



**FOCUS ARTICLE**

# Predicting land surface temperature with geographically weighed regression and deep learning

Hongfei Jia | Dehe Yang | Weiping Deng | Qing Wei | Wenliang Jiang

National Institute of Natural Hazards,  
Ministry of Emergency Management of  
China, Beijing, China

**Correspondence**

Dehe Yang and Weiping Deng, National Institute of Natural Hazards, Ministry of Emergency Management of China, No.1 Anningzhuang Road, Haidian District, Beijing 100085, China.  
Email: ydhmm@163.com (D. Y.) and Email: weaponed@126.com (W. D.).

**Funding information**

National Natural Science Foundation of China, Grant/Award Number: 41772219; Doctoral fund project of Henan polytechnic university, Grant/Award Number: B2015-22

## Abstract

For prediction of urban remote sensing surface temperature, cloud, cloud shadow and snow contamination lead to the failure of surface temperature inversion and vegetation-related index calculation. A time series prediction framework of urban surface temperature under cloud interference is proposed in this paper. This is helpful to solve the problem of the impact of data loss on surface temperature prediction. Spatial and temporal variation trends of surface temperature and vegetation index are analyzed using Landsat 7/8 remote sensing data of 2010 to 2019 from Beijing. The geographically weighed regression (GWR) method is used to realize the simulation of surface temperature based on the current date. The deep learning prediction network based on convolution and long short-term memory (LSTM) networks was constructed to predict the spatial distribution of surface temperature on the next observation date. The time series analysis shows that the NDBI is less than -0.2, which indicates that there may be cloud contamination. The land surface temperature (LST) modeling results show that the precision of estimation using GWR method on impervious surface and water bodies is superior compared to the vegetation area. For LST prediction using deep learning methods, the result of the prediction on surface temperature space distribution was relatively good. The purpose of this study is to make up for the missing data affected by cloud, snow, and other interference factors, and to be applied to the prediction of the spatial and temporal distributions of LST.

This article is categorized under:

Technologies > Machine Learning

## KEY WORDS

geographically weighed regression, land surface temperature, long short-term memory, spatial correlation, time series

## 1 | INTRODUCTION

Land surface temperature (LST) is one of the most important parameters in the physical processes of surface energy and water balance. It provides information on the temporal and spatial variations of the surface equilibrium state and affects the variety of fields including urban ecosystem and human activities (Karnieli et al., 2010; Kustas & Anderson, 2009). However, it is essential for LST to gain better understanding of the climate dynamics and forecasting

extreme events. The study is related to temperature and its connectivity with urban environment. It is categorized into two major areas: (a) the atmospheric temperature, calculated from weather station networks (Chow & Roth, 2006; Fast, Torcolini, & Redman, 2005) and (b) the surface temperature, estimated from thermal infrared remote sensing techniques (Imhoff, Zhang, Wolfe, & Bounoua, 2010; Peng et al., 2012). Ground measurements practically obtain discrete LST observation. Satellite remote sensing offers the only possibility for measuring LST with complete spatial averaged on pixel basis rather than point values (Li et al., 2013). Due to the easy access and complete spatial coverage of the earth, remotely sensed surface temperature has been a popular area of research in recent times (Clinton & Gong, 2013; Zhang, Imhoff, Wolfe, & Bounoua, 2010).

Many different approaches have been proposed to derive LST from Thermal Infrared (TIR) data based on radiative transfer equation (RTE). Some methods have been employed to deal with the emissivity and atmospheric effects (Jiménez-Muñoz & Sobrino, 2003; Tonooka, 2001). To derive LST, we use the single-channel method that is a simple inversion of the RTE. It should be noted that the LSE and the atmospheric profiles should be known for calculation of RTE based LST. LST can be calculated from Landsat 5, 7, and 8 data (Chen, Wang, & Li, 2002; Weng, 2001), which have the advantage of recording long-term surface temperature changes, and also a relatively high spatial resolution, and enhance the analysis of surface spatial temperature changes in details. However, there are some limitations to the data: for example, the long revisited period and the interference of cloud, snow and other factors make the high-quality data be observed by satellite at a specific period. In terms of reconstruction of the observed data and spatial distribution of surface temperature, there are more studies carried out on the reconstruction using Moderate Resolution Imaging Spectroradiometer data with lower spatial resolution (Markus, AndreoVerónica, & Markus, 2017; Neteler, 2010; NourEldeen et al., 2020), and fewer studies on the reconstruction using Landsat data with higher spatial resolution. In addition, other studies propose that data fusion is affected by ground object changes. For Landsat time series, a time series model was proposed to generate new band image data (Zhu, Woodcock, Holden, & Yang, 2015), which required a large number of clearly observed data to establish a time series function whose generalization ability still needs to be verified.

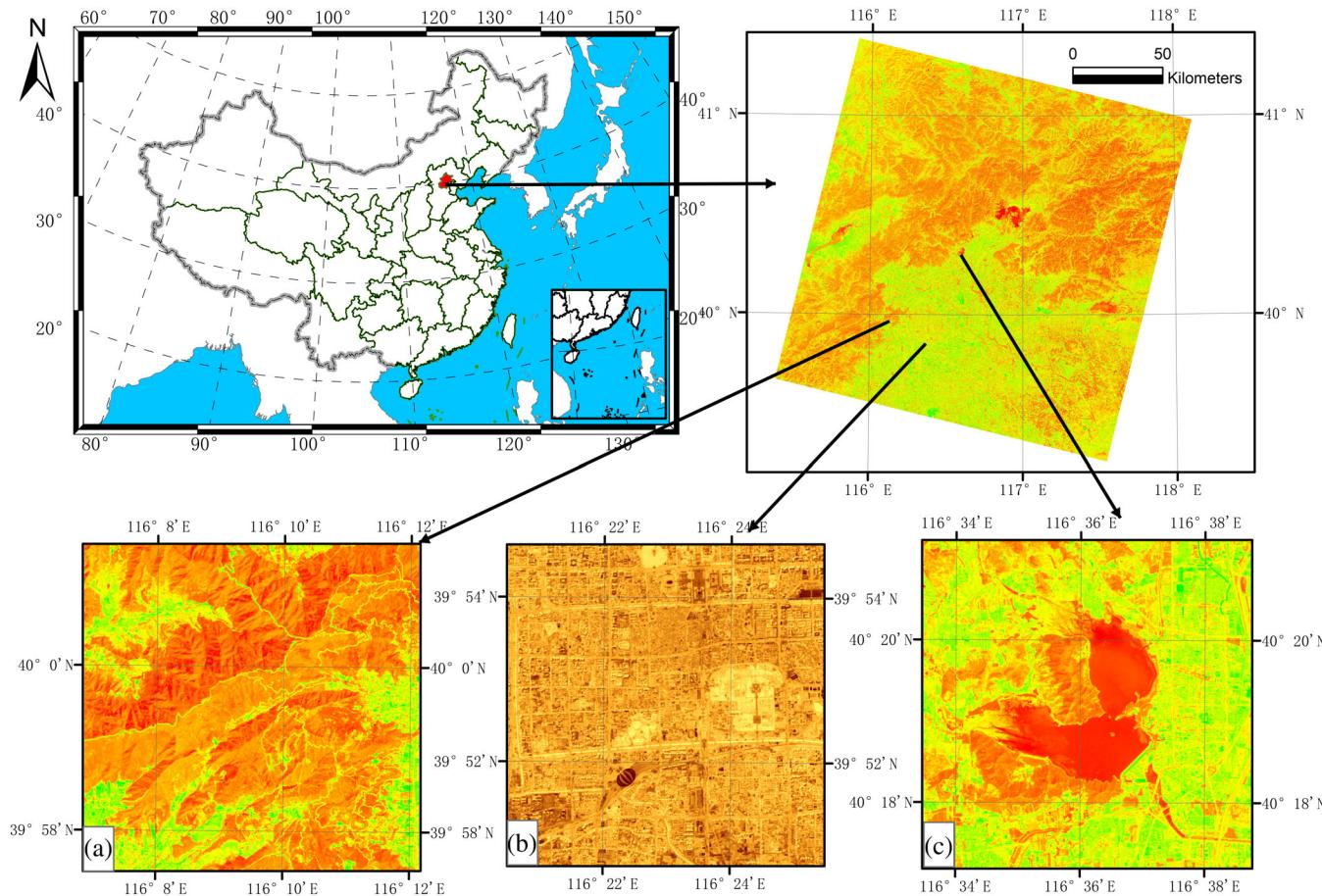
The LST in urban areas is closely related to the changes of human-induced land use/land cover (LULC) (Buyantuyev & Wu, 2010; Chao et al., 2017). This shows that the surface radiation temperature response can be expressed as a function of vegetation cover and soil water content on the surface (Owen, Carlson, & Gillies, 1998). These outcomes inspired many studies on the relationship between vegetation abundance and LST (Weng, 2001; Weng, Lu, & Jacquelyn, 2003). The ordinary least squares (OLS) regression can be used (Huang, Yuan, & Lu, 2017; Zhou, Qian, Li, Li, & Han, 2013) to analyze the relationship between LST and other LULC features (Estoque, Murayama, & Myint, 2017; Myint, Wentz, Brazel, & Quattrochi, 2013). However, the spatial dependence and spatial heterogeneity are not considered in the OLS regression and it was assumed that the process was constant throughout the entire study. To solve this problem, geographically weighed regression (GWR) was used in many current studies to figure out the possible relationships between LST and LULC composition, and also used to analyze the spatial variations and temporal changes in different parts of the globe (Wang, Fan, Zhao, & Myint, 2020; Zhao, Jensen, Weng, & Weaver, 2018). The other studies employed GWR focusing on randomly sampled points (Huang et al., 2017; Wang et al., 2020) but not region of interest (ROI) in pixel level. GWR can be applied to fill the gap in time series, to make the long-term changes of temperature be obtained in high spatial resolution for one area. Its prediction can be divided into long trend prediction, seasonal prediction, cyclic prediction, irregular variation prediction, and comprehensive prediction. Some studies focused on comprehensive prediction to estimate LST value in the length of one step (Huang, Li, & Lu, 2008; Mathew, Sreekumar, Khandelwal, Kaul, & Kumar, 2016). Moreover, the deseasonalization (irregular variation) prediction has been studied using long short-term memory (LSTM) aimed at identifying extreme events in LST time series (Xiao et al., 2019). LSTM, as a kind of recurrent neural network, is capable of predicting time series data but is incapable of modeling the LST sequence of long-term dependencies because of the problem of vanishing and exploding gradient. The recurrent structure and gating mechanisms of LSTM enhance its long-term memory, which makes it an advanced method to solve time series-related problems (Chao, Pu, Yin, Han, & Chen, 2018). Although it is being applied to the short and mid-term daily sea surface temperature (SST) (Xiao et al., 2019; Zhang, Wang, Dong, Zhong, & Sun, 2017), the single-point data are only predicted rather than the special spatial range in remote sensing image. The prediction pattern of 10 days for SST is inappropriate due to the error accumulation caused by the former predicted values. The main problem lies in the 1-day prediction error that has exceeded 30%, which makes the error of the latter days larger and larger. Therefore, the pattern of 1-day prediction should be an optimal choice based on the literature reviewed. According to the method

of one-step and comprehensive prediction, we try to use the conv2DLSTM base on the former LST image time series to predict the LST image on the next measurement date for any region of interest.

