Spatiotemporal Downscaling Method of Land Surface Temperature Based on Daily Change Model of Temperature

Penghua Hu, Aihui Wang, Yingbao Yang, Xin Pan, Xiejunde Hu, Yuncheng Chen, Xuechun Kong, Yao Bao, Xiangjin Meng, and Yang Dai

Abstract—Land surface temperature (LST) is one of the most crucial variables of surface energy processes. However, the trade-off between spatial and temporal resolutions of remote sensing data has greatly limited the availability of concurrently high-spatiotemporal resolution LST data for wide applications. Existing downscaling methods are easily affected by null values of LST data and effective time distribution of high-resolution LST data, resulting in large downscaling errors at sometimes. Within this context, this study proposes a novel spatiotemporal fusion model of LST based on diurnal variation information (BDSTFM) to predict LST data with a high temporal resolution and spatiotemporal continuity based on FY-4A and MODIS. Results indicated that the accuracy of the downscaling results was comparable to that of MODIS LST products. The BDSTFM model exhibited the following characteristics: use low-spatial resolution data to establish a diurnal temperature cycle (DTC) model for scale deduction, and retention of the temporal distribution characteristics of LST data; extend the observation time of high-spatial resolution data to improve the accuracy and stability of the model; add an invalid pixel reconstruction step that considers the LST spatiotemporal continuity, and can obtain a realistic and reliable 1-km seamless LST datasets at hourly intervals under clear skies. Compared with the enhanced spatial and temporal adaptive reflectance fusion model (ESTARFM), 4-parameter DTC model, and Random Forest model, the BDSTFM model attained a higher downscaling accuracy.

Index Terms—Diurnal temperature cycle (DTC), downscaling, FY-4A, land surface temperature (LST), MODIS.

I. INTRODUCTION

LAND surface temperature (LST) is one of the most crucial variables of surface energy processes [1], LST has been widely used in the study of the urban heat island (UHI) effect [2], [3] and the assessment of the risks associated with high temperatures in cities [4]. Due to the limitation of technology, the existing remote sensing system cannot provide the thermal infrared data (TIR) satisfying high spatiotemporal resolution. The TIR data obtained by polar-orbiting satellites may miss the optimal observation time, especially in rapidly changing areas, while TIR data obtained by geostationary satellites may not contain the spatial details of heterogeneous landscapes due to their rough spatial resolution (usually 3–5 km). This greatly limits the potential applications of satellite source LST data in various fields.

At present, various methods to improve spatial and temporal resolutions have been proposed, which can be divided into two categories [5]: using single sensor LST or multisensor LST pair. The first category contains two different ways, one is spatial downscaling for LST data with low spatial resolution, and the other is time interpolation for LST data of polar orbit satellites. The second category is LST downscaling method based on data fusion.

The spatial downscaling method has mainly been proposed based on the assumption that the relationships between LST and surface parameters (such as vegetation abundance, dynamic soil moisture, and land cover type) exhibit scale invariance. The differences between these methods include the selected regression kernel parameters and regression models. Regression kernel parameters can be divided into single scale factor and multiple scale factors. Typical single scale factor models include DisTrad method [6] and TShARP method [7]. Both models can achieve LST reconstruction by establishing a regression relationship between LST and normalized vegetation index (NDVI). However, it is often difficult to obtain a good reconstruction effect in areas with many land-cover types by using a single scale factor. In this case, using multiple scale factors to establish a regression relationship can improve the LST downscaling accuracy. Yang et al. [8] implemented LST downscaling in mixed areas with three or four surface types, by establishing a multiple linear regression relationship between LST and multiple scale factors. Linear regression is robust and easy to implement, but this method cannot be used to solve the nonlinear relationship between signals. Nonlinear statistical regression models (such as artificial neural networks) can establish nonlinear statistical relationships among multiple variables. Venkateshwarlu et al. [9] used an artificial neural network to achieve LST downscaling. As a popular method, machine learning provides unique
advantages in predicting nonlinear relationships between variables. The Random Forest model, as a machine learning method, can establish nonlinear regression relationships between LST and various land surface scale factors. Yang et al. [10] used the Random Forest model to achieve LST downscaling, and Pan et al. [11] added a new factor to their model to improve the downscaling accuracy in desert regions. Regarding the selection of land surface scale factors, studies have demonstrated that different land surface scale factors involved in regression can yield different downscaling effects, and the selection of scale factors is related to the underlying surface type [12]. However, the regression kernel parameters used in existing methods cannot fully explain the spatial and temporal distributions of LST, which affects the spatiotemporal continuity of downscaled LST data. Moreover, different spatial downscaling methods for LST provide a certain regional applicability due to the high auxiliary data requirements and spatial relationship variations between LST and scale factors [5].

The temporal interpolation method temporally interpolates discrete observations, typically for polar-orbiting satellites, which usually involves the daily/annual temperature cycle (DTC/ATC) model to interpolate LST data with a high temporal resolution and low spatial resolution [5], [14], [15]. At the diurnal scale, Duan et al. [1] established a DTC model by using four MODIS observations under clear sky conditions and obtained an LST dataset with a resolution of 1 km at each moment. Hong et al. [16] provided an overview of commonly used four-parameter DTC models, and laid a foundation for the generation of spatiotemporally continuous LST products. Due to the limitation of the amount of data available for polar-orbiting satellites, Quan et al. [17] proposed a hybrid method combining a DTC model with a linear temperature mixing model to reconstruct the diurnal temperature cycle (DTC) over consecutive days using polar-orbiting satellite data. On an annual scale, Sismanidis et al. [18] used an ATC model to reduce the geostationary LST resolution of 4 km during the day and night to 1 km. Zhan et al. [19] proposed a dynamic methodology to implement LST decomposition by combining the DTC and ATC models. Although the temporal interpolation method has achieved good results, there remain problems in these models, such as atmospheric conditions, land cover, and missing data (at least four observations are needed per day for certain models), which affects the performance of the models.

The downscaling method based on data fusion mainly includes endmember unmixing-, sparse representation-, and weight function-based fusion methods for multisresolution LST pairs at different times [5]. This method combines the advantages of different sensors, and considers the continuity in the spatial distribution of pixels to improve the accuracy of the results. The most typical weight function-based fusion method is the spatial and temporal adaptive reflectance fusion model (STARFM) proposed by Feng et al. [20]. The STARFM method can accurately predict the reflectance value of high-resolution images under a high surface uniformity. Given the deficiencies of the STARFM model, Zhu et al. [21] proposed the enhanced STARFM (ESTARFM), which improved the prediction accuracy of the STARFM on high-resolution pixels under surface coverage type changes and mutation occurrence. Although the data fusion method was originally designed for reflectance data, these methods have been proven useful for the generation of daily Landsat-like thermal data such as evapotranspiration and LST data [22], [23]. Wu et al. [24] introduced the spatiotemporal variational model to extend the STARFM, and used it to combine GOES and Landsat LST data to obtain a product with both the temporal resolution of GOES data and the spatial resolution of Landsat data. Buscail et al. [4] proposed an integrated framework to blend spatiotemporal LST by combining the above-mentioned multiple temporal and spatial fusion models, fused Landsat MODIS and FY-2, and obtained a surface temperature product with a spatial resolution of 100 m and a temporal resolution of 1 h. There have been many studies on LST downscaling based on data fusion. However, due to the influence of reference images, there are many invalid values in the downscaling results. Therefore, the scale reduction effect of the model must be further improved.

Considering the disadvantages of the above-mentioned methods, this article fully utilized the characteristics of geostationary satellites (FY-4A) with a high temporal resolution and polar orbit satellites (MODIS) with a high spatial resolution, and proposed a spatiotemporal fusion model for the surface temperature based on diurnal variation information (BDSTFM) to predict LST data with a high temporal resolution.

