A Comprehensive Flexible Spatiotemporal DAta Fusion Method (CFSDAF) for Generating High Spatiotemporal Resolution Land Surface Temperature in Urban Area

Chenlie Shi, Ninglian Wang, Quan Zhang, Zhuang Liu, and Xinming Zhu

***Abstract*—Spatiotemporal fusion of land surface temperature (LST) has a vital significance in studying the temporal and spatial variation of urban heat islands. But most existing LST fusion methods do not consider the highly heterogeneous urban surface and complexity of the spatial layout. In this study, a Comprehensive Flexible Spatiotemporal DAta Fusion (CFSDAF) method was proposed to generate a high spatiotemporal resolution urban LST image, which was an improvement of the Flexible Spatiotemporal DAta Fusion (FSDAF). The CFSDAF first adjusted the differences between coarse-resolution LST and fine-resolution LST. Then, the visible and near-infrared image of a fine resolution was introduced to execute spectral unmixing and to conduct soft classification, which considered the mixed pixel of fine-resolution LST. The inverse distance weighting (IDW) interpolation was used in improving the computational efficiency, and the constrained least square was selected to better distribute the residual. The performance of CFSDAF was compared with the temporal adaptive reflectance fusion model (STARFM) and FSDAF. The results indicate that the predicted images by CFSDAF are better than STARFM and FSDAF from both visual comparison and quantitative assessment in two experiments, and CFSDAF can reserve more spatial details and accurately restore the spatial continuity of urban LST than others. Moreover, CFSDAF has high computational efficiency and can monitor land cover changes the same as FSDAF. Due to the above advantages, the CFSDAF has great potential for studying spatiotemporal changes of LST and UHI in an urban area.**

***Index Terms*—Comprehensive flexible spatiotemporal data fusion (CFSDAF), flexible spatiotemporal data fusion (FSDAF), landsat, land surface temperature (LST), MODIS, spatiotemporal fusion.**

I. Introduction

OVER the last few decades, the world has experienced rapid urbanization, and more than 55.2% of the populations live in cities up to 2018 [1]. It is reported that the number of urban populations will expand by 68% of the total population around the world by 2050 [2]. In the meanwhile, the gathering of many people in the city caused a series of negative effects. The urban heat island (UHI) effect refers to a phenomenon that the temperature of the city is higher than the temperature in the suburbs [3], [4] and is one of the most severe urban problems [5], [6], [7]. People are exposed to the environment of the city for a long time, which can affect citizens' health and well-being. Accordingly, accurately monitoring the variation of the UHI effect in spatial and temporal scales has great significance for mitigating UHI and projecting the spatial layout of the city for policymakers [8], [9].

Land surface temperature (LST) retrieved from a remotely sensed image is regarded as effective and convenient data for monitoring the UHI [10], [11], [12], [13]. In many cases, the LST image with the high spatiotemporal resolution is indispensable for studying the UHI effect because of the high heterogeneity in an urban area and spatiotemporal variability of the urban thermal environment [14], [15]. However, due to the tradeoff of sensors between spatial resolution and temporal resolution, there is no single sensor that can simultaneously generate the LST image with both high temporal and high spatial resolution [16], [17]. For example, the spatial resolution of the Landsat 8 LST image is 30 m (resampled from 100 m to 30 m by NASA) but it has a long coverage period of 16 days [18], and the MODIS LST image can be obtained every day but the spatial resolution of 1000 m [19]. Besides, it is rare to obtain a cloudless Landsat LST image

Manuscript received 23 June 2022; revised 14 October 2022; accepted 31 October 2022. Date of publication 11 November 2022; date of current version 23 November 2022. This work was supported in part by the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant XDA19070302, in part by the National Natural Science Foundation of China under Grant 42130516, and in part by the Second Tibetan Plateau Scientific Expedition and Research Program under Grant 2019QZKK020102. *(Corresponding author: Ninglian Wang.)*

Chenlie Shi and Quan Zhang are with the Shaanxi Key Laboratory of Earth Surface System and Environmental Carrying Capacity, College of Urban and Environmental Sciences, Northwest University, Xi'an 710127, China, and also with the Institute of Earth Surface System and Hazards, College of Urban and Environmental Sciences, Northwest University, Xi'an 710127, China (e-mail: max1995@stumail.nwu.edu.cn; zhangquanzq@nwu.edu.cn).