The purpose of this research is to examine the changes in LULC features and LST in Beijing since the 2010s and to show the relationship between them in any ROI. Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), and Normalized Difference Built-up Index (NDBI) indices were used to derive LULC information from Landsat7/8 images in different times and the surface temperature inversion in thermal infrared band was obtained. The specific objectives include the following: (a) to examine extreme climate phenomenon from three remote sensing indices; (b) to analyze the temporal and spatial difference of LST among different seasons in Beijing; (c) to estimate LST from three remote sensing indices in the ROI using GWR; (d) to predict LST from former LST time series in the ROI using conv2DLSTM; (e) to predict the spatial distribution of surface temperature in the next date using deep learning method basing on the surface temperature from GWR simulation and single-channel inversion. The paper is organized as follows: Section 2 describes the study area, the Landsat data, and the preprocessing strategies used in this study. Section 3 presents the methodology and the calculation of remote sensing indices, LST, GWR, and the proposed conv2DLSTM method for LST prediction. Section 4 provides the experimental results. Thorough discussion is given in Section 5. Finally, Section 6 concludes the paper.

## 2 | STUDY AREA AND MATERIALS

The study area that we selected in this paper is Beijing, which is located in the east of China and occupies 16,410.54 km<sup>2</sup>. The LULCs include forest, shrubland, water, and urban built-up (Wang et al., 2020). Landsat 7 (2010.01–2013.05) and 8 (2013.05–2019.11) images from one scene (path/ row: 123/32) in the World Reference



**FIGURE 1** Study area in Beijing. (a) Jiufeng Forest Park, (b) the Imperial Palace, (c) Huairou Reservoir

System-2 were downloaded from the United States Geological Survey website (Figure 1). Landsat 8 thermal infrared sensor bands 10 and Landsat 7 ETM+ band 6\_VCID\_1 are resampled to 30-m resolution.

In this research, three representative sites in Beijing were selected for case study, as shown in Figure 1. The three sites are Jiufeng forest park ( $40^{\circ}0'34.82''N$ ,  $116^{\circ}9'32.57''E$ ), the Imperial Palace ( $39^{\circ}55'54.55''N$ ,  $116^{\circ}24'34.02''E$ ) and Huairou Reservoir ( $40^{\circ}18'47.46''N$ ,  $116^{\circ}35'59.46''E$ ). A  $300 \times 300$  pixel box around each site is applied for the GWR and conv2DLSTM. A  $15 \times 15$  pixel box around each site is applied for the analysis of remote sensing indices time series.

The strip noise is filtered by Landsat Gap Fill using Hayes algorithm for Landsat 7. Bands 3 (Green), 4 (Red), 5 (Near Infrared), and 6 (Swir1) of all images for Landsat 8 are calibrated to top of atmosphere reflectance for the calculation of NDVI, NDWI, and NDBI, and Bands 2 (Green), 3 (Red), 4 (NIR), and 5 (Middle Infrared) of all images for Landsat 7 are also done as Landsat 8. If NDBI value is lower than  $-0.2$  in special ROI, the corresponding date is removed from the LST time series that applied for prediction.

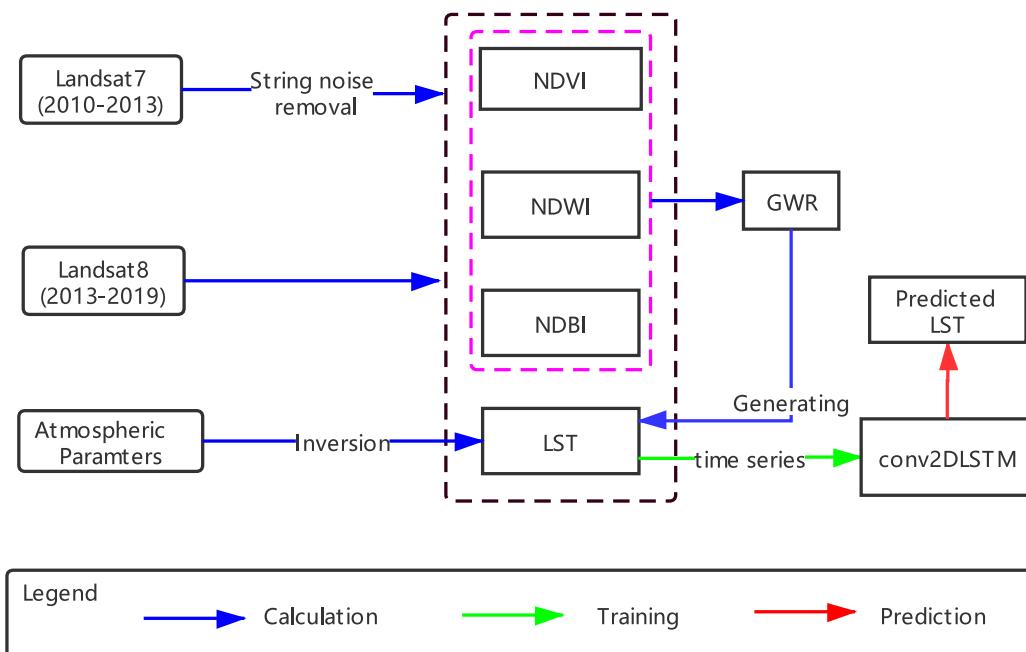
### 3 | METHODS

#### 3.1 | LST and biophysical parameters computation

Figure 2 shows the spatial-temporal analysis for the Landsat data. The total amount of thermal radiation absorbed by the Landsat 7 and Landsat 8 satellites is subtract from the estimated atmospheric influence on the surface thermal radiation. The resulting surface thermal radiation intensity is then converted into the corresponding surface temperature.

$$L(T_\lambda) = [eL(T_s) + (1-e)L \downarrow]\tau + L \uparrow, \quad (1)$$

where  $L(T_\lambda)$  is the total amount of thermal radiation observed by the satellite sensor,  $L \uparrow$  is the upward radiance of the atmosphere, the downward radiation of the atmosphere denoted by  $L \downarrow$ , the Earth's true radiance reaching the satellite sensor is  $L(T_s)$ ,  $e$  denotes the target emissivity and  $1 - e$  represents the target albedo,  $\tau$  stands for atmospheric transmittance in the thermal infrared band. The atmospheric parameters including  $\tau$ ,  $L \uparrow$ , and  $L \downarrow$  is provided by the following web site: <http://atmcntr.gsfc.nasa.gov/>. Moreover,  $e$  can be estimated using the NDVI-based threshold method (Qin, Karnieli, & Berliner, 2001). The Calvin  $-273.15$  is converted to Celsius for LST.



**FIGURE 2** Flow diagram of spatial-temporal analysis for the Landsat data

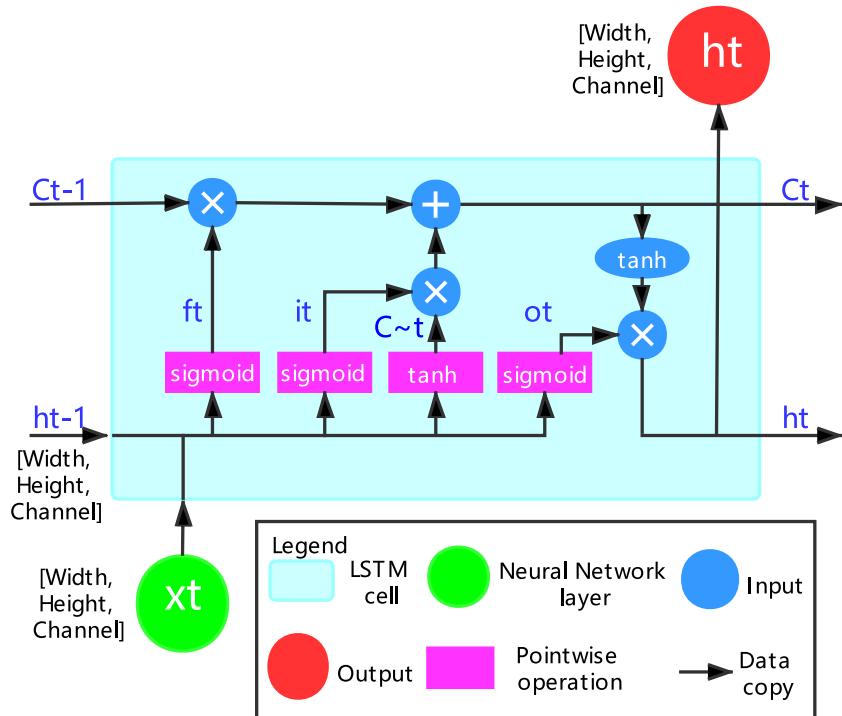
Biophysical parameters mainly include NDVI, NDBI, and NDWI. The NDVI is calculated using Top of Atmosphere (TOA) reflectance from near infrared and red bands of each image for Landsat 7/8. The TOA reflectance of green and near-infrared bands is applied to the NDWI calculation for Landsat 7/8. The NDBI is calculated based on the TOA reflectance of near infrared band and short-wave infrared band for Landsat 8. The TOA reflectance of near infrared and middle infrared bands is used in calculating NDBI for Landsat 7. It should be noted that there are discrepancies that are not appropriate for time series analysis, if these indices are calculated using DN value for Landsat 7/8.

### 3.2 | Multi-indices time series modeling with LSTM

The key block of LSTM is a cell, which is linked back and forth to constitute a chain-like network structure. There are three different types of gates employed in LSTM to control a cell state. They are forget gates ( $f_t$ ), input gates ( $i_t$ ), and output gates ( $o_t$ ) as shown in Figure 3. A gate, which consists of a fully connected neural network layer activated by a sigmoid function and a pointwise multiplication operation (Olah, 2015), is a rule to regulate how much information can be removed or memorized. The structure of altering and transmitting information can be expressed as follows:

$$\begin{aligned} f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \\ i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \\ \tilde{C}_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \\ C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t, \\ o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \\ h_t &= o_t * \tanh(C_t), \end{aligned} \quad (2)$$

where  $f_t$ ,  $i_t$ , and  $o_t$  are the three outputs of the sigmoid functions  $\sigma$ , having values between 0 and 1.  $f_t$  determines the information needed to be discarded.  $i_t$  ascertains the information to be updated.  $C_{t-1}$  denotes an old cell state.  $\tilde{C}_t$  creates new candidate values using a tanh function.  $C_t$  is a new cell state.  $o_t$  decides the information we will output.  $h_t$  represents the output information.