Specifically, the first step involved the use of FY-4A and MODIS LST datasets to establish a DTC model suitable for the MODIS resolution. The second step fully utilized image spatial and temporal neighborhood information to optimize the quality of the fusion LST dataset. A further description is provided in Section III. It is tested and evaluated using simulated data, actual satellite data, and field observations, discussion, and conclusions are provided in Sections V and VI, respectively.

II. STUDY AREA AND DATA

A. Study Area

The study area in this article is located in the Heihe River Basin in Northwest China and Jiangsu Province in eastern China (Fig. 1). The Heihe River Basin (97°24′–102°10′E, 37°41′–42°42′N) is mainly located in Qinghai Province, Gansu Province, and Inner Mongolia Autonomous Region of China. The climate is dry with rare and concentrated precipitation, strong winds, abundant sunshine, high solar radiation, and a large temperature difference between day and night. The region exhibits a complex terrain, including plains, plateaus, and mountains. The upstream area is the southern mountainous region of the Qilian Mountains, where alpine forests and shrublands are distributed, with an elevation of 3000–6000 m. The middle reaches, located in the central part of the Hexi Corridor, comprise an inland accumulation plain with several typical oases and deserts occurring at an elevation ranging from 1500–3000 m. The lower reaches are flat, mainly including hilly and Gobi desert areas, with an elevation ranging from 1000–1500 m. Especially in summer, clear-sky conditions prevail, the effect of the LST downscaling method can be suitably assessed under clear-sky conditions. In addition, there are many meteorological stations in this area, which is convenient for obtaining measured data. In this experiment, we selected observation data for 6
TABLE I
INFORMATION ON THE SIX AUTOMATIC WEATHER OBSERVATION STATIONS

| Site name       | Location       | Land cover types       | Area       |
|-----------------|----------------|------------------------|------------|
| A’rou station   | 100.4643E,38.0473N | Mountain meadow       | Upper reach|
| Dasalong station| 98.9406E,38.8399N | Alpine meadow         | Upper reach|
| Desert station  | 100.9872E,42.1135N | Red sand desert      | Lower reach|
| Sidaqiao station| 101.1374E,42.0012N | Blame the willow     | Lower reach|
| Daman station   | 100.3722E,38.8555N | Corn field            | Middle reach|
| Zhangye station | 100.4464E,38.9751N  | Reed wetland          | Middle reach|

B. Datasets

Table II lists the datasets used in this study, including an FY-4A LST product dataset (FY4A_AGRI_L2_LST), MODIS daily land temperature emissivity product dataset (MOD11B1/MYD11B1), MODIS daily land temperature emissivity product dataset (MOD11A1/MYD11A1) and in situ measurements (surface upwelling longwave radiation and atmospheric downwelling longwave radiation).

1) MODIS Data: Daily MODIS/LST products retrieved from Terra and Aqua satellites (MOD11A1 and MYD11A1, downloaded from NASA’s EOSDIS, http://earthdata.nasa.gov) with a spatial resolution of 1 km were used. Both Terra and Aqua overpass any point on the Earth’s surface twice a day in most cases, and therefore, the MODIS sensors onboard both satellites are able to provide four thermal observations per day, which enables us to interpolate the diurnal LST. These satellites overpass the equator at approximately 10:30, 13:30, 22:30, and 01:30 local solar time, while the overpass time at the other locations varies. Therefore, the overpass time was extracted from the MOD/MYD11A1 product for each pixel.

Daily MODIS/narrowband emissivity data products were obtained from Terra and Aqua satellites with a spatial resolution of 5 km. Use this data to calculate the broadband emissivity data,
TABLE II
PRODUCT DATA PARAMETERS

| Data | Variable | Spatiotemporal resolution |
|------|----------|---------------------------|
| FY4A_AGRI_L2_LST | LST | 15 min/60 min, 4 km |
| MOD11A1/MYD11A1 | LST, observation time, quality control data | Daily, 1 km |
| MOD11B1/MYD11B1 | Narrowband emissivity | Daily, 6 km |
| In situ measurements | Surface upwelling and atmospheric downwelling longwave radiation | 10 min, point |

Fig. 2. BDSTFM model framework.

the calculation formula is as follows [13]:

\[
\epsilon_b = 0.2122 \times \epsilon_{29} + 0.3859 \times \epsilon_{31} + 0.4029 \times \epsilon_{32}
\]

where \(\epsilon_b\) is the broadband emissivity, \(\epsilon_{29}, \epsilon_{31}\) and \(\epsilon_{32}\) is the narrow band emissivity of MODIS 29, 31, 32 bands.

2) FY-4A_AGRI_L2_LST Data: FY-4A, a new generation of China’s geostationary meteorological satellite, was successfully set over the equator at 99.5 degrees east longitude on December 17, 2016. This article uses the LST data downloaded from cloud satellite remote sensing data service network (http://satellite.nsmc.org.cn/PortalSite/Default.aspx). It can provide LST products with spatial resolution of 4 km and temporal resolution of 15 min, 60 min, and irregular intervals. In this study, FY-4A LST data from September 23, 2019, to October 1, 2019 (DOY266-DOY274) were used to analyze the model performance, and FY-4A LST data from September 24, 2019, to September 30, 2019 (DOY267-DOY273), as well as MODIS LST data, are used to predict spatiotemporally continuous datasets with a 1-km resolution.

3) Site Actual LST: The actual LST was estimated from upwelling and downwelling longwave radiation observations obtained by pyranometers using the following equation:

\[
T_s = \left[ \frac{R_{lu} - (1 - \epsilon) \cdot R_{ld}}{\epsilon \cdot \delta} \right]^{0.25}
\]

where \(R_{lu}\) is the surface upwelling longwave radiation, \(R_{ld}\) is the atmospheric downwelling long-wave radiation, \(\epsilon\) is the land surface emissivity calculated from (1), \(T_s\) is the actual LST, and \(\delta\) is the Stefan–Boltzmann constant.

III. METHODOLOGY

The BDSTFM comprises two steps (Fig. 2). The first step entails the use of FY-4A and MODIS LST datasets to establish a DTC model suitable for the MODIS resolution and completion of preliminary downscaling of the LST dataset (referred to as LST downscaling, as detailed in Section III.A). The second step aims to fully utilize spatial and temporal neighborhood information of images to optimize the quality of the predicted LST...
dataset. The quality optimization process comprises two parts, namely the reconstruction of invalid values and correction of the residual error, and a MODIS-like LST dataset is finally obtained with a continuous spatiotemporal hourly resolution (denoted as LST quality optimization, as detailed in Section III.B). Before LST downscaling using the proposed model, LST datasets with different resolutions should be resampled to the same spatial resolution, and the resampling method is the nearest neighbor method.

A. LST Downscaling

The theoretical assumption for LST downscaling based on the DTC model is as follows: for the same pixel, the parameters of the DTC model corresponding to LST datasets with different spatial resolutions vary, but the differences are similar over a certain period. A concrete downscaling scheme was designed based on this hypothesis, and the correctness of this hypothesis was verified. The design scheme mainly used MODIS LST data to correct the DTC model established based on FY-4A LST data. A DTC model suitable for MODIS was obtained, and the implementation was conducted pixel by pixel. The specific implementation process was as follows.

