competing methods. However, we acknowledge several limitations of the proposed method. First, insufficient auxiliary data (e.g., SATs and NDVI) may fail to completely capture the short-term fluctuations. Second, the inaccurate interpolation of SATs can introduce a certain level of uncertainties to the model performance. We believe this study can benefit from an improved interpolation method SATs. In addition, our proposed method presents better suitability under clear-sky conditions. Additional data should be used and expriments should be conducted to reflect conditions that include surface radiation fluxes, different cloud percentages, and microwave images. Finally, the effectiveness of the proposed LST reconstruction method should be further evaluated in more studies areas using multi-year data.

6. Conclusion

LST is a key indicator for understanding various important issues. However, the pixel quality of satellite-observed LSTs is largely affected by cloud, cloud shadow, and other atmospheric conditions, resulting in spatial discontinuity and temporal irregularity and posing great challenges to the spatiotemporal analysis in urban thermal environments. Most existing reconstruction algorithms target small spatial scale, single scene/several scene images, and large-scale (city/regional) and long-time series of LSTs data reconstruction are not universal.

In this study, we use SATs as auxiliary data and propose a hierarchical reconstruction method combining ATC models and 3D-CNN methods for reconstructing high spatiotemporal LSTs in urban areas. To validate the feasibility and effectiveness of the proposed method, we

retrieve 22 scenes of Landsat LSTs and 365 scenes of MODIS LSTs that cover Beijing, China, in the year 2017. We conduct experiments on six groups, i.e., ATCs, ATCe, ATCs&CNN, ATCe&CNN, ATCs&SATs&CNN, and ATCe&SATs&CNN. The LST reconstruction results are verified under different clear sky pixel percentages and different land cover types.

The results show that the reconstruction results involving auxiliary data and the RMSEs for the 3D-CNN model are at least three times lower than the ones from ATCs and ATCe models (RMSE: Landsat ATCe&SATs&CNN = 0.96 K, Landsat ATCs = 4.21 K; MODIS ATCs = 3.83 K; MODIS ATCs&SATs&CNN = 0.61 K). The RMSE is less than 2 K for Landsat and 1 K for MODIS, between the reconstructed LSTs and the observed LSTs. The conceptual, methodological, and experimental knowledge of this study is expected to benefit a variety of domains that require spatiotemporal continuous LSTs.

CRediT authorship contribution statement

Huyan Fu: Conceptualization, Methodology, Data curation, Formal analysis, Writing – original draft, Writing – review & editing, Investigation, Software, Funding acquisition. **Zhenfeng Shao:** Conceptualization, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition. **Peng Fu:** Writing – review & editing, Methodology, Formal analysis. **Xiao Huang:** Writing – review & editing. **Tao Cheng:** Software. **Yewen Fan:** Formal analysis, Funding acquisition.

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.

Acknowledgement

This work is supported in part by the National Key Research and Development Program of China with grant number 2018YFB2100501, the National Natural Science Foundation of China with grant numbers 42090012, 41890820 and 42101322, in part by 03 special research and 5G project of Jiangxi Province in China (20212ABC03A09); Zhuhai industry university research cooperation project of China (ZH22017001210098PWC); Key R & D project of Sichuan Science and technology plan (2022YFN0031).

The authors are grateful to USGS for making the Landsat archive freely available, the NASA EOSDIS team for providing the daily 1 km land surface temperature product, and Dr. Peng Gong's team (Tsinghua University) for using their latest land cover product in this study. We also would like to give our thanks to the editors and anonymous reviewers for their comments and suggestions on the earlier version of this manuscript.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jag.2022.102733.

References

Anderson, M.C., Allen, R.G., Morse, A., Kustas, W.P., 2012. Use of Landsat thermal imagery in monitoring evapotranspiration and managing water resources. Remote Sens. Environ. 122, 50–65. https://doi.org/10.1016/j.rse.2011.08.025.

Arslan, N., Sekertekin, A., 2019. Application of Long Short-Term Memory neural network model for the reconstruction of MODIS Land Surface Temperature images. J. Atmos. Sol. Terr. Phys. 194, 105100. https://doi.org/10.1016/j.jastp.2019.105100.

Bastiaanssen, W.G.M., Pelgrum, H., Wang, J., Ma, Y., Moreno, J.F., Roerink, G.J., Van Der Wal, T., 1998. A remote sensing surface energy balance algorithm for land (SEBAL): 1. Formulation. J. Hydrol. 212, 198–212. https://doi.org/10.1016/S0022-1694(98)00254-6.

Bechtel, B., 2012. Robustness of annual cycle parameters to characterize the urban thermal landscapes. IEEE Geosci. Remote Sens. Lett. 9 (5), 876–880. https://doi.org/10.1109/LGRS.2012.2185034.