**FIGURE 3** Structure of long short-term memory (LSTM)

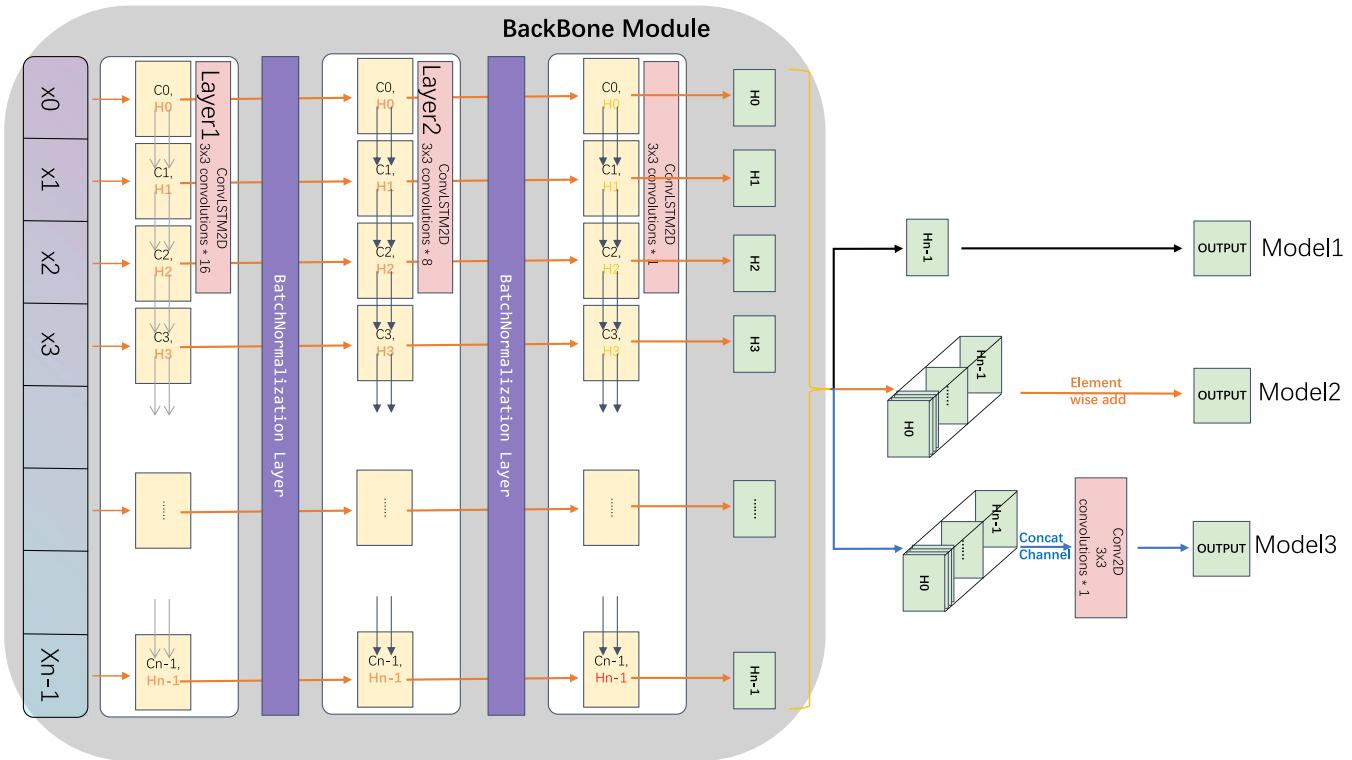
$$\text{loss function : loss} = \frac{1}{N} \int_{i=1}^N (\text{predict}_i - \text{label}_i)^2 \\ = \frac{1}{h * w} \int_{h=1}^H \int_{w=1}^W (\text{predict Pixel}_{hw} - \text{label Pixel}_{hw})^2. \quad (3)$$

As shown in Equation (3), the loss function was constructed to update the LSTM parameters through Adam optimizer. The predict Pixel is the output information  $h_t$  in Function 7, and  $N$  is the number of pixels of image matrix.  $h$  and  $w$  denote the height and width of the image, respectively.

Figure 4 shows a built LSTM and a two-dimensional convolution. It includes a five-layer neural network, comprising of one input layer, one output layer, and three conv2DLSTM layers. Experiments were designed to assess this neural network and the results show that the best prediction performance was achieved. The input of this neural network could be expressed as 4D tensors, which included Step, Height, Width, and Channels. The Step represents the size of the time window used in the prediction of the LST. We set it to 50, this indicates that the previous 50 measurements LST sequence was used to predict the 51st LST. Height and Width denote the size of a ROI image for LST. Channels are set to 1 because the LST was the only predictor. After several experiments, the number of cells and convolution in each conv2DLSTM was 50. The size of convolution kernel was  $3 \times 3$ , the number of which was set to 16. The first 85% of samples in the data set were used as training samples, 20% of which were split out for validation, and the remaining 15% samples were set as the held-out testing samples. The batch size is 4 and learning rate is set to  $2e^{-4}$  for training. Cross-entropy is applied to the standard of accuracy evaluation and the number of epochs is set to 800.

### 3.3 | Multi-indices spatial correlation modeling with GWR

The GWR model is an advanced technique of conventional standard regression methods like OLS. It supports the estimation of local parameters for the analysis of spatial nonstationarity instead of global parameters (Fotheringham,



**FIGURE 4** Structure of conv2DLSTM. The Model 1, Model 2, and Model 3 are defined as S1, S2, and S3, respectively. If not specified, the model used in this paper is the S1 model

Charlton, & Brunsdon, 2001). GWR enables the spatial visualization of local relationships and a better understanding of local variations by adding a parameter of the geographical location to the traditional regression. GWR model is given by

$$y_i = \beta_0(\mu_i, \nu_i) + \sum_{k=1}^p \beta_k(\mu_i, \nu_i)x_{ik} + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (4)$$

where  $y_i$  is the dependent variable,  $x_{ik}$  is the  $k$ th independent variable at position  $i$ , and  $\varepsilon_i$  is a random error term at location  $i$ .  $(\mu_i, \nu_i)$  are the coordinates of the observed location  $i$  and  $\beta_k(\mu_i, \nu_i)$  is the coefficient associated with the  $k$ th variable at location  $i$ .

GWR estimates parameters according to the degree to which all observed values are spatially close to a particular point  $i$ . The closer the observation points to one point  $i$ , the greater impacts it will have on the local parameter estimates. This indicates that a point nearer to the point  $i$  has more weight than the one far from it in the local parameter estimation.

The significance and residual spatial autocorrelation of GWR coefficients were tested using Akaike Information Criterion (AIC), which is a relative measure of performance comparison of models. AIC along with Monte Carlo, Residual squares, and Sigma squares diagnostics (Leung, Mei, & Zhang, 2003), are the most commonly used statistical methods to determine the coefficient significance. The smaller AIC is, the superior a model is. Corrected AIC ( $AIC_C$ ) denotes AIC which corrects the finite sample sizes.

$$AIC_C = 2n\ln(\hat{\sigma}) + n\ln(2\pi) + n\left(\frac{n + \text{tr}(S)}{n - 2 - \text{tr}(S)}\right), \quad (5)$$

where  $n$  is the number of the sample size,  $\hat{\sigma}$  is the estimate of the standard deviation of the residuals, and  $\text{tr}(S)$  denotes the trace of the hat matrix. The model accuracy is evaluated using the  $AIC_C$ . Lower values of  $AIC_C$  indicate a better model fit. Hence, the smaller the values of this statistics, the closer the fit of the GWR model to the observed data.

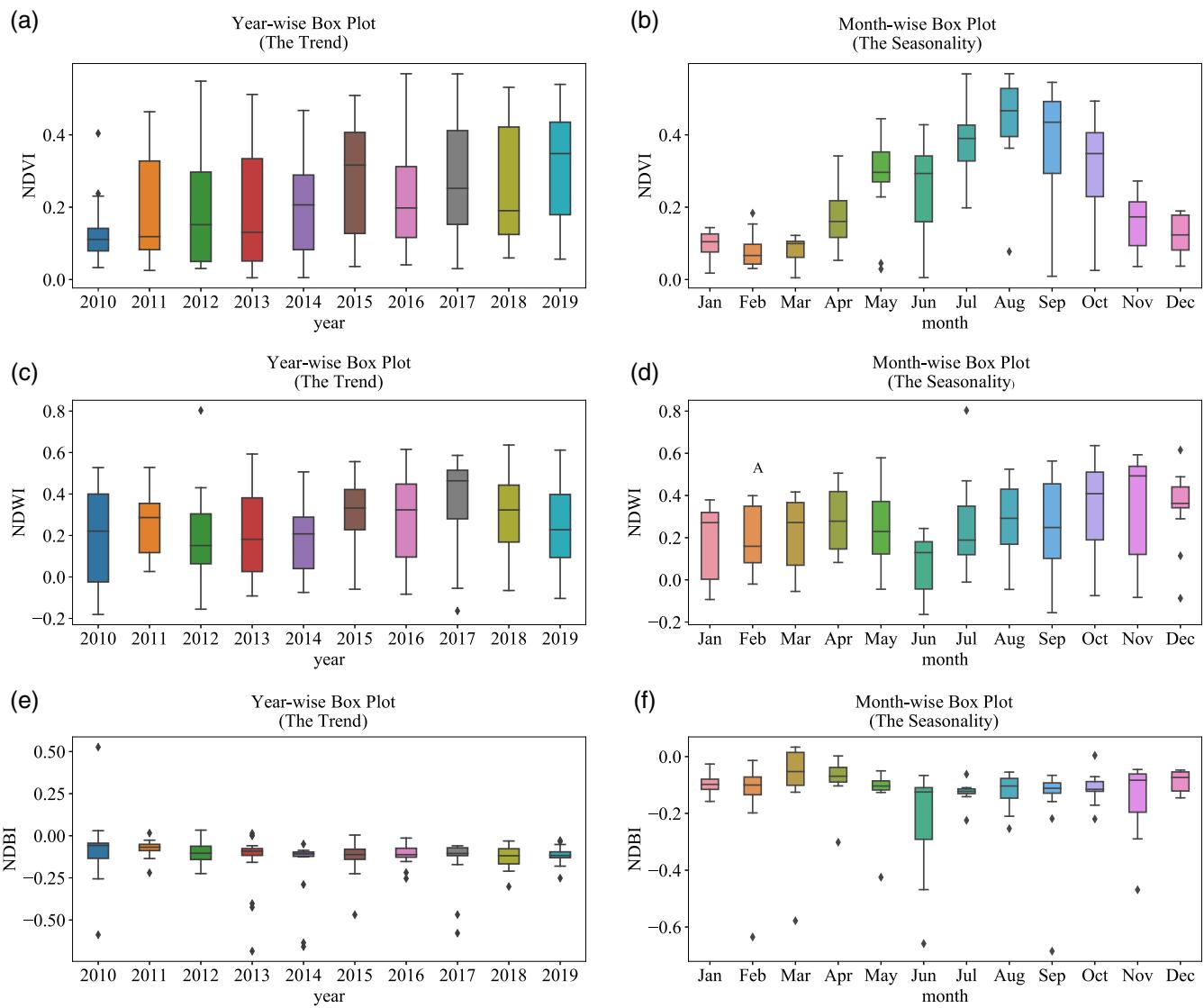
### 3.4 | Performance evaluation

To evaluate the accuracy of the simulated surface temperature, we compared it to the inverted surface temperature. The difference was calculated quantitatively using relative error. To evaluate the accuracy of deep learning on the sample area, the similarity and differences of the predicted surface temperature were calculated quantitatively. The differential indicator function is denoted by

$$\text{diff} = \left\| \frac{\mathbf{X} - \mathbf{Y}}{N} \right\|_F = \sqrt{\frac{\sum_{i,j} (X_{ij} - Y_{ij})^2}{N}}, \quad (6)$$

where  $\mathbf{X}$  and  $\mathbf{Y}$  stands for the actual and predicted image matrix and  $N$  is the number of pixels in the image matrix.

We use the histograms between the images to evaluate the similarity between the surface temperature predicted by deep learning and the actual inverted surface temperature. Firstly, the histogram distributions of the two kind images were calculated. Secondly, the histogram was divided into 64 zones and further divided into 4 consecutive levels. Summing up the 4 values of each zone, the data vectors with size of 64 were obtained for two images. Lastly, cosine similarity,  $\cos(\theta) = \frac{\mathbf{X} \cdot \mathbf{Y}}{\|\mathbf{X}\| \|\mathbf{Y}\|}$ , was used to calculate the similarity between two vectors, which can be used to measure the similarity between images. Two methods were used in this paper to measure the similarity of images: one is to directly calculate the cosine similarity of two images; the other is to calculate the cosine similarity of two images after subtracting the difference from each pixel of the image. The similarity between two images is evaluated by degree of angles.