1) Build DTC Model Pixel by Pixel: DTC model describes the change of LST over time in a day. Duan et al. [14] compared and evaluated the simulation effects of six DTC models under the condition of clear sky, and the results showed that the JNG06 model proposed by Jiang et al. [25] had a high overall fitting accuracy. Therefore, in this study, the JNG06 model is used to describe the diurnal variation information of surface temperature, with the following specific expression:

\[
T_{day}(t) = T_0 + T_a \cos(\beta (t - t_m)), t < t_s
\]
\[
T_{night}(t) = b_1 + b_2 e^{\alpha (t-t_s)}, t \geq t_s
\]

with
\[
b_2 = [-\beta T_a \sin(\beta (t_s - t_m))]/\alpha
\]
\[
b_1 = T_0 + T_a \cos(\beta (t_s - t_m)) - b_2 \tag{3}
\]

where \(T_{day}\) and \(T_{night}\) are the LST values during the day and at night, respectively. \(T_0\) is the LST at sunrise (K); \(T_a\) is the temperature amplitude (K), meaning the difference between \(T_0\) and the maximum temperature during a diurnal cycle; \(t_m\) is the time of maximum temperature within a diurnal cycle; \(t_s\) is the start time of the attenuation function; \(\alpha\) and \(\beta\) have no clear physical meaning. \(t\) is the time, all moments in the equation are expressed in hours, and the time system is the coordinated universal time (UTC) system.

Because of the nonlinearity of the DTC model, the model parameters cannot be directly determined with the corresponding equations. In this study, Levenberg–Marquardt minimization schemes were utilized to fit the model.

According to Duan et al. [14], the initial values of the model parameters were set. Since this model assumes LST change under clear-sky conditions and contains 6 free parameters, it is required that the model can be established only considering at least 6 available moments on a given day.

When the obtained FY-4A LST data meet the conditions of DTC model establishment, the parameters of the DTC model can be calculated pixel by pixel.

\[
LST_{fy4a}(t, d) = DTC\left(T_{fy4a,d}, t_{fy4a,d}, t_{fy4a,d}, \alpha_{fy4a,d}, \beta_{fy4a,d}\right) \tag{4}
\]

where \(d\) denotes the date corresponding to the DTC model, \(t\) is the time, and \(LST_{fy4a}(t, d)\) denotes the FY-4A LST data at time \(t\) on date \(d\). FY-4A LST data at each moment can be calculated by using the 6 free parameters obtained via solution, denoted as \(LST_{f0}(t, d)\).

2) Increase the MODIS Dataset Frequency: If the MODIS LST dataset was directly used to correct the established DTC model, the uncertainty in LST values would lead to local optimization [1], and a large error would occur at a time far from the MODIS transit time, which would affect the stability of the model. To solve this problem, the observation frequency of MODIS was increased before data correction so that the maximum number of effective moments available for the MODIS data was changed from 4 to 6. The frequency-increasing scheme for the MODIS dataset was as follows.

MODIS crosses the equator at approximately 10:30, 13:30, 22:30, and 01:30 (\(T_1, T_2, T_3, \text{ and } T_4\), respectively) at local solar time. Through this study, it was found that adding two moments at approximately 8:30 (i.e., the moment 2 h away from time \(T_1\), which is denoted as \(T_0\)) and approximately 18:00 (i.e., the middle moment between times \(T_2\) and \(T_3\), denoted as \(T_{2.5}\)) could greatly improve the reconstruction effect of the LST dataset because these six moments are uniformly distributed within a given diurnal variation cycle. It could be found from the DTC model that the temperature changes from \(T_0\) to \(T_1\) and from \(T_{2.5}\) to \(T_3\) are approximately linear [26], so it could be considered that the difference values of LST with the different spatial resolutions are approximately equal at this stage, given as follows (in the calculation, all local solar times are converted into UTC times):

\[
LST_{\text{modis}}(T_0, d) - LST_{f0}(T_0, d) = LST_{\text{modis}}(T_1, d) - LST_{f0}(T_1, d)
\]
\[
LST_{\text{modis}}(T_2.5, d) - LST_{f0}(T_2.5, d) = LST_{\text{modis}}(T_3, d) - LST_{f0}(T_3, d) \tag{5}
\]

where \(LST_{f0}\) is the FY-4A LST value at the different moments calculated with DTC model, \(LST_{\text{modis}}\) is the MODIS LST value at the different times, and the LST values at times \(T_0\) and \(T_{2.5}\) corresponding to the MODIS resolution can be calculated as follows:

\[
LST_{\text{modis}}(T_0, d) = LST_{f0}(T_0, d) + (LST_{\text{modis}}(T_1, d) - LST_{f0}(T_1, d))
\]
\[
LST_{\text{modis}}(T_{2.5}, d) = LST_{f0}(T_{2.5}, d) + (LST_{\text{modis}}(T_3, d) - LST_{f0}(T_3, d)) \tag{6}
\]
It should be noted that due to the lack of MODIS data, even though the resolution of the MODIS dataset was increased, each pixel did not contain 6 available moments every day, and the quality of the LST dataset at other extended moments still resulted in certain errors. Therefore, a DTC model corresponding to MODIS data could not be fully established just by using the extended MODIS dataset.

3) Calculate the Deviation Coefficient: The DTC model corresponding to the MODIS LST dataset of on date d is

\[
LST_{modis}(t, d) = DTC(T_0^{mod is,d}, T_a^{mod is,d}, t_m^{mod is,d}, t_s^{mod is,d}, \alpha_{mod is,d}, \beta_{mod is,d}, t)
\]

where \((T_0^{mod is,d}, T_a^{mod is,d}, t_m^{mod is,d}, t_s^{mod is,d}, \alpha_{mod is,d}, \beta_{mod is,d}, t)\) are DTC model parameters corresponding to date d of the MODIS LST dataset.

Due to the systematic deviation between FY-4A and MODIS LST data, the DTC model parameters of FY-4A on date d cannot be directly applied to the MODIS LST data. This model corrects the parameters by adding offset coefficient, and believes that DTC parameters corresponding to FY-4A and MODIS of the same pixel have deviation [as shown in (8)], and this deviation is the same on multiple adjacent days. The deviation can be calculated as follows:

\[
LST_{modis}(t, d) = DTC(T_0^{f1,d}, T_a^{f1,d}, t_m^{f1,d}, t_s^{f1,d}, \alpha^{f1,d}, \beta^{f1,d}, t) + \xi
\]

\[
T_0^{f1,d} = T_0^{fy4a,d} + \varepsilon_1
\]

\[
t_m^{f1,d} = t_m^{fy4a,d} + \varepsilon_3
\]

\[
t_s^{f1,d} = t_s^{fy4a,d} + \varepsilon_4
\]

\[
\alpha^{f1,d} = \alpha^{fy4a,d} + \varepsilon_5
\]

\[
\beta^{f1,d} = \beta^{fy4a,d} + \varepsilon_6
\]

where \(\xi\) is the regression residual; \(d\) stands for different dates, and multiple dates can be taken; the value of t is \(T_0, T_a, T_m, T_s\) at the 6 moments corresponding to MODIS; \((\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5, \varepsilon_6)\) are the 6 deviation coefficients to be calculated; \((T_0^{f1,d}, T_a^{f1,d}, t_m^{f1,d}, t_s^{f1,d}, \alpha^{f1,d}, \beta^{f1,d})\) is the approximate parameters of DTC model corresponding to MODIS data.

Since DTC model parameter \((T_0^{fy4a,d}, T_a^{fy4a,d}, t_m^{fy4a,d}, t_s^{fy4a,d}, \alpha^{fy4a,d}, \beta^{fy4a,d})\) corresponding to FY-4A's LST dataset have been calculated, the data needed to solve the deviation coefficient is at least six. Therefore, the deviation coefficients can be solved when the number of available MODIS data in the date participating in the fitting is greater than six. As it is a nonlinear model, Levenberg-Marquardt least square algorithm is used to solve the deviation coefficient. The initial value of 6 deviation coefficients \((\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5, \varepsilon_6)\) is set to 0, 1, 0, 0, 1, and 1, respectively.