Bechtel, B., 2011. Multitemporal Landsat data for urban heat island assessment and classification of local climate zones. Joint Urban Remote Sensing Event 2011, 129–132. https://doi.org/10.1109/JURSE.2011.5764736.

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

Deng, C., Wu, C., 2013. Examining the impacts of urban biophysical compositions on surface urban heat island: A spectral unmixing and thermal mixing approach. Remote Sens. Environ. 131, 262–274. https://doi.org/10.1016/j.rse.2012.12.020.

Estoque, R.C., Murayama, Y., 2017. Monitoring surface urban heat island formation in a tropical mountain city using Landsat data (1987–2015). ISPRS J. Photogramm. Remote Sens. 133, 18–29. https://doi.org/10.1016/j.isprsjprs.2017.09.008.

Fu, H., Shao, Z., Fu, P., Zhan, W., Xie, Y., Cheng, T., 2021. Reconciling the inconsistency of annual temperature cycles modelled from Landsat and MODIS LSTs through a percentile approach. Int. J. Remote Sens. 42 (20), 7907–7930. https://doi.org/10.1080/01431161.2021.1966854.

Fu, P., Weng, Q.H., 2015. Temporal Dynamics of Land Surface Temperature From Landsat TIR Time Series Images. Ieee Geosci. Remote Sens. Lett. 12, 2175–2179. https://doi.org/10.1109/Lgrs.2015.2455019.

Gallo, K., Hale, R., Tarpley, D., Yu, Y., 2011. Evaluation of the relationship between air and land surface temperature under clear- and cloudy-sky conditions. J. Appl. Meteorol. Climatol. 50, 767–775. https://doi.org/10.1175/2010JAMC2460.1.

Gevaert, C.M., García-Haro, F.J., 2015. A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion. Remote Sens. Environ. 156, 34–44. https://doi.org/10.1016/j.rse.2014.09.012.

Gong, P., Liu, H., Zhang, M., Li, C., Wang, J., Huang, H., Clinton, N., et al., 2019. Stable Classification with Limited Sample: Transferring a 30-m Resolution Sample Set Collected in 2015 to Mapping 10-m Resolution Global Land Cover in 2017. Sci. Bull. 64, 370–373. https://doi:10.1016/j.scib.2019.03.002.

Han, X.-J., Duan, S.-B., Huang, C., Li, Z.-L., 2019. Cloudy land surface temperature retrieval from three-channel microwave data. Int. J. Remote Sens. 40 (5-6), 1793–1807. https://doi.org/10.1080/01431161.2018.1471552.

Hengl, T., Heuvelink, G.B.M., Perčec Tadić, M., Pebesma, E.J., 2012. Spatio-temporal prediction of daily temperatures using time-series of MODIS LST images. Theor. Appl. Climatol. 107 (1-2), 265–277. https://doi.org/10.1007/s00704-011-0464-2.

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 (5-6), 1828–1843. https://doi.org/10.1080/01431161.2018.1508920.

IPCC, 2018. IPCC Special Report on Global Warming of 1.5°C. https://www.ipcc.ch/.

Ji, S., Xu, W., Yang, M., Yu, K., 2013. 3D Convolutional neural networks for human action recognition. IEEE Trans. Pattern Anal. Mach. Intell. 35 (1), 221–231. https://doi.org/10.1109/TPAMI.2012.59.

Jiménez-Muñoz, J.C., Sobrino, J.A., 2003. A generalized single-channel method for retrieving land surface temperature from remote sensing data. J. Geophys. Res. 108 (D22) https://doi.org/10.1029/2003JD003480.

Jin, M., Dickinson, R.E., 2000. A generalized algorithm for retrieving cloudy sky skin temperature from satellite thermal infrared radiances. J. Geophys. Res. Atmos. 105 (D22), 27037–27047. https://doi.org/10.1029/2000JD900318.

Key, J., 2002. The Cloud and Surface Parameter Retrieval (CASPR) System for Polar AVHRR: Users Guide. Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin 1225, 33–69.

Key, J.R., Wong, A.M., 1999. Estimating the Cloudy Sky Surface Temperature of Sea Ice with Optical Satellite Data. IEEE Int. Geosci. Remote Sens. Symposium 00, 320–322.

Kumar, B., Abhishek, N., Chattopadhyay, R., George, S., Singh, B.B., Samanta, A., Patnaik, B.S. V., Gill, S.S., Nanjundiah, R.S., Singh, M., 2021. Deep Learning Based Forecasting of Indian Summer Monsoon Rainfall. arXiv preprint arXiv:2107.04270.