Ninglian Wang is with the Shaanxi Key Laboratory of Earth Surface System and Environmental Carrying Capacity, College of Urban and Environmental Sciences, Northwest University, Xi'an 710127, China, also with the Institute of Earth Surface System and Hazards, College of Urban and Environmental Sciences, Northwest University, Xi'an 710127, China, and also with the Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China (e-mail: nlwang@nwu.edu.cn).

Zhuang Liu is with the State Key Laboratory of Geo-Information Engineering, Xi'an 710127, China, and also with the Xi'an Research Institute of Surveying and Mapping, Xi'an 710127, China (e-mail: 1065661608@qq.com).

Xinming Zhu is with the College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: zhuxinming19@mails.ucas.ac.cn).

Digital Object Identifier 10.1109/JSTARS.2022.3220897

within one year because of the influence of clouds [20], [21], especially in low-latitude areas. Regarding the above issues, the scholars around the world have proposed many effective ways including downscaling and spatiotemporal fusion of the LST image in the past 20 years to generate the high spatiotemporal resolution LST image [14], [15], [22], [23], [24], [25].

Downscaling of an LST image from low spatial resolution (hereafter referred to "coarse-resolution") to high spatial resolution (hereafter "fine-resolution") is commonly used to provide high spatiotemporal LST image, which is also named thermal disaggregation or sharpening [26], [27], [28]. The main idea for LST downscaling supposes that the scale effect between coarse-resolution LST (e.g., MODIS LST image) and fine-resolution LST (e.g., Landsat LST image) is negligible [29], [30] and then utilizing the correlations between LST with the coarse resolution and visible–near-infrared image with the coarse resolution to generate a downscaled LST image with fine resolution. The accuracy of the downscaled LST image was mainly affected by the downscaling methods (also called regression tools) and auxiliary data (also called regression factors) [27], [31], and most of the LST downscaling studies are improvements of the above two aspects in the past [26], [30], [32], [33], [34], [35], [36]. However, previous studies on LST downscaling were mainly applied to nonurban areas, and few reports for urban LST studies where there are mainly three reasons. First, the spatial resolution of the downscaled LST image is still too coarse (e.g., downscaling MODIS 1000 m LST to 250 m LST) for studying the UHIs because the auxiliary image and the LST image are obtained from the same satellite [37]. Second, the spatial distributions of urban LST are largely influenced by human activities, and only the auxiliary image from the visible and near-infrared image is not sufficient for LST downscaling in urban areas. Third, the assumption of invariant scale is not suitable for the high heterogeneous region, and the errors of downscaled LST will be greater when the resolution ratio from low-resolution LST to downscaled LST is too large [38]. Therefore, the LST downscaling has some limitations for generating high spatiotemporal LST in urban areas.

Spatiotemporal fusion, another method that can generate high spatiotemporal resolution LST, has been rapidly developed in the past ten years [24], [25], [39], [40]. The core idea of spatiotemporal fusion is to fully utilize the spatial and temporal information of input images with different resolutions to generate an image with both high spatial and temporal resolution, which was early developed for reflectance fusion [17], [25], [41], [42], [43], [44], [45], [46]. The STARFM proposed by Gao et al. [25] was the first spatiotemporal fusion method, which predicted the Landsat-like reflectance image by a weighted function with the information of neighborhood. To improve the prediction accuracy of STARFM in heterogeneous landscapes and disturbance events, the enhanced STARFM (ESTARFM) and a spatial–temporal adaptive algorithm for mapping reflectance change were developed [47]. At present, the improvements in spatiotemporal fusion method for surface reflectance mainly focus on preserving spatial details, monitoring land cover change events, and improving calculation efficiency [48], [49], [50]. Although the original fusion methods were proposed for reflectance, some scholars introduced the spatiotemporal fusion method for LST fusion [51], [52], [53], [54], [55]. Liu et al [56] directly applied the STARFM to fusion MODIS LST and ASTER LST to generate the ASTER-like LST image. Wu et al. [39] improved the STARFM for LST fusion that considered the sensor observation differences in different land cover types. A new fusion method based on bilateral filtering for urban LST fusion was proposed [24], which accounted for the interaction of LST at the different surface boundary. Weng et al. [57] presented the Spatiotemporal Adaptive Data Fusion Algorithm for Temperature mapping (SADFAT) by introducing the annual temperature cycle (ATC) for STARFM, which was assessed by blending MODIS LST and Landsat LST in heterogeneous areas. In addition, the spatiotemporal fusion methods for LST that integrated geostationary satellite and Polar orbiting satellite were proposed to create hourly Landsat-like LST data recently [58], [59]. It is noteworthy that fusing microwave LST and thermal infrared LST to generate all-weather and seamless LST products, which can promote spatiotemporal fusion of LST for global expansion, and it is a hot topic in recent years as well [60], [61].