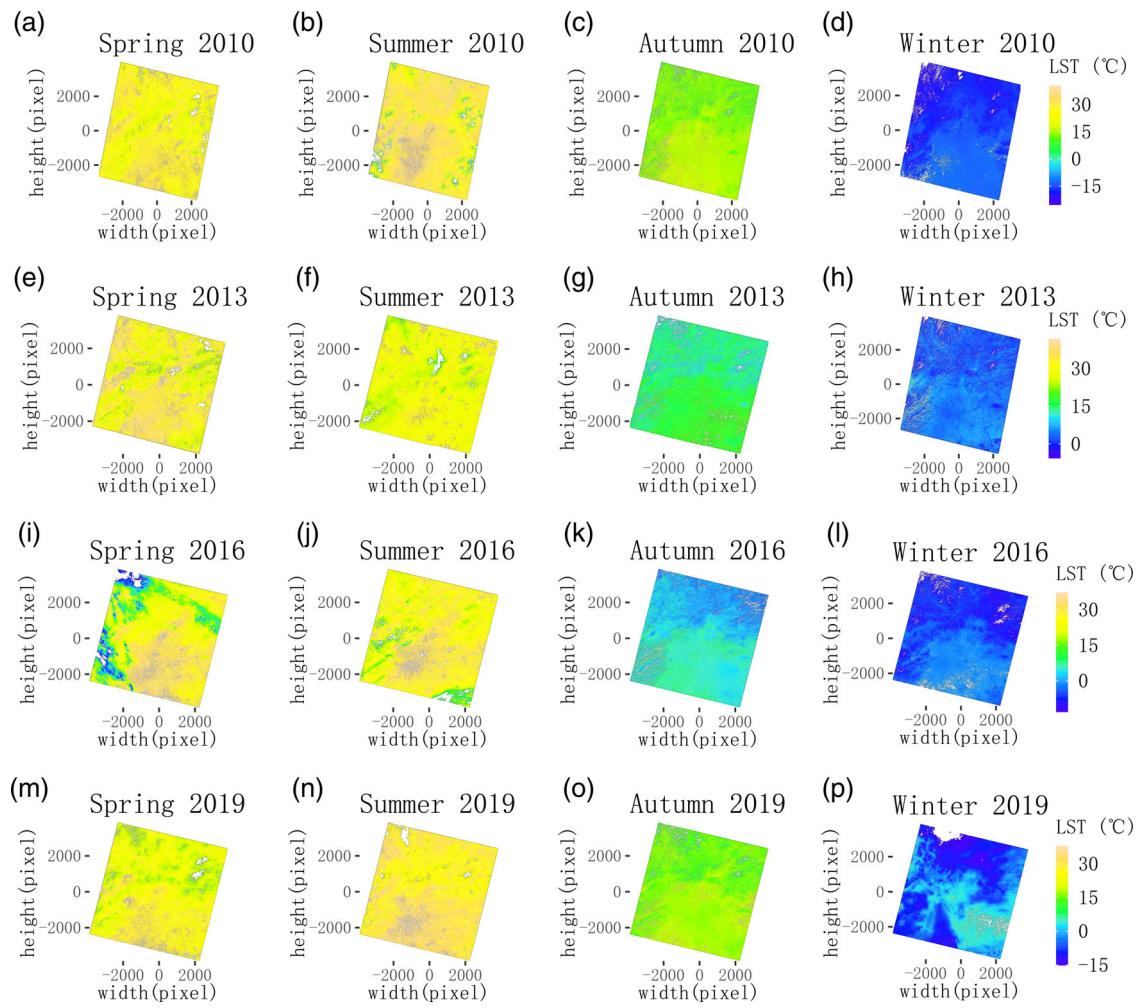
**FIGURE 5** Remote sensing indices extracted from Landsat observed at Jiufeng forest park between 2010 and 2019. (a) Year-wise box plot for Normalized Difference Vegetation Index at Jiufeng forest park. (b) Month-wise box plot for Normalized Difference Vegetation Index at Jiufeng forest park; (c) Year-wise box plot for Normalized Difference Water Index at Huairou reservoir. (d) Month-wise box plot for Normalized Difference Water Index at Huairou reservoir. (e) Year-wise box plot for Normalized Difference Built-up Index at Imperial Palace. (f) Month-wise box plot for Normalized Difference Built-up Index at Imperial Palace

## 4 | RESULTS AND ANALYSIS

### 4.1 | Temporal analysis of the multi-indices

Figure 5 shows the temporal change of remote sensing indices, including NDVI, NDWI, and NDBI, ranging from 2010 to 2019. Figure 5a,c,e groups the data at year intervals, which demonstrates the distribution of values in a given year from 2010 to 2019. Figure 5b,d,f shows the distribution of values in different months among 1 year. These pictures show the distribution by year and by month.

Figure 5a shows that there is an increasing slope for the trend of NDVI time series at Jiufeng forest park. In Figure 5b, seasonality is observed; there is a steep rise of NDVI from May to June in 2019. This implies that vegetation coverage increases when compared with the same date in other years, validated using a precipitation from the National Oceanic and Atmospheric Administration (NOAA). In addition, the NDVI values in both July and August are significantly higher in the 1-month boxplot, which is attributed to the annual precipitation concentrated months. We



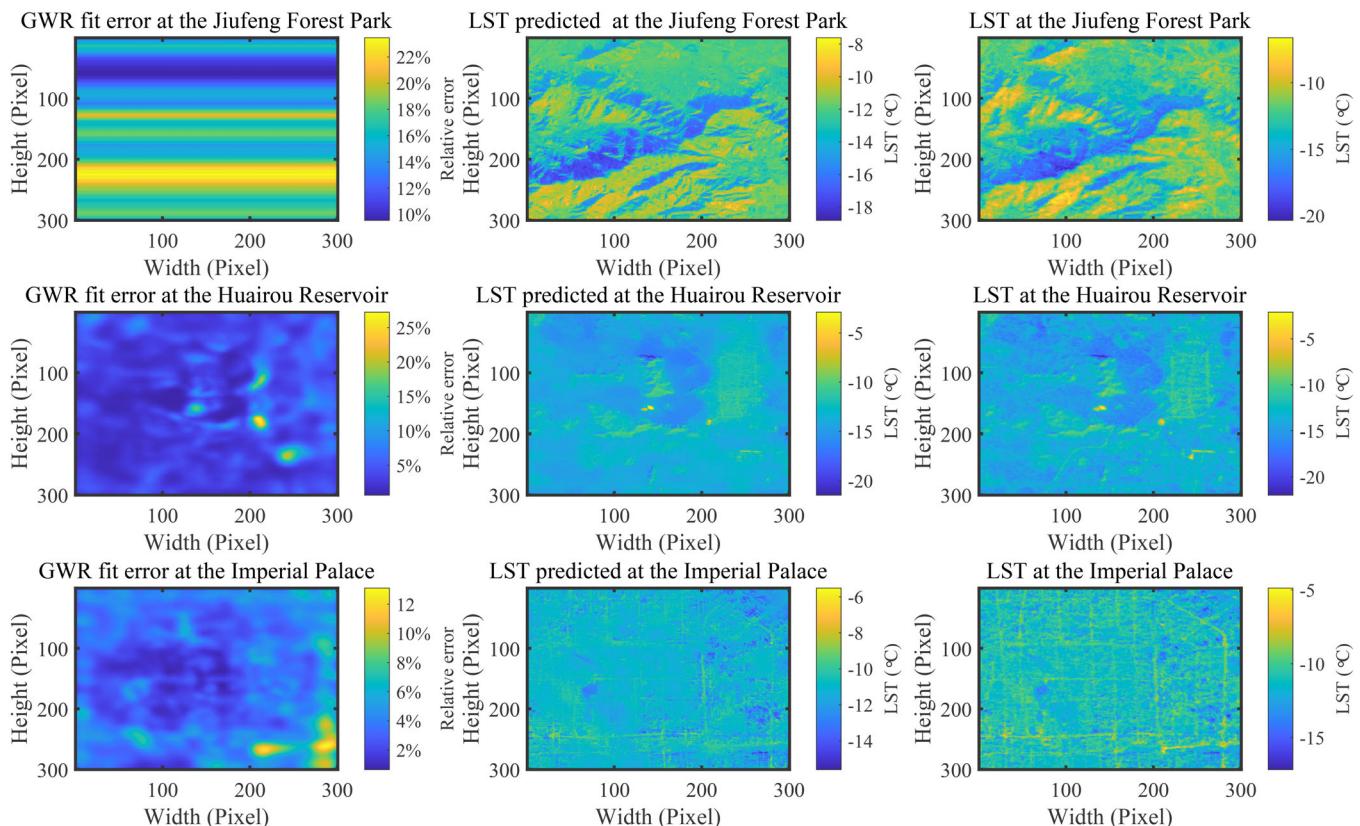
**FIGURE 6** Temporal and spatial difference of land surface temperature (LST) between different seasons in Beijing

observed that Figure 5d shows an outlier value on July 20, 2019, which is not as a result of the cloud because the NDWI values affected by cloud has a lower value that have been validated using data contaminated with cloud observed on another date. The boxplots at Figure 5c,d shows that NDWI mean values of all years for each month is higher than 0.1 because NIR is absorbed strongly by water (McFeeters, 1996). There is also a significant seasonality for the NDBI time series at the Imperial Palace as shown in Figure 5(f). This may be affected indirectly by the presence of other covers whose reflectance is seasonal. Although built-up areas and barren land experience a drastic increment in their reflectance, most of the NDBI values are lower than 0 owing to other types of pixels. The results are different compared with the result from Landsat5 TM imagery (Zha, Gao, & Ni, 2003). If the NDBI is lower than -0.2, the image could be contaminated by cloud through visual validation in Figure 5e,f. The mean value of NDBI for every year keeps a stable level that is about -0.1, as shown in the bottom left picture of Figure 5. In addition, the NDBI values in both July and August are significantly lower in the one-month boxplot.

Figure 6 shows several LSTs derived from the Landsat 7/8 images in Beijing from 2010 to 2019. It shows a clear gradient between the urban areas and other areas, which is due to the fact that the urban surface materials have higher radiant temperatures (Fonseka et al., 2019). Observing all the seasons, it is found that the maximum LST is distributed over urban areas. The surface temperature over north mountain is found to be lower than urban areas in all the seasons for those years. Figure 6 shows that there is an obvious seasonal characteristic for the surface temperature on spring and summer, whose LST is higher than that on autumn and winter. In 2016 and 2019, summer temperatures was significantly higher, and in 2019, the winter temperature was higher than other years, which is consistent with announced temperature of summer and winter from the NOAA and the China Meteorological Administration. Table 1 shows the maximum and minimum LST for months selected in 2010, 2013, 2016, and 2019. In general, the maximum LST for spring and summer clearly shows higher amplitude than other seasons. It is observed that the temperature is above 33 in May for those years in Beijing. The satellite

Year	Type	Winter	Spring	Summer	Autumn
2010	Min	-25.38	14.06	-9.34	5.32
	Max	-10.16	35.48	40.48	20.43
2013	Min	-5.69	21.72	19.48	6.53
	Max	11.43	41.82	35.71	21.59
2016	Min	-13.11	-10.96	8.92	-4.56
	Max	1.14	36.95	37.22	10.17
2019	Min	-15.16	15.03	22.51	8.61
	Max	5.80	33.77	37.96	20.63

**TABLE 1** The maximum and minimum land surface temperature (LST) for months selected in 2010, 2013, 2016, and 2019



**FIGURE 7** Land surface temperature (LST) predicted using geographically weighted regression (GWR) in a pixel level

and ground-based measured LST was compared. The comparison for Landsat 8 estimated LST with ground data is found to be within an error of about 3 °C (compared for the same solar time of satellite pass).