After calculating deviation coefficients, the approximate parameters \((T_0^{f1,d}, T_a^{f1,d}, t_m^{f1,d}, t_s^{f1,d}, \alpha^{f1,d}, \beta^{f1,d})\) of the DTC model corresponding to MODIS data can be obtained according to (8), and then the DTC model can be used to realize the downsampling of the LST dataset with high spatial and temporal resolutions. The calculated data of each pixel at time t on date d can be denoted as \(LST_{f1}(t, d)\), and its corresponding parameter \((T_0^{f1,d}, T_a^{f1,d}, t_m^{f1,d}, t_s^{f1,d}, \alpha^{f1,d}, \beta^{f1,d})\) can be denoted as \(DTC_{f1}^{d}\).

B. LST Quality Optimization

The theoretical assumption for quality optimization of LST datasets is that LST changes are spatially continuous and similar adjacent pixels exhibit similar LST values. There are two main objectives of quality optimization: the first objective is to reconstruct empty pixels due to atmospheric and terrain effects; the second objective is to fully utilize spatial neighborhood information of LST to correct the residual error. The design scheme of pixel-by-pixel operation is as follows.

1) Filtering Similar Pixels: To fully utilize spatial neighborhood information of the LST dataset, similar pixels must be filtered before calculation, and the filtering method is as follows.

A 9 × 9 window (containing approximately four coarse resolution pixels) was first built for the pixel to be treated; then, the standard deviation between each pixel and the central pixel in the window was calculated. The standard deviation was calculated using high-resolution LST time series data. In the end, the mean value of the standard deviation was calculated. A pixel whose standard deviation was lower than the mean value could be regarded as a similar pixel.

2) Reconstructing Invalid Pixels: Due to the influence of terrain and shadow factors, invalid values exist in the image, which seriously affects the quality of the predicted LST dataset. However, due to the error in MODIS data, pixels could still contain errors during regression, resulting in the phenomenon of plaques. Therefore, it is necessary to reconstruct invalid values and correct the patch phenomenon.

The operation was performed pixel by pixel and day by day. First, a neighborhood window was established to identify similar pixels. Then predicted LST data of the similar and the central pixels were used to calculate the new DTC model parameters \(T_0^{f2,d}, T_a^{f2,d}, t_m^{f2,d}, t_s^{f2,d}, \alpha^{f2,d}, \beta^{f2,d}\) corresponding to the central pixel, and the new DTC model parameters \(T_0^{f2,d}, T_a^{f2,d}, t_m^{f2,d}, t_s^{f2,d}, \alpha^{f2,d}, \beta^{f2,d}\) are denoted as \(DTC_{f2}^{d}\). If the central pixel is null, then only similar pixels are used to calculate the new DTC model parameter corresponding to the central pixel on date d. The deviation is calculated as follows:

\[
LST_{f2}^{d}(t, d) = DTC(T_0^{f2,d}, T_a^{f2,d}, t_m^{f2,d}, t_s^{f2,d}, \alpha^{f2,d}, \beta^{f2,d}, t)
\]

where k is 0, 1, 2 ...K when the central pixel is not null; when the central pixel is null, the value of k is 1, 2 ...K (K is the number of similar pixels); t is the valid time of the corresponding date d.

Then, a new LST image without null value can be calculated by using the calculated model parameters, denoted as \(LST_{f2}(t, d)\).
C. Correcting Regression Residuals

Currently, there remain errors between the LST dataset with high spatial and temporal resolutions and the real LST dataset. To reduce the effect of the regression error, the regression residuals must be calculated.

First, using (8) and (9) to obtain new deviation coefficients $\xi_1, \xi_2, \xi_3, \xi_4, \xi_5$ and $\xi_6$ corresponding to the central pixel, and the DTC parameter $T_0^{f_3, d, p_0}$, $T_{f_3, d, p_0}^{f_3, d, p_0}$, $T_{f_3, d, p_0}^{f_3, d, p_0}$, $T_{f_3, d, p_0}^{f_3, d, p_0}$, $T_{f_3, d, p_0}^{f_3, d, p_0}$, and $T_{f_3, d, p_0}^{f_3, d, p_0}$ corresponding to different dates of the central pixel, denoted as $DTC_{f_3}^{f_3, d, p_0}$. The difference from Section III-A.3 is that in addition to the central pixel, the similar pixel is also involved in the calculation of the deviation coefficient. Then, the LST data $LST_{f_3}^{p_0}(t, d)$ corresponding to the central pixel can be calculated by using the obtained DTC parameters. By using (11), the residual $R_{f_3}^{p_0}(t, d)$ at the corresponding moment of the central pixel can be calculated.

$$R_{f_3}^{p_0}(t, d) = LST_{mod \_is}^{p_0}(t, d) - LST_{f_3}^{p_0}(t, d)$$  \hspace{1cm} (11)

where the value of $t$ is $T_{f_3, d, p_0}^{f_3, d, p_0}$.

Based on the previous step, the residual of daily regression $R_{f_3}(t, d)$ could be obtained. However, the difference between other moments $LST_{f_3}(t, d)$ and the actual MODIS resolution LST data remains unknown.

Therefore, to reduce errors, linear interpolation is used to calculate the errors at other times, and the residual $R_{f_3}(t, d)$ corresponding to each time of each pixel is finally obtained. According to (12), the spatiotemporal continuous LST dataset $LST_{f_4}(t, d)$ could be obtained, namely the final reconstruction result.

$$LST_{f_4}(t, d) = LST_{f_3}(t, d) + R_{f_3}(t, d).$$  \hspace{1cm} (12)

Finally, the error correction method could correct the system deviation in the fused LST values under clear-sky conditions. The BDSTFM model does not require additional data, such as NDVI data, resulting in a reduced data input, which is critical for the transferability of this method to other similar regions of the world.

D. Evaluation Strategies

To evaluate BDSTFM more comprehensively, two strategies are employed to evaluate its accuracy.

1) Using Generated Simulation Data: The first strategy uses the aggregation-disaggregation method using simulation data (Section V-A describes how to generate simulated dataset) as input, which aims to comprehensively assess BDSTFM. The evaluation results are presented in Section V-A. Because the accuracy of the predicted LST dataset is affected by three factors: the accuracy of the BDSTFM model, the accuracy of the DTC model, and the accuracy of the high spatial resolution LST dataset involved in the fitting. We excluded the impact of the accuracy of the DTC model and the accuracy of the high spatial resolution LST dataset involved in the fitting on the downscaling accuracy when using this strategy for evaluation. In the evaluation, the performance of the model is evaluated from three aspects, namely: 1) the influence of the extended LST dataset on the downscaling accuracy; 2) the influence of the number of days involved in the fitting on the downscaling accuracy of the LST dataset; and 3) the applicability of model to regions with different climate types.

2) Using Real Satellite Data and Measured Site Data for Evaluation: The second strategy used real satellite data as input to evaluate the performance of the BDSTFM model under real scenarios; at the same time, the model was compared with the ESTARFM model and Random Forest model, and the results are provided in Section IV. The 4-km-resolution FY-4A LST dataset was used as the input low-spatial resolution dataset; the 1-km-MODIS LST dataset was used as the input high-spatial resolution data. Because MODIS LST values were not available at the target time, we used field measurements to evaluate the accuracy of the BDSTFM model.

Several statistical indices are used to give a quantitative assessment. The first index is mean error (ME), it is defined as follows:

$$ME = \frac{\sum^n (LST_D - LST_R)}{n}$$  \hspace{1cm} (13)

where $n$ is the total number of pixels or observations, $LST_D$ is the downscaled LST, and $LST_R$ is the reference LST.