Although spatiotemporal fusion methods for LST have been developed in recent years, most of which cannot restore the spatial details of LST, cannot predict the abrupt event, and cannot reserve the spatial continuous of LST in an urban area simultaneously [62]. Furthermore, due to the high heterogeneity of urban areas, the proposed fusion methods for LST were rarely applied and tested for the urban LST study in the past. The Flexible Spatiotemporal DAta Fusion (FSDAF) method was proposed by Zhu et al. [63], which can simultaneously predict both gradual change and abrupt change events and became one of the most popular spatiotemporal fusion methods [64], [65], [66]. But in FSDAF, the fine-resolution pixels were regarded as pure pixels and hard classification was executed for fine-resolution data, which cause the spatial discontinuity of LST and obvious boundary lining in the different land cover, and this phenomenon particularly occurred in urban areas. In addition, for large areas or long-term studies, the computational time of FSDAF has exponential growth due to the usage of Thin Plate Spline (TPS) interpolation. Finally, the distribution of residual in FSDAF was empirical, which can be further improved.

In this study, a Comprehensive Flexible Spatiotemporal DAta Fusion (CFSDAF) method was proposed to create a high spatiotemporal resolution LST image in an urban area, which considers the differences in sensors and reserve the spatial continuity and spatial details of LST. Specifically, CFSDAF has the following improvements versus FSDAF:

1) adjust the differences between coarse-resolution LST and fine-resolution LST using a linear model at a coarse-resolution scale;
2) consider mixed pixels of fine-resolution LST due to the high heterogeneity in urban areas by introducing the visible and near-infrared images of a fine resolution, which can reserve spatial continuity of LST in an urban area;
3) employ inverse distance weighting (IDW) interpolation instead of TPS interpolation in FSDAF to improve the computational efficiency;

4) combine temporal increments and spatial increments at a fine-resolution scale through the constrained least squares (CLS) theory to better distribute residuals.

The performance of CFSDAF was tested and evaluated in two urban sites including heterogeneous area and land cover change regions with STARFM and FSDAF, which are popular fusion methods and are expediently executed for free codes.

II. Methodology

A. Some Notations for CSFDAF

1) T_B/T_P : base date T_B and prediction date T_P;
2) $A_F^B(x_{ij}, y_{ij})/A_F^P(x_{ij}, y_{ij})$: abundances of one fine-resolution pixel on T_B or T_P;
3) $R_F^B(x_{ij}, y_{ij}, m)/R_F^P(x_{ij}, y_{ij}, m)$: numerical values of LST for each endmember in one fine-resolution pixel on T_B or T_P;
4) n_m: the number of endmembers;
5) $F_B(x_{ij}, y_{ij})/\ F_P(x_{ij}, y_{ij})$: numerical values of a fine-resolution LST image on T_B or T_P;
6) N: the number of fine-resolution pixels in one coarse-resolution pixel;
7) $(x_i,\ yi)$/(x_{ij}, y_{ij}): coordinate of one coarse-resolution pixel and coordinate of fine-resolution pixels in one coarse-resolution pixel;
8) $\Delta C(x_i, y_i)/\Delta F(x_{ij}, y_{ij})$: Change values for coarse-resolution pixel and fine-resolution pixel from T_B to T_P;
9) $\Delta R_C(x_i, y_i, m)/\Delta R_F(x_{ij}, y_{ij}, m)$: Change values of endmembers for one coarse-resolution pixel and fine-resolution pixels from T_B to T_P;
10) $A_C^B(x_i, y_i)$: abundances of one coarse-resolution pixel on T_B.