## 4.2 | Spatial patterns analysis of LST using GWR

To elaborate the effectiveness of GWR, we pick date at random to compare the real LST and predicted LST. Figure 7 shows the LST values of three locations in a pixel level for one date estimated from NDVI, NDWI, and NDBI using GWR, which we observed that there is significant spatial heterogeneity between LST and other indices. In Figure 7, the pattern of fit errors estimated by GWR is exhibited as a horizontal stripe at Jiufeng Forest Park. The maximum of fit errors has reached about 24%. Figure 7 shows that the maximum of fit errors reaches 26%, which is close to the Huairou Reservoir. The above result indicates that there are local multicollinearity problems. The fit errors for the Imperial Palace are less than 13% compared to the above locations due to the relative consistency of reflectivity for the built-up area.

**TABLE 2** Parameters obtained from GWR and OLS in January 04, 2010

	OLS				GWR	
	Coef	SE	p Value	VIF	Mean coef	SD
Normalized Difference Vegetation Index	-24,449.8	471.7	≈0	1	7.847	21.59
Normalized Difference Water Index	-15,506.8	312.5	≈0	1	1.105	19.18
Normalized Difference Built-up Index	-8,818.4	48.8	≈0	1	2.981	10.7
AIC	702,464				47,144	
Adjusted $R^2$	0.130				0.949	
Bandwidth	-				42	

Abbreviations: AIC, Akaike Information Criterion; GWR, geographically weighed regression; OLS, ordinary least squares; VIF, variance inflation factor.

Since the prediction error is lower, the spatial pattern of LST is closer to the actual spatial pattern when predicted using GWR methods. According to the spatial distribution of fitting error in Figure 7, GWR's estimation effect on impervious water surface and water body is better than that of vegetation area.

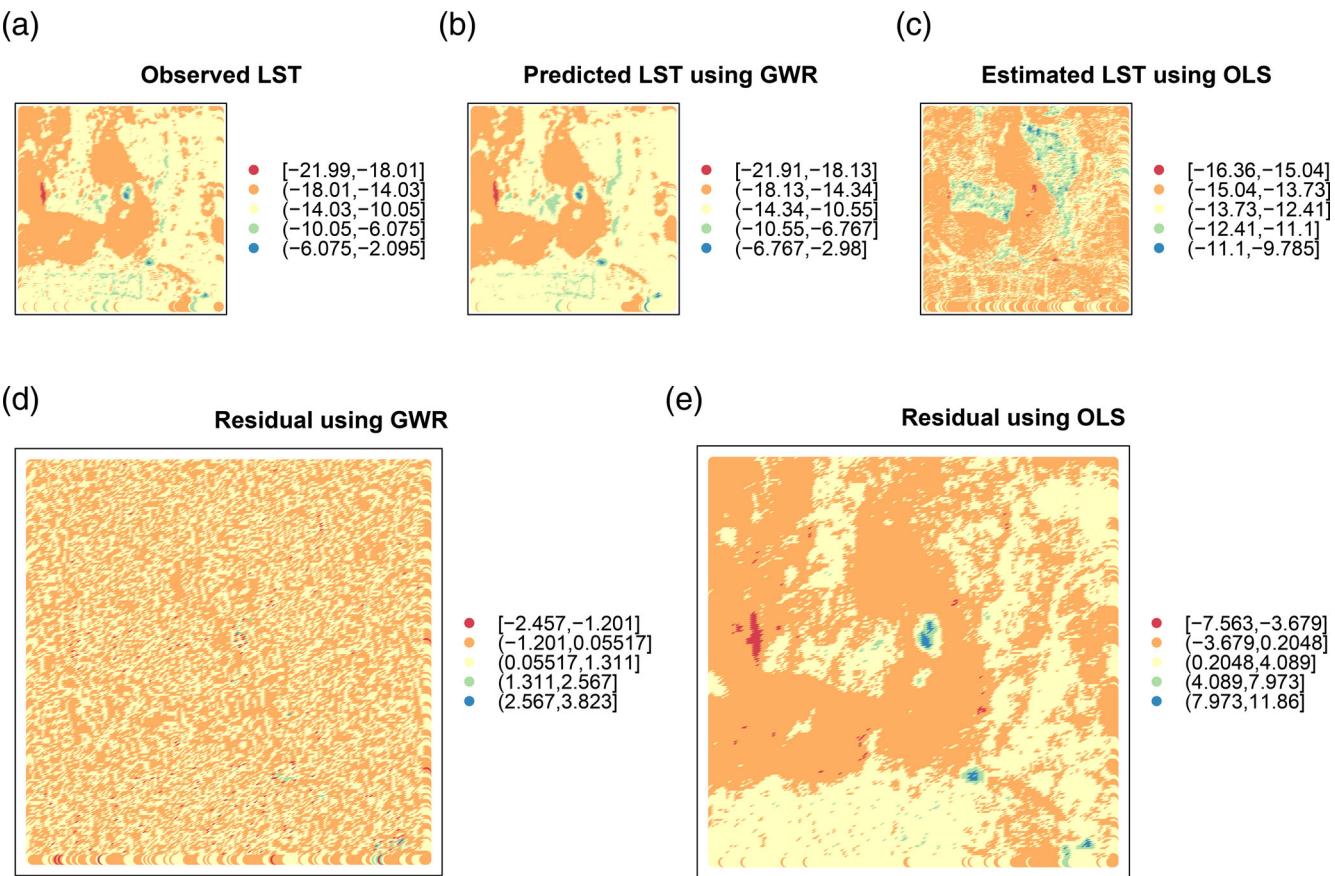
Table 2 lists the parameters obtained from OLS and GWR for one ROI image. Using the variance inflation factor (VIF), there was no significant multicollinearity issue in the models we employed. According to the linear regression model, a negative correlation between LST and biophysical parameters was obtained, which was statistically significant for this ROI image. The OLS model set up the global estimation for the entire ROI image, while the GWR model made predictions to the specific local region. Table 2 lists the mean (coef) and SD obtained from all the local predictions. With an adjusted  $R^2 = .130$ , the OLS model failed to provide a good fit for the ROI image. Hence, the results of the global OLS were inconclusive, which calls for the need of a local estimation. In the view of model diagnostics, GWR improved model performance significantly compared to OLS by raising the mean adjusted  $R^2$  from 0.130 to 0.949 in pixel level. The average AIC for OLS was 702,464, and GWR had a better performance with an AIC of 47,144. Local processes are identified by the GWR through bandwidth = 42 (pixel) in models, which stand for very localized spatial relationships between LST and three indices.

Figure 8 shows the predicted spatial pattern of LST by both GWR and OLS models. The GWR model demonstrates a significant improvement over the OLS model with the same predictor. The spatial pattern of LST predicted by GWR is closer to the actual spatial pattern as is indicated by the lower prediction error. When LST values are low, the simulation residual of the OLS model is higher than that of the GWR model.

#### 4.3 | LST prediction using conv2DLSTM

LSTM is capable of predicting the time series, which is validated by many research studies (Xiao et al., 2019; Zhang et al., 2017). We have tested the model for the NDVI, NDWI, and NDBI at several locations and considered that the above time series should be decomposed by additive or multiplicative methods. Otherwise, the predicted value would extremely deviate from the real value in phase or amplitude, which would not be provided in limited space. Previous studies of relationship between LST and other remote sensing indices concentrate on scatter plots (Chen, Zhao, Li, & Yin, 2006; Manjula, Estoque, & Yuji, 2017), which is applied to their spatial correlations. Figure 9 shows the relationships of time series between LST and other remote sensing indices in the Imperial Palace. NDWI of the surface moisture is negatively correlated with NDVI of vegetation by the correlation value of -0.98. There are linear relationships between NDVI, NDWI, and LST. NDBI for the built-up area and other three indices have very weak correlations, as shown in Figure 9. This indicates that the predictive ability of NDBI with regard to the temporal pattern of LST in the study area is inferior to NDVI and NDWI.

As shown in Figure 10, the relative error of conv2DLSTM observed indicates that the prediction accuracy of each pixel varies ranging from about 45% to zero for two dates. It can be observed that the range predicted is larger than that of real value. Although the error is large enough for some pixels, most of the errors are below 20%. Obviously, the spatial pattern of LST predicted is similar to the real LST image according to the histogram. The river on the left part inside the image is clearly presented in the LST image predicted, which may be due to the advantage of convolution on the feature extraction from the image. The reason for the discrepancy of the range of relative error for two images used

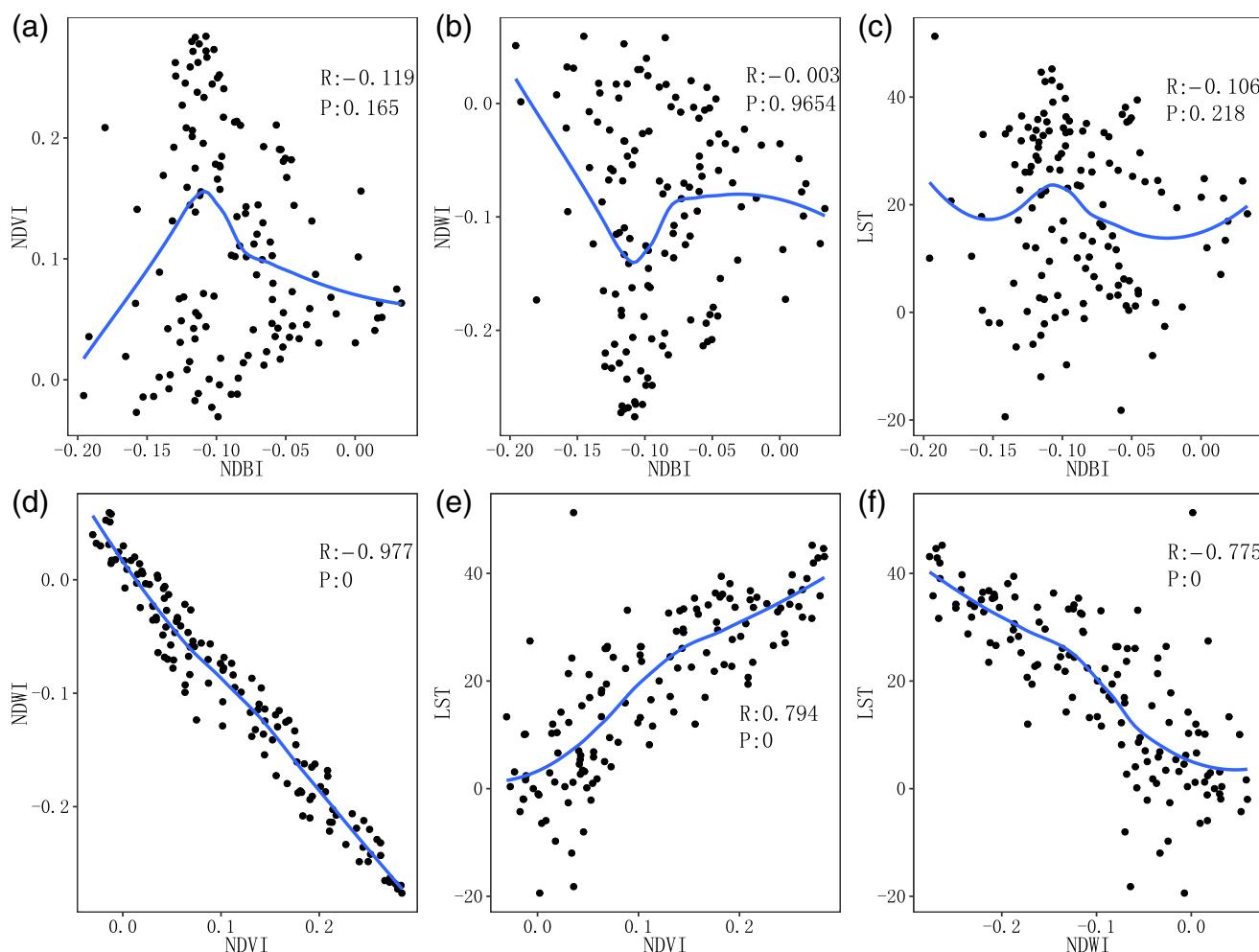


**FIGURE 8** Land surface temperature (LST) analysis results obtained using geographically weighed regression (GWR) and ordinary least squares (OLS) at Imperial Palace on January 04, 2010

may be insufficient training samples resulting in overfitting. In addition, it is found that the prediction results of the water body were closer to the inversion value, which was possibly because the thermal inertia of the water was higher than that of other surface features making the temperature change of the water area lag behind. As a result, the single-step prediction was more accurate.