The second index is mean-absolute error (MAE), it is defined as follows:

$$MAE = \frac{\sum^n |LST_D - LST_R|}{n}.$$  \hspace{1cm} (14)

The third index is root-mean-square error (RMSE), it is defined as follows:

$$RMSE = \sqrt{\frac{\sum^n (LST_D - LST_R)^2}{n}}.$$  \hspace{1cm} (15)

The fourth index is the Pearson correlation coefficient ($r$), it is defined as follows:

$$r = \frac{\sum^n (LST_D - LST_D) (LST_R - LST_R)}{\sqrt{\sum^n (LST_D - LST_D)^2} \sqrt{\sum^n (LST_R - LST_R)^2}}$$  \hspace{1cm} (16)

where $LST_D$ and $LST_R$ are the averages of $LST_D$ and $LST_R$.

IV. RESULTS

The BDSTFM model is used to obtain LST datasets with high spatial and temporal resolution. Therefore, FY-4A and MODIS LST datasets from September 24 to September 30, 2019 (DOY 267-273) were used to test in this section. The measured data of the six sites were used to evaluate the quality of LST datasets predicted with the BDSTFM model. In addition, the downscaling results were compared with those of ESTARFM, Random Forest, and 4-parameter DTC model.
A. MODIS Dataset Frequency-Increasing Results

To evaluate the performance of the developed method in Section III-A.2, the MODIS LST retrievals and MODIS-like LSTs predicted by BDSTFM were validated with the measured data of the six sites during the study periods.

Table III provides the accuracy of the MODIS LST retrievals and MODIS-like LST values predicted with the BDSTFM model compared with the in situ observations at the six weather stations from September 24 to 30, 2019. Overall, most of the MODIS LST retrievals exhibited excellent agreement with the in situ LST measurements at the six flux tower sites, with MAE (RMSE) ranging from 0.91 K (1.14 K) to 2.69 K (3.45 K) and r ranging from 0.985 (Desert station) to 0.998 (Zhangye station).

For the MODIS-like data made by BDSTFM, the MAE (RMSE) varies from 1.06 K (1.41 K) to 2.55 K (3.35 K), and r ranges from 0.968 to 0.988. The accuracy of the fused LST at the A’rou station is lower than that of the other stations. This may be attributed to the large error of MODIS LST retrievals used in BDSTFM, which directly affects the accuracy of the fused LST. The accuracy of MODIS LST retrievals may rely on land cover types, as some studies [27] reported that the accuracy of MODIS LST retrievals in arid areas (e.g., sandy regions) may be lower than those over other areas due to the uncertainty in emissivity estimation. Overall, the performance of the fused LST at the six flux towers is generally comparable to that of MODIS LST. These results indicate that the developed method is reliable for all stations with different land cover types.

B. Image Downscaling Results

Fig. 3 shows 1-km downscaled LST results obtained with the three models (BDSTFM, ESTARFM, and Random Forest models) for the middle reaches of the Heihe River Basin. Compared with MODIS and original FY-4A LST images at the corresponding time, in general, these three models successfully reproduced the spatial patterns of the FY-4A data in detail.

For the FY-4A LST images, there are some invalid pixels left in the LST images for some time due to cloud contamination. Missing pixels of FY-4A LST images are often randomly distributed due to cloud contamination and other atmospheric conditions. Therefore, FY-4A LST itself has limitations in reflecting temporal and spatial variations in LST. Especially, the weight function-based fusion methods are affected by the reference images, resulting in spatial incompleteness and temporal discontinuity of the predicted LST.

For ESTARFM, there are some null values in the downscaling results, because the spatiotemporal fusion methods like ESTARFM are affected by the MODIS reference images and original FY-4A images (Fig. 3). Similar pixels screening criteria based on window matching make invalid pixels affect other pixels in the window, which results in the spatial incompleteness and temporal discontinuity of downscaled LST. Compared with ESTARFM, BDSTFM not only accurately reflects the temporal variations in LST but also provides complete spatial details. As shown in Fig. 3, there were many invalid values in FY-4A LST images at 13:40, but BDSTFM successfully predicted the MODIS-like LST images, while ESTARFM did not.

In terms of the Random Forest model, although the reconstruction of invalid values is complete, this method ignores the spatial continuity of LST, which results in a low smoothness of the downscaling results and many outliers. And the Random Forest model requires secondary data, in this study, we use reflectance of the red band, green band, blue band, NIR band and SWIR band, the quality of secondary data will also influence experimental results. Compared with the Random Forest model, the spatial patterns of LST predicted with BDSTFM are highly consistent with the near time, reflecting the excellent performance of the proposed approach.

Fig. 4 shows the hourly downscaling effect of the BDSTFM method from 8:00 on September 26 to 6:00 on September 27, 2019 (Beijing time). As shown in the figure, LST in the study area reaches its highest value around 14:00. From 0:00 to 8:00, LST is the lowest period of the day, and LST changes little, which is consistent with the change law of LST. LST is relatively stable from night to early morning. From the downscaling results, the hourly surface temperature reconstructed by BDSTFM conforms to the diurnal variation law of LST, and the spatial reconstruction is basically seamless with strong continuity.
Fig. 3. Diurnal variation in the FY-4A LST images and LST images predicted with the ESTARFM, BDSTFM, and Random Forest models (both at a 1 km × 1 km spatial resolution) on September 25 and 26, 2019 (Beijing Time). Images on the other six dates are not shown here due to length constraints. (a) FY-4A LST images. (b) MODIS LST images. (c) LST images predicted by Random Forest Model. (d) LST images predicted by ESTARFM. (e) LST images predicted by BDSTFM.

C. Validation Against Site Data

Fig. 5 compares the FY-4A LST, diurnal MODIS-like LST data predicted with BDSTFM, ESTARFM, Random Forest, and 4-parameter DTC model [1] compared with the in situ observations at the six weather stations from September 24th to September 30th, 2019. Four methods all complete the prediction of LST at the sites, but LST predicted by the Random Forest model is higher than actual at Daman station. On the contrary, the results of the Random Forest model are lower than the actual at the Desert station (Fig. 5). In terms of verification, LST predicted with the BDSTFM has the smallest deviation from site data, especially at Daman, Desert, and Dashalong stations. Affected by invalid values, results available for testing of ESTARFM are less than the other three models, especially at the Sidaqiao station, which could ensure more accurate results.

Based on the verification results of six stations, the downscaling results of BDSTFM are better than the other three models, probably due to its consistent DTC modeling.

Furthermore, the LST sample numbers predicted with the BDSTFM and ESTARFM models for validation are not equal, with much fewer LST sample number predicted with the ESTARFM model than that predicted with the BDSTFM model due to the effects of invalid values. Although FY-4A LST data was not available on September 29, the BDSTFM model still successfully predicted the MODIS-like LSTs on that day. Besides, missing pixels of MODIS LST images are often randomly distributed due to atmospheric conditions. When ESTARFM
model is predicted in LST, the target time of the target pixel and target day should be valid data, and only complete image pairs of high and low spatial resolution involved in the operation can successfully reconstruct the target pixel. Therefore, the missing target time data or incomplete image pairs involved in the operation will lead to the failure of LST reconstruction. Compared with ESTARFM, BDSTFM can not only realize LST downscaling but also reconstruct the invalid pixels caused by atmospheric effect and terrain effect.

To further evaluate the downscaling performance of BDSTFM, ESTARFM, Random Forest, and 4-parameter DTC model in all sites, the LSTs predicted by the four models at the same time were selected and compared. Table IV shows the accuracy of LST predicted by BDSTFM, ESTARFM, Random Forest, and 4-parameter DTC model compared with the in situ observations over six weather stations on September 24th to September 30th, 2019. Overall, the MODIS-like LSTs are all recovered by four models, where BDSTFM shows better agreement with the in situ data (1.34 K < RMSE < 3.04 K, average 1.93 K) than ESTARFM (1.55 K < RMSE < 3.39 K, average 2.31 K), Random Forest (2.63 K < RMSE < 3.56 K, average 3.24 K) and 4-parameter DTC model (2.50 K < RMSE < 4.06 K, average 3.03 K). In particular, at the stations Desert, the RMSE of BDSTFM is lower than ESTARFM by more than 0.9 K, lower than Random Forest by over 1.9 K, and lower than 4-parameter DTC model by over 1.3 K.