B. CFSDAF

The proposed CFSDAF method first adjusts the differences in different resolution LST and then considers mixed pixels of a fine-resolution LST image in an urban area by introducing visible and near-infrared image with fine resolution for performing soft classification to obtain the endmembers' abundances of a fine-resolution image, and the predicted images by CFSDAF reserve more spatial details and spatial continuity of LST. In addition, the CFSDAF requires a pair of coarse-resolution LST image and one fine-resolution LST image on T_B and one coarse-resolution LST image on T_P same as FSDAF, as well as additional fine-resolution visible and near-infrared image on T_B. Abnormal values of coarse-resolution LST and fine-resolution LST images are corrected because the inversion process of LST will bring the inevitable outliers before executing the CFSDAF. The MODIS LST image and Landsat LST image were selected as coarse-resolution LST and fine-resolution LST to test the performance of CFSDAF in this study.

The CFSDAF includes the main six steps as follows:

1) adjust the differences between coarse-resolution LST and fine-resolution LST;
2) extract the endmembers and perform soft classification for fine-resolution images;

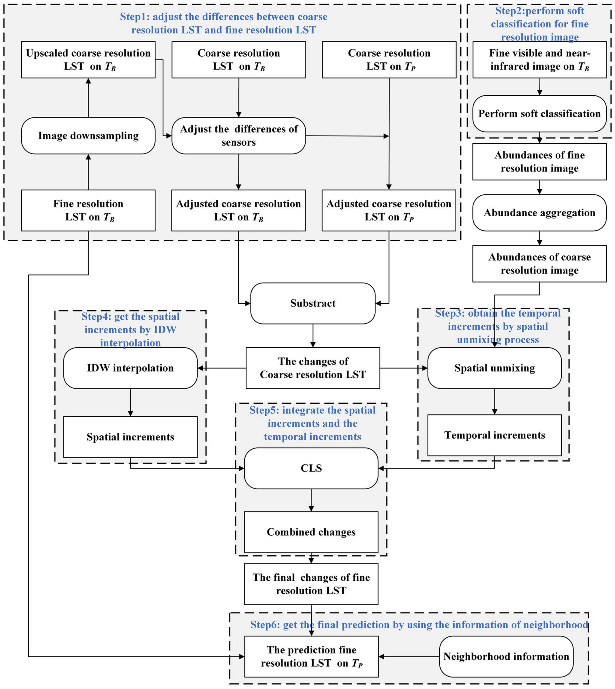

Step1: adjust the differences between coarse resolution LST and fine resolution LST
Upscaled coarse resolution LST on T_B
Coarse resolution LST on T_B
Coarse resolution LST on T_P
Image downsampling
Adjust the differences of sensors
Fine resolution LST on T_B
Adjusted coarse resolution LST on T_B
Adjusted coarse resolution LST on T_P
Step2:perform soft classification for fine resolution image
Fine visible and near-infrared image on T_B
Perform soft classification
Abundances of fine resolution image
Abundance aggregation
Abundances of coarse resolution image
Substract
Step4: get the spatial increments by IDW interpolation
IDW interpolation
Spatial increments
The changes of Coarse resolution LST
Step3: obtain the temporal increments by spatial unmixing process
Spatial unmixing
Temporal increments
Step5: integrate the spatial increments and the temporal increments
CLS
Combined changes
The final changes of fine resolution LST
Step6: get the final prediction by using the information of neighborhood
The prediction fine resolution LST on T_P
Neighborhood information

Fig. 1. Flowchart of the proposed CFSDAF.

3) obtain the temporal increments by a spatial unmixing process;
4) get the spatial increments by IDW interpolation;
5) integrate the spatial increments and temporal increments;
6) get the predicted values by using the information of neighborhood.

The flowchart of the CFSDAF is shown in Fig. 1.

1) Adjust the Differences Between Coarse-Resolution LST and Fine-Resolution LST: In general, the differences in LST from different sensors are not only affected by atmospheric conditions, performance and posture of sensors, and observation angle [17], which are the same as surface reflectance data, but also the transit time of different sensors and the inversion accuracy of LST. For example, the LST image retrieved from the MODIS image via a generalized split-window algorithm in which errors can reach 1K under ideal conditions [67], [68], [69], [70], and the retrieved LST data from Landsat image is based on single channel algorithm in which the errors will be about 1.5K under strict condition control [71], [72], and a recent research finds the retrieved errors from Landsat 5/7/8 LST products can reach 2–3K [73]. Furthermore, the time difference between the LST retrieved from different satellite images is about 30 min. For the above reasons, the differences in LST images in multisources should be corrected, which are the first and crucial step for CFSDAF. In this article, a linear model is proposed to normalize the differences between coarse-resolution LST and fine-resolution LST. Specifically, the fine-resolution LST on T_B upscale to the coarse resolution at first (called upscaled coarse-resolution LST), then establishing a linear relationship between the upscaled LST and the coarse-resolution LST, and last, the linear model is applied to the coarse-resolution LST image to generate the adjusted coarse-resolution LST image on T_B and T_P. The reason why a linear model is utilized is listed in Section V-A.