To verify the reliability of the prediction using deep learning method constructed by convolution and LSTM, we use the time sequence with the window length of 50 and step length of 1 to the conv2DLSTM model after learning completed, the similarities and differences of 14 randomly results were calculated and the inversion result achieved. Figure 11a shows the differences of the 14 predicted results and inversion results, which ranges from 1 to 4 °C. After removing the differences, as shown in Figure 11c, the similarity of the overall distribution is relatively high, that is, the angle between the image gray histogram vectors is smaller (less than 8°). The results showed that this method can be effectively applied to the single-step prediction of the surface temperature image. The prediction results also showed that this method can be used to distinguish the different ground objects, which mainly because of the advantage of identifying shape features using convolution, and the stability of different ground objects in the sample area in the prediction time range. Due to the difference in the range of the lowest and highest surface temperatures of different images, the predicted images without removing the differences have a low similarity (with a large angle degree, about 20°–50°).

To obtain the parameters of the best model, we perform cross check to avoid the influence of data distribution on our method. The total amount of samples is of 94. The training and predicting samples account for 85 and 15%, respectively. Different data sets were selected randomly in 100 times to get the best parameters. Figure 12 shows that the model is the best when angle degree is of <8. As shown in Figure 13, Val\_loss and Tran\_loss represent the L2 loss during valuation and training process, respectively. Val\_ce and Tran\_ce stand for the cross entropy curve during those processes, respectively. With the increase of the iteration number, the conv2DLSTM model begins to converge gradually after 100 generations. It shows that large perturbation exists before 100 generations due to the characteristics of



**FIGURE 9** Relations between NDBI, NDVI, NDWI, and land surface temperature (LST) for all years at the Imperial Palace

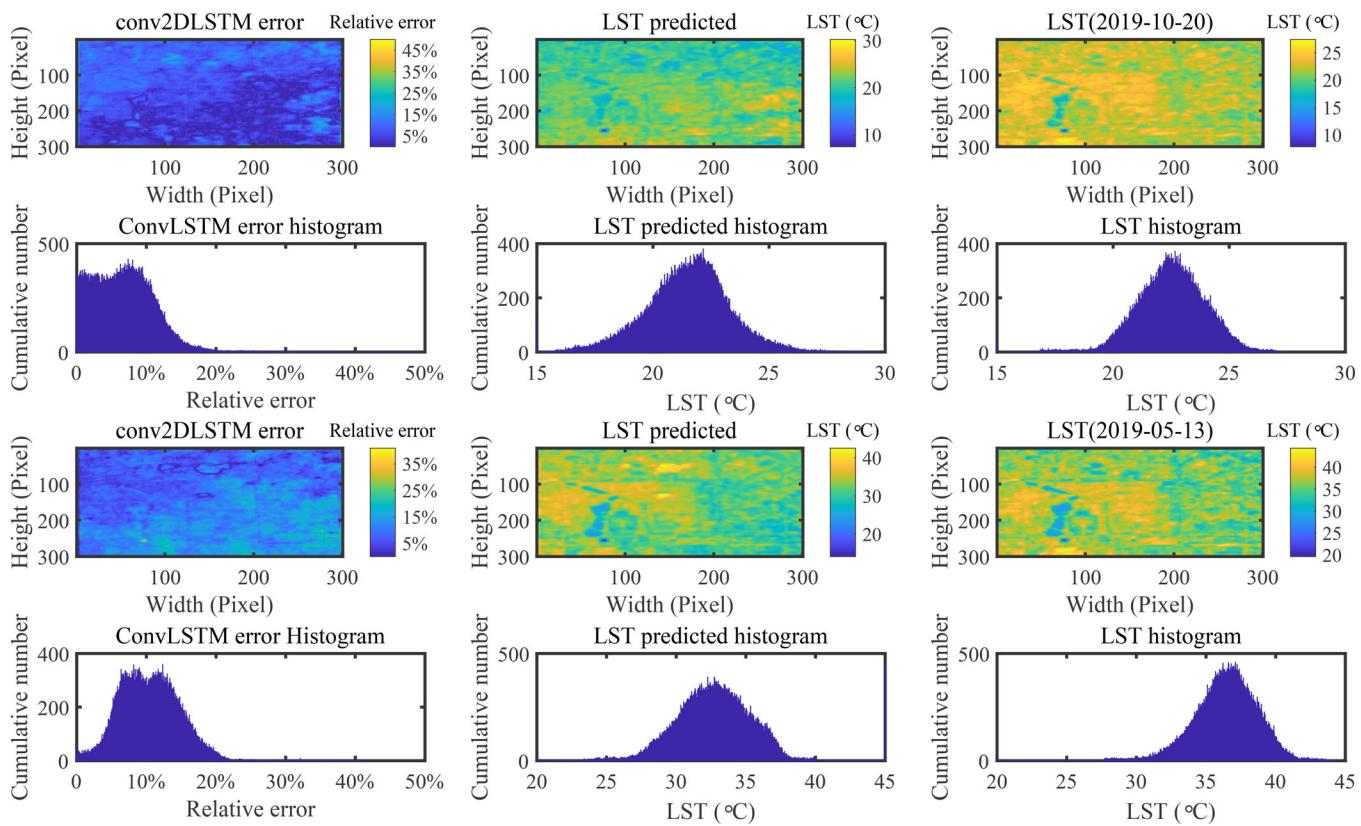
optimizer. The lower the cross entropy is, the more accurate the prediction model is. With the increase of iteration number, the length of training and validation samples for each batch has an influence on the convergence trend of cross entropy and determines the level of stability.

To evaluate the effect of network structures and parameters on the accuracy, we designed three network models to predict the LST through changing learning rate (Lrate), the size of samples (Stps) and kernel size (Krnls). Table 3 shows the accuracy of the data set selected. We checked the extent to which the factors affected the model resulting accuracy, for evaluating whether the S1 model performance is better than those of the two models of both S2 and S3. The accuracy of S1 model (as seen in Figure 4) with Lrate =  $2e^{-4}$ , Stps = 50, and Krnls = 3, is higher than other models. The output of S2 model (as seen in Figure 4) is the sum of all outputs in the Layer 3, which is the only difference with the output of the S1. Similarly, the output of S3 model (as seen in Figure 4) can be obtained from the convolution of all outputs in the Layer 3.

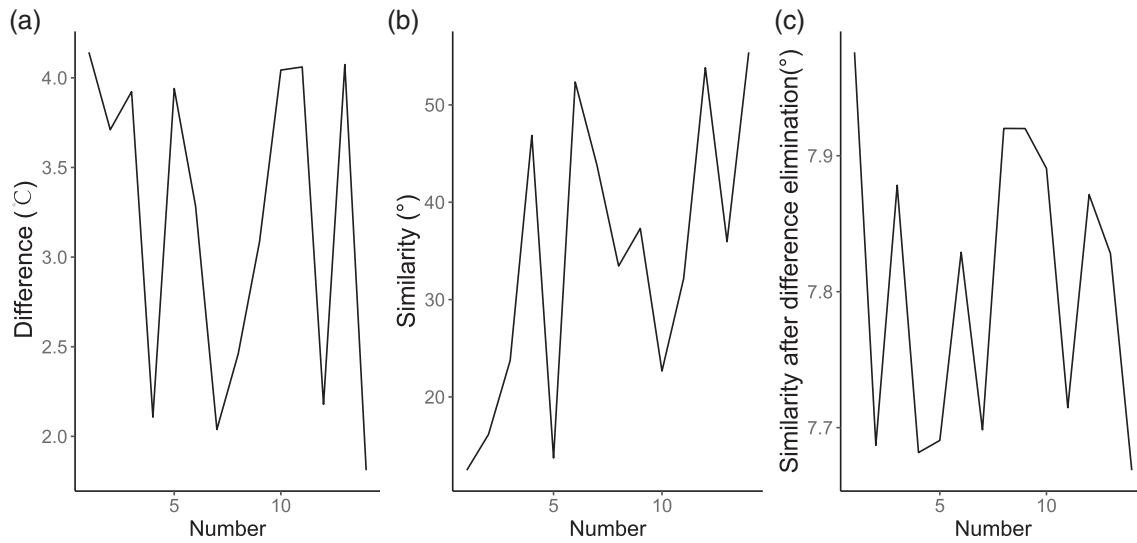
## 5 | DISCUSSION

### 5.1 | Analysis of times model

Visual analysis of the time series of remote sensing indices (NDVI, NDWI, and NDBI) from 2010 to 2019, some extreme events were observed directly, such as the strong variation of vegetation cover and water body. The experiments show that there are severe phase lag and amplitude changes when using LSTM directly to predict time series of different remote sensing indices. As a result, it is necessary to conduct the prediction using the trend, seasonal and residual term