Compared with the accuracy of MODIS product itself shown in Table III, BDSTFM is more similar. The proposed method achieved a higher downscaling accuracy than that of ESTARFM, Random Forest, and 4-parameter DTC model, especially at the Daman and Zhangye stations.

V. MODEL ANALYSIS AND DISCUSSION

A. Analysis of the Model Performance Via Simulation Data

In this section, simulation data are used to analyze the model performance. The generation method of simulation data is...
Fig. 5. Hourly LST values predicted with BDSTFM, ESTARFM, Random Forest, and 4-parameter DTC model compared with the in situ observations over six weather stations from September 24 to 30, 2019.
as follows: first, the 4-km-resolution FY-4A LST data (high-resolution data) were downscaled into 16-km-resolution data (coarse-resolution data) via the aggregate average method; then, the DTC model was used to reconstruct the obtained LST dataset to generate a new LST dataset at the corresponding time.

The 4-km-resolution original FY-4A LST data were also predicted with the DTC model to obtain a new LST dataset at the corresponding time. The data at the four MODIS overpass times corresponding to this dataset (the overpass times in the study area are approximately 4:00, 6:00, 15:00, and 19:00) were input as high-spatial resolution data, and the data at other times were used as high-resolution verification data.

1) Influence of Extending High-Spatial Resolution Dataset: When the MODIS dataset was expanded in the initial downscaled of the LST dataset (Section III-A), BDSTFM model attained a high robustness. To verify this, we select a dataset with a period of September 24 to September 30, 2019 (DOY267-273) to participate in the verification. Each pixel has 7 dates of high-resolution data to participate in the calculation of the deviation coefficient. For 16 km resolution images (coarse-resolution data), there are almost no pure pixels, therefore, two pixels in the simulated image (labeled pixels A and B) were selected to evaluate the results. By comparing the differences between 16 km and 4 km resolution LST time series datasets, it can be found that the differences of different pixels are different. As seen in Table V, pixels A and B are pixel types with large and small differences in LST time series, RMSE of pixels A and B are 1.29 K and 4.01 K, respectively.

Table V shows the accuracy of the 4 km LST data predicted by two typical pixels, the downscaling accuracy of the LST dataset is significantly improved by the high-resolution dataset is expanded. For pixel A, the MAE is reduced by 0.04 K after

| Site name | ME(K) | MAE(K) |
|-----------|-------|--------|
|           | BDSTFM | ESTARFM | RF | 4-para DTC | BDSTFM | ESTARFM | RF | 4-para DTC |
| Arou      | -1.65  | -1.64   | -1.59 | -1.09     | 2.20    | 2.65    | 2.69 | 2.47     |
| Dashalong | -0.53  | 0.31    | -1.26 | -3.26     | 1.48    | 1.61    | 2.22 | 3.28     |
| Desert    | -0.15  | 0.55    | 2.43  | 2.33      | 1.02    | 1.58    | 2.63 | 2.31     |
| Sidaqiao  | -1.01  | -0.54   | 0.10  | 0.01      | 1.48    | 1.62    | 2.22 | 1.82     |
| Daman     | -0.45  | 0.49    | -2.85 | -1.21     | 1.37    | 1.52    | 3.14 | 2.01     |
| Zhangye   | -0.21  | 1.02    | -1.84 | -1.86     | 1.32    | 1.69    | 2.55 | 2.37     |

| Site name | RMSE(K) | r     |
|-----------|---------|-------|
|           | BDSTFM | ESTARFM | RF | 4-para DTC | BDSTFM | ESTARFM | RF | 4-para DTC |
| Arou      | 3.04    | 3.39   | 3.38 | 3.23      | 0.982  | 0.979   | 0.955 | 0.970    |
| Dashalong | 1.90    | 2.13   | 2.63 | 4.06      | 0.981  | 0.974   | 0.971 | 0.973    |
| Desert    | 1.34    | 2.25   | 3.31 | 2.70      | 0.990  | 0.973   | 0.960 | 0.984    |
| Sidaqiao  | 1.89    | 1.55   | 3.10 | 2.50      | 0.989  | 0.984   | 0.963 | 0.975    |
| Daman     | 1.71    | 2.21   | 3.48 | 2.62      | 0.989  | 0.970   | 0.952 | 0.971    |
| Zhangye   | 1.70    | 2.36   | 3.56 | 3.10      | 0.977  | 0.967   | 0.950 | 0.969    |

Table IV

**TABLE IV**

**COMPARISONS OF STATISTICAL METRICS BETWEEN BDSTFM, ESTARFM, RANDOM FOREST, AND 4-PARAMETER DTC MODEL AGAINST IN SITU LST MEASUREMENTS ON THE SAME DAYS AT SIX FLUX TOWERS**

| Site name | Site name | ME(K) | MAE(K) |
|-----------|-----------|-------|--------|
|           | BDSTFM    | ESTARFM | RF | 4-para DTC | BDSTFM | ESTARFM | RF | 4-para DTC |
| Arou      | -1.65     | -1.64   | -1.59 | -1.09     | 2.20    | 2.65    | 2.69 | 2.47     |
| Dashalong | -0.53     | 0.31    | -1.26 | -3.26     | 1.48    | 1.61    | 2.22 | 3.28     |
| Desert    | -0.15     | 0.55    | 2.43  | 2.33      | 1.02    | 1.58    | 2.63 | 2.31     |
| Sidaqiao  | -1.01     | -0.54   | 0.10  | 0.01      | 1.48    | 1.62    | 2.22 | 1.82     |
| Daman     | -0.45     | 0.49    | -2.85 | -1.21     | 1.37    | 1.52    | 3.14 | 2.01     |
| Zhangye   | -0.21     | 1.02    | -1.84 | -1.86     | 1.32    | 1.69    | 2.55 | 2.37     |

**TABLE V**

**INFLUENCE OF EXTENDING LOW SPATIAL RESOLUTION DATASET ON RECONSTRUCTION ACCURACY**

| Pixel | Type                  | MAE(K) | RMSE(K) | r     |
|-------|-----------------------|--------|---------|-------|
|       |                       |        |         |       |
| Arou  | Coarse-resolution data| 1.19   | 1.29    | 0.999 |
|       | Result of not extending data | 0.24 | 0.36 | 0.999 |
|       | Result of extending data | 0.28 | 0.38 | 0.999 |
| B     | Coarse-resolution data| 3.77   | 4.01    | 0.993 |
|       | Result of no extending data | 1.68 | 2.49 | 0.977 |
|       | Result of extending data | 0.77 | 1.19 | 0.994 |
expansion, and the RMSE is reduced by 0.02 K; for pixel B, the MAE is reduced by 0.91 K after expansion, and the RMSE is reduced by 1.30 K. By expanding the high-resolution LST dataset, the reconstruction accuracy of pixels with large temporal differences was significantly improved, while the impact on pixels with small temporal differences was limited. The effectiveness of extending a high-spatial resolution dataset was demonstrated.

Fig. 6 shows the effect of expanding the high-spatial-resolution LST dataset on the reconstruction results. The results show that, for pixel A, the accuracy did not change significantly by expanding the dataset, for pixel B, the downscaling accuracy has been significantly improved at other times far away from $T_1$, $T_2$, $T_3$, and $T_4$ (this dataset corresponds to 4:00, 6:00, 15:00, and 19:00) by expanding the dataset. Therefore, by expanding the effective time of the high-resolution dataset, the stability of the model is improved, so that the downscaling results always exhibited a higher accuracy.