Kustas, W., Anderson, M., 2009. Advances in thermal infrared remote sensing for land surface modeling. Agric. For. Meteorol. 149 (12), 2071–2081. https://doi.org/10.1016/j.agrformet.2009.05.016.

Kustas, W.P., Norman, J.M., Anderson, M.C., French, A.N., 2003. Estimating subpixel surface temperatures and energy fluxes from the vegetation index-radiometric temperature relationship. Remote Sens. Environ. 85 (4), 429–440. https://doi.org/10.1016/S0034-4257(03)00036-1.

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 https://doi.org/10.1016/j.rse.2021.112707.

Li, X., Zhou, Y., Asrar, G.R., 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. https://doi.org/10.1016/j.rse.2017.12.010.

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. https://doi.org/10.1016/j.isprsjprs.2019.03.013.

Maimaitiyiming, M., Ghulam, A., Tiyip, T., Pla, F., Latorre-Carmona, P., Halik, Ü., Sawut, M., Caetano, M., 2014. Effects of green space spatial pattern on land surface temperature: Implications for sustainable urban planning and climate change adaptation. ISPRS J. Photogramm. Remote Sens. 89, 59–66. https://doi.org/10.1016/j.isprsjprs.2013.12.010.

Meng, Q., Zhang, L., Sun, Z., Meng, F., Wang, L., Sun, Y., 2018. Characterizing spatial and temporal trends of surface urban heat island effect in an urban main built-up area: A 12-year case study in Beijing, China. Remote Sens. Environ. 204, 826–837. https://doi.org/10.1016/j.rse.2017.09.019.

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

Pede, T., Mountrakis, G., 2018. An empirical comparison of interpolation methods for MODIS 8-day land surface temperature composites across the conterminous Unites States. ISPRS J. Photogramm. Remote Sens. 142, 137–150. https://doi.org/10.1016/j.isprsjprs.2018.06.003.

Qi, W., Zhang, X., Wang, N., Zhang, M., Cen, Y.i., 2019. A spectral-spatial Cascaded 3D convolutional neural network with a convolutional long short-term memory network for hyperspectral image classification. Remote Sens. 11 (20), 2363. https://doi.org/10.3390/rs11202363.

Qiao, Z., Tian, G., Xiao, L., 2013. Diurnal and seasonal impacts of urbanization on the urban thermal environment: A case study of Beijing using MODIS data. ISPRS J. Photogramm. Remote Sens. 85, 93–101. https://doi.org/10.1016/j.isprsjprs.2013.08.010.

Qin, Z., Dall'Olmo, G., Karnieli, A., Berliner, P., 2001a. Derivation of split window algorithm and its sensitivity analysis for retrieving land surface temperature from NOAA-advanced very high resolution radiometer data. J. Geophys. Res. 106 (D19), 22655–22670. https://doi.org/10.1029/2000JD900452.

Qin, Z., Karnieli, A., Berliner, P., 2001b. A mono-window algorithm for retrieving land surface temperature from Landsat TM data and its application to the Israel-Egypt border region. Int. J. Remote Sens. 22 (18), 3719–3746. https://doi.org/10.1080/01431160010006971.

Qiu, C., Tong, X., Schmitt, M., Bechtel, B., Zhu, X.X., 2020. Multilevel feature fusion-based CNN for local climate zone classification from sentinel-2 images: Benchmark results on the So2Sat LCZ42 dataset. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 2793–2806. https://doi.org/10.1109/JSTARS.2020.2995711.

Quan, J., Chen, Y., Zhan, W., Wang, J., Voogt, J., Wang, M., 2014. Multi-temporal trajectory of the urban heat island centroid in Beijing, China based on a Gaussian volume model. Remote Sens. Environ. 149, 33–46. https://doi.org/10.1016/j.rse.2014.03.037.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566, 195–204. https://doi.org/10.1038/s41586-019-0912-1.

Shuai, T., Zhang, X., Wang, S., Zhang, L., Shang, K., Chen, X., Wang, J., 2014. A spectral angle distance-weighting reconstruction method for filled pixels of the MODIS land surface temperature product. IEEE Geosci. Remote Sens. Lett. 11, 1514–1518. https://doi.org/10.1109/LGRS.2013.2297735.

Singh, B.B., Singh, M., Singh, D., 2021. An Overview of Climate Change Over South Asia: Observations, Projections, and Recent Advances. In: Singh, R.B., Chatterjee, S., Mishra, M., de Lucena, A.J. (Eds.), Practices in Regional Science and Sustainable Regional Development. Springer Singapore. https://doi.org/10.1007/978-981-16-2221-2.

Sobrino, J.A., Jiménez-Muñoz, J.C., Paolini, L., 2004. Land surface temperature retrieval from LANDSAT TM 5. Remote Sens. Environ. 90 (4), 434–440. https://doi.org/10.1016/j.rse.2004.02.003.