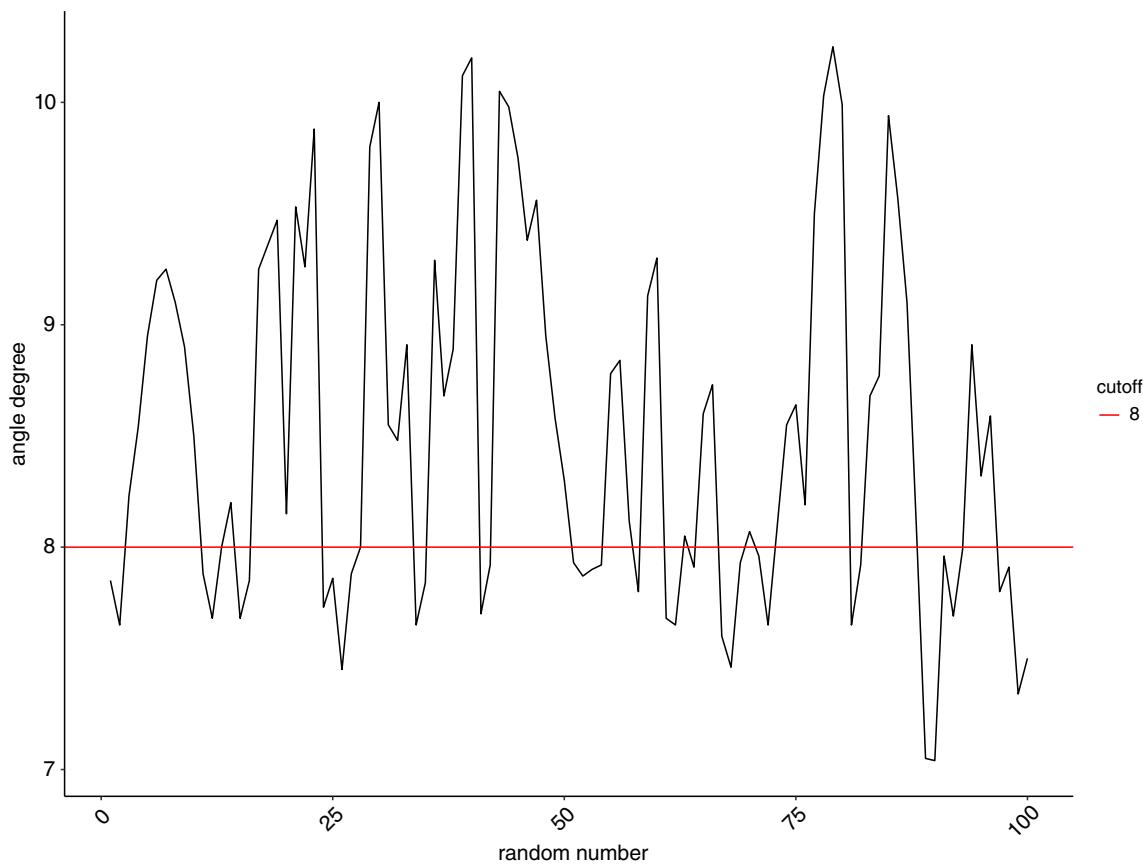


**FIGURE 10** Land surface temperature (LST) prediction at the Imperial Palace using conv2DLSTM



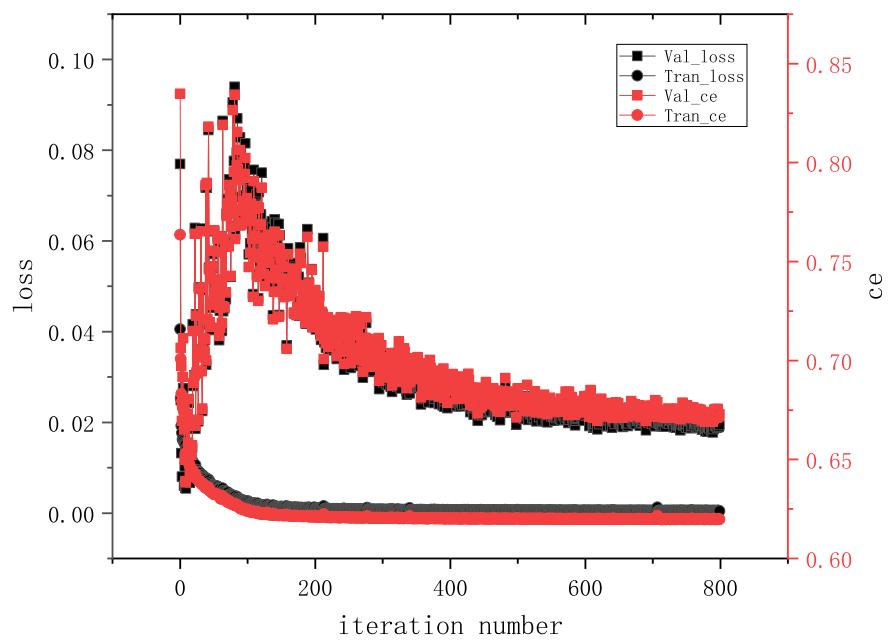
**FIGURE 11** Similarity and difference evaluation of prediction

after decomposing the addition or multiplication time series model. The periodic temporal analysis on remote sensing indices of different surface objects, provided not only a basis on choosing the window size of deep learning, but also the criterion for the accuracy difference in time series prediction. This resulted from the periodic difference between different surface objects. Whether there is a similar periodic difference mode in different water body areas remains to be further verified in more water body areas. The time series visualization analysis method cannot only discover the trend and abnormal changes in biological parameters but also provide process guidance for the subsequent simulation of surface temperature based on multiple indices using GWR. Since images observed from Landsat satellite over Beijing for each



**FIGURE 12** Cross check through random method

**FIGURE 13** Training and validation process



year can easily be contaminated by thin could and cloud shadow to more than 20%, the recognition of contamination before the LST calculation is an important application on LST time series prediction. Deep learning technology can be developed for automated cloud detection, but it requires a certain level of knowledge of clouds. Considering the importance of cloud identification in analyzing LST time series, a more intelligent and efficient method will be investigated in the future work.

Lrate	Stps	Krnls	Accuracy		
			S1	S2	S3
$1e^{-4}$	25	3	9.59	9.74	9.78
		5	9.65	9.73	9.79
		7	9.70	9.73	9.75
	40	3	8.57	8.62	8.54
		5	8.55	8.65	8.60
		7	8.63	8.64	8.62
	50	3	8.34	8.44	8.32
		5	8.32	8.43	8.29
		7	8.49	8.52	8.51
$2e^{-4}$	25	3	9.21	9.21	9.19
		5	9.34	9.21	9.21
		7	9.32	9.25	9.23
	40	3	8.23	8.20	8.15
		5	8.34	8.45	8.20
		7	8.29	8.30	8.39
	50	3	<b>7.81</b>	7.96	<b>7.81</b>
		5	<b>7.88</b>	7.98	<b>7.89</b>
		7	7.97	7.97	7.96
$5e^{-4}$	25	3	9.53	9.68	9.52
		5	9.69	9.75	9.70
		7	9.88	9.95	9.89
	40	3	8.55	8.70	8.60
		5	8.69	8.71	8.64
		7	8.63	8.74	8.71
	50	3	8.43	8.48	8.42
		5	8.41	8.50	8.42
		7	8.45	8.65	8.51

**TABLE 3** The accuracy of land surface temperature prediction using different structures and parameters

## 5.2 | Analysis of spatial model

Comparing the LST of the same period in different years, it is observed that LST of urban areas is significantly higher than that of other areas. Since the heat island effect is not our focus on in this paper, the ground object is not clarified to study the spatial difference of LST. Based on the inverted LST images of the Beijing area, two prediction studies were carried out in this paper. On the one hand, the relationship between LST and NDVI, NDWI, and NDBI was studied using GWR at pixel level. The results further prove the effectiveness of applying GWR model at a pixel level to analyze the correlation between LST and remote sensing indices, while the relationship between LST and the combination of different explanatory variables is not studied. In addition, the selected spatial region in this paper is small because of the high time consumption of GWR analysis at pixel level, but in reality, a relatively larger area is required for spatial nonuniformity study, which calls for the enhancement of time efficiency. The result that the surface temperature prediction on vegetation area is inferior to that on water body and impervious surface using GWR requires further analysis. Compared to OLS, GWR has a strong ability to depict spatial distribution patterns and details, which leads to a better effect on surface temperature simulation and provides high-quality sample data for subsequent deep learning to predict surface temperature. Secondly, we proposed to predict LST temporal images using conv2DLSTM method and single-channel method. When applying multi-channel method to predict LST image in certain areas, the loss function was transformed from single-channel loss to multi-channel loss. Conducting the LST prediction at our selected spatial area

almost maximize the GPU load, to predict LST time series at larger area requires to reduce the training parameters by compressing deep neural networks. In addition, the method of singular value decomposition and non-negative matrix decomposition can be used to measure the similarity between surface temperature image prediction and surface temperature image inversion, besides using histogram vectors.

### 5.3 | Limitations

We found that the coverage of clouds is different between the images of two adjacent time intervals of the same urban area, therefore the splicing process is required to analyze the data. This paper proposes a conv2DLSTM model to predict LST time series images in a single-channel method. However, the accuracy of predication needs to be improved, especially at the aspects of optimizing the network model and increasing the number of samples. The correlation among different time series of remote sensing indices was analyzed, while the spatial correlation analysis which should be concerned when studying urban heat island effects, was not carried out in this paper. Hence, the spatial relationship between LULC type and LST needs to be analyzed in different seasons of the year. In particular, it remains to be further explored that whether there is a fixed pattern for the gradient difference of spatial distribution between water body, vegetation and impervious water surface in the same region over the same period of each year. We will pay attention to research on trend prediction of LST values of special areas. In Landsat 7 LST inversion, some low temperature values (below  $-50^{\circ}\text{C}$ ) were obtained. Two percent linear stretch method was applied to process these values which resulted to the removal of abnormal low temperature values. Further study should be carried out to locate the cause of low temperature values from Landsat 7 LST inversion. However, through the similarity and difference of accuracy verification, it showed that the precision of this method is higher when the temperature on water body is predicted than other surface features. To realize high precision prediction on the temperature of other surface objects, more data should be used to optimize the method. In terms of improving the generalization of our model, data augmentation methods including generating image contaminated with noise and making selection randomly that are widely applied on natural images, are not suited to train the LST prediction model using LSTM. The best choice may be to add the LST measurements obtained from different satellite to fill the time interval and increase the length of observation. Using Landsat 5 and Sentinel in the future, will likely provide better results on surface temperature simulation and prediction.

## 6 | CONCLUSIONS

In this paper, the spatial-temporal variation characteristics of NDVI, NDWI, NDBI, and LST in Beijing were analyzed. LST is predicted using GWR method at pixel level to compensate for the image of failed inversion or numerical error of surface temperature. Also a conv2DLSTM deep learning prediction model for LST time series images was constructed.

1. By a series of temporal curve visualization methods, the abnormal values in NDVI and NDWI time series were found in the area of interest and the effectiveness of NDBI threshold on identifying air contamination was verified by manual comparison. NDBI ( $< -0.2$ ) in construction area indicates the possibility of cloud contamination.
2. We demonstrated that GWR is a superior model to analyze the relationships between NDVI, NDWI, NDBI, and LST patterns at the pixel level. We found that GWR's estimation accuracy of water and impervious surface is better than that of vegetation regions.
3. The predicted image time series indicates that conv2DLSTM proposed has a good predictive power for the local LST. The convolution is a high-performance image feature extractor, it plays an important role in the prediction of an image with abundant texture structure.

With the help of geographical spatial analysis and deep learning technology, the LST in a local spatial range will be predicted effectively in a certain period, which will help decision makers to formulate policies in time to deal with the negative effects of temperature changes. Moreover, it is meaningful and necessary to enhance the robustness of the predicting performance with respect to network structures and hyperparameters in the design of conv2DLSTM. A kind of parameter tuning strategy, such as grid search, should be paid attention to in further studies.

## ACKNOWLEDGMENT

This study was supported by National Natural Science Foundation of China (41772219) and Doctoral fund project of Henan Polytechnic University (B2015-22).

## CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

## AUTHOR CONTRIBUTIONS

**Hongfei Jia:** Writing-original draft. **Dehe Yang:** Methodology; supervision. **Weiping Deng:** Data curation; software. **Qing Wei:** Investigation; methodology. **Wenliang Jiang:** Formal analysis; funding acquisition.