2) Requirements for the Amount of LST Observations: To further investigate when the combined method performs well, we assessed the relationship between the number of days of high-resolution data involved in regression fitting during deviation coefficient calculation and the model downscaling accuracy during preliminary downscaling of the LST dataset (Section III-A). Table VI summarizes the specific dates corresponding to the different numbers of dates and the accuracy of the predicted data on DOY270.

Table VI presents the influence of the number of dates involved in the fitting process (3, 5, 7, and 9 days) on the downscaling effect of LST data. For pixel A, the RMSE value of the LST with a low spatial resolution was 1.35 K, and the RMSE value of the LST dataset obtained via fitting and downscaling on the different days ranged from 0.38 to 0.42 K; for pixel B, the RMSE value of the low-spatial resolution LST dataset was 3.87 K, and the RMSE value of the LST dataset obtained via fitting and downscaling on the different days ranged from 0.92 to 1.09 K. As the number of dates involved in the fitting process varied, the magnitude of the downscaling accuracy RMSE remained relatively stable. This indicates that the hypothesis proposed in this study is reasonable; for the same pixel, the DTC models corresponding to the LST datasets of different spatial resolutions differed, but this difference remained similar over a certain period.

The Levenberg–Marquardt method was used to calculate the coefficients of deviation, which is a nonlinear least squares regression method. When the amount of data participating in the fitting process is small, the error in part of the data can lead to a large error in the calculated coefficients of deviation, which can result in errors in the initial reconstruction of the LST data. To reduce the influence of data errors on the reconstruction results, more dates are required to participate in the fitting. In addition, the simulated data pertain to a period of continuous clear-sky conditions, but it is often difficult to achieve continuous clear-sky conditions in real data. Based on the verification results of the
TABLE VI
INFLUENCE OF THE NUMBER OF DATES INVOLVED IN FITTING ON THE PRELIMINARY RECONSTRUCTION EFFECT OF LST DATA

| Pixel  | Data      | MAE(K) | RMSE(K) | r  |
|--------|-----------|--------|---------|----|
|        | Coarse resolution data | 1.25   | 1.35    | 0.999 |
|        | DOY269-271 | 0.29   | 0.42    | 0.999 |
| Pixel A| DOY268-272 | 0.30   | 0.40    | 0.999 |
|        | DOY267-273 | 0.32   | 0.39    | 0.999 |
|        | DOY266-274 | 0.31   | 0.38    | 0.999 |
|        | Coarse resolution data | 3.71   | 3.87    | 0.997 |
| Pixel B| DOY269-271 | 0.64   | 1.09    | 0.996 |
|        | DOY268-272 | 0.55   | 0.98    | 0.996 |
|        | DOY267-273 | 0.52   | 0.92    | 0.997 |
|        | DOY266-274 | 0.57   | 1.01    | 0.996 |

Fig. 7. Experimental results of simulation data in arid area and humid area are verified. (a) and (b) are the distribution of two accuracy verification indicators, respectively. (a) MAE distribution of the two regions. (b) RMSE distribution of the two regions.

Simulation data for pixels A and B, the number of fitting days involved in the regression process in the subsequent experiments was set to 7 days.

3) Applicability of the Model to Different Climate Types: To verify the downscaling ability of the model in regions of different climate types, a simulation data experiment in Jiangsu Province was included.

The method of obtaining the simulation data is described in Section V-A. The climate adaptability of the model can be verified and the overall downscaling accuracy can be evaluated by comparing the reconstruction effects of the simulated data in these two regions.

After downscaling, the simulated data for the two regions to a spatial resolution of 4 km using the BDSTFM model, the MAE value of each pixel was calculated with the real 4-km spatial resolution LST data, the RMSE value was obtained, and MAE and RMSE distributions were plotted, as shown in Fig. 7. Based on the distribution map, the MAE and RMSE distributions of the LST downscaling results were relatively uniform, and the distribution difference between the arid and humid areas was small. Except for abnormal errors caused by a poor data quality, the errors in most pixels remained within a relatively small range. The Heihe region includes part of the Qilian Mountains, which are covered by ice and snow year round, so the observation error of the surface temperature can be large; the diurnal variation in LST in the glacial region differs from that in other regions, the fitting error of the DTC model in this region is relatively large, and the downscaling accuracy of the BDSTFM model decreases, which explains the large error in continuity observed in Fig. 7. In Jiangsu Province, there were only a few points with large errors, which may be attributed to the poor data quality of these points, resulting in abnormal downscaling results.

The spatial distribution of errors can only reflect the model performance under different climate types, and the statistics of errors can mathematically represent the pros and cons of the downscaling results more intuitively. The MAE and RMSE
values in the two regions were calculated separately, as shown in Fig. 8. Based on the statistical results, the error distributions of these two climate types exhibited strong similarities. The specific performance indicated that the MAE value of each pixel largely varied between 0 and 2 (K); pixels with RMSE varying between 1.5 and 2 (K) were the most abundant. In terms of the spatial distribution and statistical trend of comprehensive errors, the BDSTFM model achieved similar performance levels in typical arid and humid regions, and there occurred no significant difference. Therefore, the model attained a suitable applicability in regions with dry and humid climate types.

B. Performance of BDSTFM Under Cloudy Skies

When BDSTFM reconstructs invalid pixels, it considers the temporal and spatial continuity of LST data, and the temporal continuity is reflected by the DTC models. But, DTC models are applicable only under clear skies, when pixels are missing due to cloud occlusion, the LST reconstructed by the BDSTFM may be different from the real LST. Due to the weather, the FY-4A images on September 29 were missing nearly half of the time, and the BDSTFM successfully reconstructed the invalid pixels at these times. As a comparison, the comparison results of the two sites are shown in Fig. 9. From the comparison results, although BDSTFM successfully reconstructed the LST data at the missing moment, there is a deviation from the actual station measurements. Compared with the downscaling results under clear sky conditions, the RMSE of the desert station is increased by 1.44 K, and the RMSE of the Sidaqiao station is increased by 1.09 K.

From the verification results, BDSTFM can realize the reconstruction of missing pixels under partial cloud conditions, but there is a certain deviation between the reconstruction results and the actual LST, which is probably because the LST reconstructed with the help of the DTC model cannot reflect the LST under the real cloudy skies.

C. Advantages of BDSTFM

The results demonstrated that the BDSTFM model attained a satisfactory spatial applicability and generalization ability via simulation data, real satellite image data, and weather station data for testing. Compared with the ESTARFM model, the BDSTFM model achieved a higher accuracy, yielded more spatial details, and exhibited obvious temporal evolution. Especially when there were invalid pixels, the model could successfully obtain a high-resolution LST dataset continuous in both time and space. Compared with the Random Forest model, the LST values predicted with the BDSTFM model were more continuous in space and exhibited a higher accuracy. Although the Random Forest model uses statistical relationships to obtain seamless LST data, there still occurred much abnormal noise in the downscaling results.
Based on theories and methods such as the DTC model and data fusion method, the performance of the BDSTFM model was comprehensively improved. Specifically, this model mainly includes the following three aspects:

First, by using the DTC model, the BDSTFM model fully utilizes high-frequency information of low-resolution data, and the predicted high-resolution data exhibit spatiotemporal continuity under clear-sky conditions and a high downscaling accuracy. The BDSTFM model yields the DTC model of the high-resolution LST dataset. Therefore, high-resolution LST data can be finally obtained at any time instead of being limited to data corresponding to the considered low-resolution data. This is fundamentally different from the existing spatiotemporal fusion models, such as the ESTARFM model. The downscaling results of the classical ESTARFM model rely heavily on spatiotemporal continuity in the low-resolution dataset, and high-resolution data cannot be achieved in areas containing invalid pixels. In addition, existing spatiotemporal fusion models downscale low-resolution data separately at each moment. Although high-frequency information of these low-spatial resolution data is used, the change trend of high-frequency information (such as LST continuously changing over time) is not fully utilized, resulting in unsatisfactory downscaling results at certain moments. The BDSTFM model fully utilizes the temporal tendency revealed by the high-frequency and low-resolution LST data and describes this trend with the DTC model, which ensures more accurate downscaling results.