Su, Z., 2002. The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. Hydrol. Earth Syst. Sci. 6, 85–100. https://doi.org/10.5194/hess-6-85-2002.

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. https://doi.org/10.1016/j.cageo.2017.04.007.

Sun, R., Lü, Y., Yang, X., Chen, L., 2019. Understanding the variability of urban heat islands from local background climate and urbanization. J. Cleaner Prod. 208, 743–752. https://doi.org/10.1016/j.jclepro.2018.10.178.

Tan, J., Che, T., Wang, J., Liang, J.i., Zhang, Y., Ren, Z., 2021. Reconstruction of the daily modis land surface temperature product using the two-step improved similar pixels method. Remote Sensing 13 (9), 1671. https://doi.org/10.3390/rs13091671.

Tayyebi, A., Jenerette, G.D., 2018. Assessing diel urban climate dynamics using a land surface temperature harmonization model. Int. J. Remote Sens. 39 (9), 3010–3028. https://doi.org/10.1080/01431161.2018.1437292.

Van Nguyen, O.n., Kawamura, K., Trong, D.P., Gong, Z., Suwandana, E., 2015. Temporal change and its spatial variety on land surface temperature and land use changes in the Red River Delta, Vietnam, using MODIS time-series imagery. Environ. Monit. Assess. 187 (7) https://doi.org/10.1007/s10661-015-4691-3.

Wan, Z., 2014. 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.

Wang, L., Koike, T., Yang, K., Yeh, P.-F., 2009. Assessment of a distributed biosphere hydrological model against streamflow and MODIS land surface temperature in the upper Tone River Basin. J. Hydrol. 377 (1-2), 21–34. https://doi.org/10.1016/j.jhydrol.2009.08.005.

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

Weng, Q., Fu, P., 2014. Modeling annual parameters of clear-sky land surface temperature variations and evaluating the impact of cloud cover using time series of Landsat TIR data. Remote Sens. Environ. 140, 267–278. https://doi.org/10.1016/j.rse.2013.09.002.

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 Sensing 11, 300. https://doi.org/10.3390/rs11030300.

Xiao, Y., Zhao, W., Ma, M., He, K., 2021. Gap-free lst generation for modis/terra lst product using a random forest-based reconstruction method. Remote Sensing 13 (14), 2828. https://doi.org/10.3390/rs13142828.

Xu, Y., Shen, Y., 2013. 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.

Xu, Z., Guan, K., Casler, N., Peng, B., Wang, S., 2018. A 3D convolutional neural network method for land cover classification using LiDAR and multi-temporal Landsat imagery. ISPRS J. Photogramm. Remote Sens. 144, 423–434. https://doi.org/10.1016/j.isprsjprs.2018.08.005.

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 (3), 1026–1040. https://doi.org/10.1109/JSTARS.2019.2896455.

Yu, F., Hao, H., Li, Q., 2021. An Ensemble 3D Convolutional Neural Network for Spatiotemporal Soil Temperature Forecasting. Sustainability 13, 9174. https://doi.org/10.3390/su13169174.

Yu, W., Wu, T., Nan, Z., Zhao, L., Wang, Z., 2014. A novel interpolation method for MODIS land surface temperature data on the Tibetan Plateau. Proc. SPIE – Int. Soc. Opt. Eng. 9260 https://doi.org/10.1117/12.2068892.

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. https://doi.org/10.1016/j.isprsjprs.2018.04.005.

Zeng, C., Shen, H., Zhong, M., Zhang, L., Wu, P., 2015. Reconstructing MODIS LST based on multitemporal classification and robust regression. IEEE Geosci. Remote Sens. Lett. 12, 512–516. https://doi.org/10.1109/LGRS.2014.2348651.

Zhang, X., Li, P., 2018. A temperature and vegetation adjusted NTL urban index for urban area mapping and analysis. ISPRS J. Photogramm. Remote Sens. 135, 93–111. https://doi.org/10.1016/j.isprsjprs.2017.11.016.

Zou, Z., Zhan, W., Jiang, L., 2017. Enhanced modeling of remotely sensed Annual Land surface temperature cycle. Int. Arch. Photogrammetry, Remote Sens. Spatial Inform. Sci. – ISPRS Archives 42, 1067–1074. https://doi.org/10.5194/isprs-archives-XLII-2-W7-1067-2017.

Zou, Z., Zhan, W., Liu, Z., Bechtel, B., Gao, L., Hong, F., Huang, F., Lai, J., 2018. Enhanced modeling of annual temperature cycles with temporally discrete remotely sensed thermal observations. Remote Sensing 10, 1–12. https://doi.org/10.3390/rs10040650.