## ORCID

Dehe Yang  <https://orcid.org/0000-0002-4467-3798>

## RELATED WIREs ARTICLES

- [Process-Guided Deep Learning Predictions of Lake Water Temperature.](#)  
[A Rainfall-Runoff Model With LSTM-Based Sequence-to-Sequence Learning.](#)

## REFERENCES

- Buyantuyev, A., & Wu, J. (2010). Urban heat islands and landscape heterogeneity: Linking spatiotemporal variations in surface temperatures to land-cover and socioeconomic patterns. *Landscape Ecology*, 25(1), 17–33. <https://doi.org/10.1007/s10980-009-9402-4>
- Chao, F., Soe, M., Shai, K., Ariane, M., Baojuan, Z., Huei-Ping, H., Anthony, B., & Dan, B. (2017). Understanding the impact of urbanization on surface urban heat islands—A longitudinal analysis of the oasis effect in subtropical desert cities. *Remote Sensing*, 9(7), 672. <https://doi.org/10.3390/rs9070672>
- Chao, Z., Pu, F., Yin, Y., Han, B., & Chen, X. (2018). Research on real-time local rainfall prediction based on MEMS sensors. *Journal of Sensors*, 1–9. <https://doi.org/10.1155/2018/6184713>.
- Chen, X.-L., Zhao, H.-M., Li, P.-X., & Yin, Z.-Y. (2006). Remote sensing image-based analysis of the relationship between urban heat Island and land use/cover changes. *Remote Sensing of Environment*, 104(2), 133–146. <https://doi.org/10.1016/j.rse.2005.11.016>
- Chen, Y., Wang, J., & Li, X. (2002). A study on urban thermal field in summer based on satellite remote sensing. *Remote Sensing for Land and Resources*, 4, 55–59. <https://doi.org/10.6046/gtzyyg.2002.04.12>
- Chow, W. T. L., & Roth, M. (2006). Temporal dynamics of the urban heat Island of Singapore. *International Journal of Climatology*, 26, 2243–2260. <https://doi.org/10.1002/joc.1364>
- Clinton, N., & Gong, P. (2013). MODIS detected surface urban heat islands and sinks: Global locations and controls. *Remote Sensing of Environment*, 134, 294–304. <https://doi.org/10.1016/j.rse.2013.03.008>
- Estoque, R. C., Murayama, Y., & Myint, S. W. (2017). Effects of landscape composition and pattern on land surface temperature: An urban heat island study in the megacities of Southeast Asia. *Science of the Total Environment*, 577, 349–359. <https://doi.org/10.1016/j.scitotenv.2016.10.195>
- Fast, J. D., Torcolini, J. C., & Redman, R. (2005). Pseudovertical temperature profiles and the urban heat Island measured by a temperature datalogger network in Phoenix, Arizona. *Journal of Applied Meteorology*, 44, 3–13. <https://doi.org/10.1175/JAM-2176.1>
- Fonseka, H. P. U., Zhang, H., Sun, Y., Su, H., Lin, H., & Lin, Y. (2019). Urbanization and its impacts on land surface temperature in Colombo metropolitan area, Sri Lanka, from 1988 to 2016. *Remote Sensing*, 11(8), 957. <https://doi.org/10.3390/rs11080957>
- Fotheringham, A. S., Charlton, M. E., & Brunsdon, C. (2001). Spatial variations in school performance: A local analysis using geographically weighted regression. *Geographical and Environmental Modelling*, 5(1), 43–66. <https://doi.org/10.1080/13615930120032617>
- Huang, C., Li, X., & Lu, L. (2008). Retrieving soil temperature profile by assimilating MODIS LST products with ensemble Kalman filter. *Remote Sensing of Environment*, 112(4), 1320–1336. <https://doi.org/10.1016/j.rse.2007.03.028>
- Huang, Y., Yuan, M., & Lu, Y. (2017). Spatially varying relationships between surface urban heat islands and driving factors across cities in China. *Environment and Planning B: Urban Analytics and City Science*, 46, 377–394. <https://doi.org/10.1177/2399808317716935>
- Imhoff, M. L., Zhang, P., Wolfe, R. E., & Bounoua, L. (2010). Remote sensing of the urban heat Island effect across biomes in the continental USA. *Remote Sensing of Environment*, 114, 504–513. <https://doi.org/10.1016/j.rse.2009.10.008>
- Jiménez-Muñoz, J. C., & Sobrino, J. A. (2003). A generalized single-channel method for retrieving land surface temperature from remote sensing data. *Journal of Geophysical Research*, 108, 4688–4695. <https://doi.org/10.1029/2003JD003480>
- Karnieli, A., Agam, N., Pinker, R. T., Anderson, M., Imhoff, M. L., & Gutman, G. G. (2010). Use of NDVI and land surface temperature for drought assessment: Merits and limitations. *Journal of Climate*, 23, 618–633. <https://doi.org/10.1175/2009JCLI2900.1>
- Kustas, W., & Anderson, M. (2009). Advances in thermal infrared remote sensing for land surface modeling. *Agricultural and Forest Meteorology*, 149, 2071–2081. <https://doi.org/10.1016/j.agrformet.2009.05.016>
- Leung, Y., Mei, C. L., & Zhang, W. X. . (2003). Statistical test for local patterns of spatial association. *Environment and Planning A*, 35(4), 725–744. <https://doi.org/10.1068/a3550>

- Li, Z. L., Tang, B. H., Wu, H., Ren, H., Yan, G., Wan, Z., ... Sobrino, A. J. (2013). Satellite-derived land surface temperature: Current status and perspectives. *Remote Sensing of Environment*, 131, 14–37. <https://doi.org/10.1016/j.rse.2012.12.008>
- Manjula, R., Estoque, R. C., & Yuji, M. (2017). An urban heat Island study of the Colombo metropolitan area, SriLanka, based on landsat data (1997–2017). *ISPRS International Journal of Geo-Information*, 6(7), 189. <https://doi.org/10.3390/ijgi6070189>
- Markus, M., AndreoVerónica, & Markus, N. (2017). A new fully gap-free time series of land surface temperature from modislt data. *Remote Sensing*, 9(12), 1333. <https://doi.org/10.3390/rs9121333>
- Mathew, A., Sreekumar, S., Khandelwal, S., Kaul, N., & Kumar, R. (2016). Prediction of land-surface temperatures of Jaipur city using linear time series model. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, 9(8), 3546–3552. <https://doi.org/10.1109/JSTARS.2016.2523552>
- McFeeters, S. K. (1996). The use of the normalized difference water index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7), 1425–1432. <https://doi.org/10.1080/01431169608948714>
- Myint, S. W., Wentz, E. A., Brazel, A. J., & Quattrochi, D. A. (2013). The impact of distinct anthropogenic and vegetation features on urban warming. *Landscape Ecology*, 28(5), 959–978. <https://doi.org/10.1007/s10980-013-9868-y>
- Neteler, M. (2010). Estimating daily land surface temperatures in mountainous environments by reconstructed MODIS LST data. *Remote Sensing*, 2(1), 333–351. <https://doi.org/10.3390/rs1020333>
- NourEldeen, N., Mao, K., Yuan, Z., Shen, X., Xu, T., & Qin, Z. (2020). Analysis of the spatiotemporal change in land surface temperature for a long-term sequence in Africa (2003–2017). *Remote Sensing*, 12, 488. <https://doi.org/10.3390/rs12030488>
- Olah, C., 2015. Understanding LSTM networks. Retrieved from <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>.
- Owen, T. W., Carlson, T. N., & Gillies, R. R. (1998). An assessment of satellite remotely-sensed land cover parameters in quantitatively describing the climatic effect of urbanization. *International Journal of Remote Sensing*, 19, 1663–1681. <https://doi.org/10.1080/014311698215171>
- Peng, S., Piao, S., Ciais, P., Friedlingstein, P., Ottle, C., Bréon, F. M., ... Myneni, R. B. (2012). Surface urban heat Island across 419 global big cities. *Environmental Science & Technology*, 46, 696–703. <https://doi.org/10.1021/es2030438>
- Qin, Z., Karnieli, A., & Berliner, P. (2001). A mono-window algorithm for retrieving land surface temperature from landsat tm data and its application to the Israel–Egypt border region. *International Journal of Remote Sensing*, 22(18), 3719–3746. <https://doi.org/10.1080/01431160010006971>
- Tonooka, H. (2001). An atmospheric correction algorithm for thermal infrared multi-spectral data over land—A water-vapor scaling method. *IEEE Transactions on Geoscience and Remote Sensing*, 39, 682–692. <https://doi.org/10.1109/36.911125>
- Wang, Z., Fan, C., Zhao, Q., & Myint, S. W. (2020). A geographically weighted regression approach to understanding urbanization impacts on urban warming and cooling: A case study of Las Vegas. *Remote Sensing*, 12(2), 222. <https://doi.org/10.3390/rs12020222>
- Weng, Q. (2001). A remote sensing-GIS evaluation of urban expansion and its impact on surface temperature in Zhujiang Delta, China. *International Journal of Remote Sensing*, 22(10), 1999–2014. <https://doi.org/10.1080/713860788>
- Weng, Q., Lu, D., & Jacquelyn, S. (2003). Estimation of land surface temperature–vegetation abundance relationship for urban heat Island studies. *Remote Sensing of Environment*, 89, 467–483 <https://doi.org/10.1016/j.rse.2003.11.005>.
- Xiao, C., Chen, N., Hu, C., Wang, K., Gong, J., & Chen, Z. (2019). Short and mid-term sea surface temperature prediction using time-series satellite data and LSTM-AdaBoost combination approach. *Remote Sensing of Environment*, 233, 111358. <https://doi.org/10.1016/j.rse.2019.111358>
- Zha, Y., Gao, J., & Ni, S. (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing*, 24(3), 583–594. <https://doi.org/10.1080/01431160304987>
- Zhang, P., Imhoff, M. L., Wolfe, R. E., & Bounoua, L. (2010). Characterizing urban heat islands of global settlements using MODIS and night-time lights products. *Canadian Journal of Remote Sensing*, 36, 185–196. <https://doi.org/10.5589/m10-039>
- Zhang, Q., Wang, H., Dong, J., Zhong, G., & Sun, X. (2017). Prediction of sea surface temperature using long short-term memory. *IEEE Geoscience and Remote Sensing Letters*, 14, 1745–1749. <https://doi.org/10.1109/LGRS.2017.2733548>
- Zhao, C., Jensen, J., Weng, Q., & Weaver, R. (2018). A geographically weighted regression analysis of the underlying factors related to the surface urban heat Island phenomenon. *Remote Sensing*, 10(9), 1428. <https://doi.org/10.3390/rs10091428>
- Zhou, W., Qian, Y., Li, X., Li, W., & Han, L. (2013). Relationships between land cover and the surface urban heat Island: Seasonal variability and effects of spatial and thematic resolution of land cover data on predicting land surface temperatures. *Landscape Ecology*, 29(1), 153–167. <https://doi.org/10.1007/s10980-013-9950-5>
- Zhu, Z., Woodcock, C. E., Holden, C., & Yang, Z. (2015). Generating synthetic Landsat images based on all available Landsat data: Predicting Landsat surface reflectance at any given time. *Remote Sensing of Environment*, 162, 67–83. <https://doi.org/10.1016/j.rse.2015.02.009>

**How to cite this article:** Jia H, Yang D, Deng W, Wei Q, Jiang W. Predicting land surface temperature with geographically weighed regression and deep learning. *WIREs Data Mining Knowl Discov*. 2020;e1396. <https://doi.org/10.1002/widm.1396>