Second, the use of high-spatial-resolution data at multiple times improves the spatial applicability of the model and the downscaling stability and accuracy at each time. In contrast to existing spatiotemporal fusion models, which can only use high-resolution LST at one or two times during the downscaling of low-resolution LST data, the BDSTFM model uses high-resolution LST data on multiple days and multiple times. In addition, existing spatiotemporal fusion models downscaling models exhibit requirements for the time of the reference image, i.e., the times of reference images with different spatial resolutions must be the same (usually realized by normalizing the time of the low-spatial resolution data), but the BDSTFM model does not require the same time of the high-spatial resolution images, which allows high-resolution data at any time to participate in downscaling. In this study, the downscaling accuracy of the BDSTFM model at each time was improved by extending the existing time of MODIS because this study only used one high-resolution LST data source. In fact, the BDSTFM model could perform well considering any high-spatial resolution data.

Third, by adding an invalid pixel reconstruction method considering LST spatiotemporal continuity, the BDSTFM model can obtain continuous and 1-km-resolution LST data at hourly intervals under clear-sky conditions. The quality of the dataset could be optimized mainly by using neighborhood information to reconstruct invalid pixels and by linearly interpolating regression residuals to enhance the downscaling accuracy. The change rule of LST is continuous in both time and space; with the use of this rule, missing pixels can be predicted and substituted by combining temporal and spatial neighborhood information components to achieve seamless downscaling of the LST. The traditional spatiotemporal fusion model can only calculate valid data but cannot predict missing data, so there are many gaps in downscaled high-resolution LST data, which cannot be seamlessly applied. Although the statistical downscaling method can achieve an excellent performance, a purely statistical relationship cannot describe the continuity in the spatial distribution of the LST data. The BDSTFM model not only realizes the reconstruction of invalid values but also considers the spatiotemporal characteristics of LST data, which benefits from the invalid pixel reconstruction method in the model.

D. Prospects and Future Issues of BDSTFM

The BDSTFM model proposed in this study obtains a high-resolution LST dataset under clear-sky conditions. This method first establishes a DCT model using a high-temporal-resolution and low-spatial-resolution LST dataset and then uses high-spatial-resolution data at multiple times to correct the model to generate a high-spatial-resolution and high-temporal-resolution LST dataset. This study generates LST datasets at a 1-km spatial resolution at hourly intervals by using only MODIS and FY-4A LST data and can use higher-spatial-resolution LST datasets retrieved from other sources to obtain spatiotemporally continuous LST datasets with a higher spatial resolution.

VI. Conclusion

This research proposed a spatiotemporal downscaling model of LST, namely the BDSTFM model, which can obtain preliminary downscaling results by using high-resolution MODIS LST data on multiple days and multiple times to correct the DTC model established based on FY-4A LST data. Then, the model fully utilizes information of the spatiotemporal neighborhood to obtain a continuous LST dataset with a high spatiotemporal resolution.

The BDSTFM model was assessed against simulated data, real satellite image data, and site measurement data and compared with the commonly used spatiotemporal downscaling model (ESTARFM), machine learning method (Random Forest), and temporal interpolation method (4-parameter DTC model). The results indicated that the BDSTFM model can not only successfully obtain LST data with a 1-km resolution at hourly intervals under clear-sky conditions but can also attain a higher downscaling accuracy, which is consistent with the accuracy of MODIS products. The simulation data test revealed that the BDSTFM model achieved a suitable stability, was less affected by climatic conditions and was applicable under both arid and humid climate types.

The BDSTFM model was proposed to compensate for the shortcomings of existing downsampling methods and to produce continuous LST datasets with high spatial and temporal resolutions to provide richer LST datasets for research on the UHI effect. Future work will probe the influence of short-term meteorological fluctuations on the downscaled LST dataset and further improve the model downscaling accuracy under complex weather conditions.
[43] W. Zhao, H. Wu, G. Yin, and S.-B. Duan, “Normalization of the temporal effect on the MODIS land surface temperature product using random forest regression,” *ISPRS J. Photogrammetry Remote Sens.*, vol. 152, pp. 109–118, 2019.

[44] W. Zhao and S.-B. Duan, “Reconstruction of daytime land surface temperatures under cloud-covered conditions using integrated MODIS/Terra land products and MSG geostationary satellite data,” *Remote Sens. Environ.*, vol. 247, 2020, Art. no. 112931.

[45] X. Zheng, Z. L. Li, T. Wang, H. Huang, and F. Nerry, “Determination of global land surface temperature using data from only five selected thermal infrared channels: Method extension and accuracy assessment,” *Remote Sens. Environ.*, vol. 268, 2022, Art. no. 112774.

[46] Y. Liao et al., “Surface urban heat island detected by all-weather satellite land surface temperature,” *Sci. Total Environ.*, vol. 811, Mar. 2022, Art. no. 151405.

[47] Z. Xing et al., “Estimation of daily mean land surface temperature at global scale using pairs of daytime and nighttime MODIS instantaneous observations,” *ISPRS J. Photogrammetry Remote Sens.*, vol. 178, pp. 51–67, 2021.

[48] Z. Zhang and Q. Du, “Hourly mapping of surface air temperature by blending geostationary datasets from the two-satellite system of GOES-R series,” *ISPRS J. Photogrammetry Remote Sens.*, vol. 183, pp. 111–128, 2022.

**Penghua Hu** received the B.S. degree in surveying and mapping engineering in 2021 from Hohai University, Nanjing, China, where he is currently working toward the master’s degree in surveying and mapping science and technology with the School of Earth Science and Engineering. His main research interest is thermal infrared remote sensing data fusion.

**Aihui Wang** received the M.S. degree in surveying and mapping science and technology from Hohai University, Nanjing, China. He is currently working with China Railway Design Corporation, Tianjin, China.

**Yingbao Yang** received the Ph.D. degree in cartography and geographical information system from the Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing, China, in 2005. She was a Visiting Scholar with Michigan State University, East Lansing, MI, USA, in 2013. She is currently a Professor with the School of Earth Science and Engineering, Hohai University, Nanjing, China. Her main research interests include thermal infrared remote sensing, processing of remote sensing image data, remote sensing of surface water, and thermal environment.

**Xin Pan** received the Ph.D. degree in cartography and geographical information system from the Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing, China, in 2016. He is currently an Associate Professor with Hohai University, Nanjing, China. His research interests include remote-sensing retrieval and validation of land surface evapotranspiration and net radiation.

**Xiejunde Hu** received the M.S. degree in surveying and mapping science and technology from Hohai University, Nanjing, China, in 2022. His research interests include land surface temperature downscaling.

**Yuncheng Chen** received the B.S. degree in surveying and mapping engineering from the Shandong University of Science and Technology, Taian, China, in 2016. He is currently working toward the master’s degree in surveying and mapping science and technology with the School of Earth Science and Engineering, Hohai University, Nanjing, China. His main research interest is spatiotemporal fusion network for land surface temperature based on deep learning.

**Xuechun Kong**, photograph and biography not available at the time of publication.

**Yao Bao**, photograph and biography not available at the time of publication.

**Xiangjin Meng**, photograph and biography not available at the time of publication.

**Yang Dai**, photograph and biography not available at the time of publication.