2) Extract the Endmembers and Perform Soft Classification for Fine-Resolution Image: There has more than one land cover type for one fine-resolution pixel (e.g., Landsat pixel) because of strong heterogeneity in an urban area where LST in different land cover types has disparate values, also called mixed pixel and it is necessary to resolve the phenomenon of mixed pixel for the fine-resolution LST image. Landsat image is regarded as the fine-resolution image in this article because it has more than 40 years of data with fine resolution, which has great significance in studying the long time series changes of UHI. The endmembers and endmembers' abundances extracted from Landsat visible and near-infrared data can be treated as endmembers and endmembers' abundances of LST in one Landsat pixel because the surface reflectance image and LST image have the same spatial resolution (Landsat LST was resampled to 30 m by NASA) and the same acquisition time. To generate the long-time synthetic Landsat-like LST data expediently, a globally representative spectral linear mixture model (SVD model) for Landsat surface reflectance image shared by Sousa et al. [74] was chosen as endmembers for performing soft classification, and the SVD model includes the substrate (S), vegetation (V), and dark surface (D) three types. A fully CLSs (FCLS) [75] method which is a linear model, was selected to apply for the soft classification. The abundances for the Landsat LST image can be calculated and the sum of abundances equal to 1 for one Landsat LST pixel, and the range of abundances is from 0 to 1. For other fine-resolution LST image such as ASTER LST image [76], extraction of endmembers is the first step. After accomplishing the above calculation, the abundances of fine-resolution LST image can be obtained.

3) Obtain the Temporal Increments by Spatial Unmixing Process: According to the linear mixture theory [77], supposing that the LST value in one fine-resolution pixel is a linear mixture with endmembers and endmembers' abundances of the LST image, the values of the fine-resolution LST image $F_T(x_{ij}, y_{ij})$ on T_B and T_P can be expressed as follows:

$$F_T\ (x_{ij}, y_{ij}) = \sum_{n=1}^{n_m} A_F^T\,(x_{ij}, y_{ij}) \times R_F^T\,(x_{ij}, y_{ij}, m) + \varepsilon \quad \text{with } T = B \text{ or } P \tag{1}$$

where $A_F^T(x_{ij}, y_{ij})$ are the abundances of one fine-resolution LST pixel on T_B and T_P. $R_F^T(m)$ store the endmembers values of the fine-resolution LST image on T_B and T_P. n_m and m are the number of endmembers and the mth endmember separately. Supposing no land cover change from T_B to T_P, and the abundances will not change between T_B and T_P, i.e., $A_F^B\,(x_{ij}, y_{ij}) = A_F^P\,(x_{ij}, y_{ij})$ and ε is constant. The changes in the fine-resolution LST image $\Delta F(x_{ij}, y_{ij})$ can be calculated with

$$\Delta F\ (x_{ij}, y_{ij}) = A_F^B\ (x_{ij}, y_{ij}) \times \Delta R_F\,(x_{ij}, y_{ij}, m) \tag{2}$$

where $\Delta R_F(m)$ represent the changes of fine-resolution LST for each endmember. Similarly, the changes in the coarse-resolution LST image $\Delta C(xi, y_i)$ are described with

$$\Delta C\ (x_i, y_i) = A_C^B\ (x_i, y_i) \times \Delta R_C\,(x_i, y_i, m)\,. \tag{3}$$

$A_C^B(x_i, y_i)$ are the abundances of one coarse-resolution LST pixel on T_B and T_P, and $A_C^B(x_i, y_i)$ are obtained by averaging the endmembers' abundances of the fine-resolution LST image in one coarse-resolution LST pixel. But $A_C^B(x_i, y_i)$ in FSDAF are the ratio between the numbers of each class m for fine-resolution LST pixels in one coarse-resolution pixel and the total numbers of fine-resolution LST pixel in one coarse-resolution pixel. Theoretically speaking, if the $\Delta R_C(x_i, y_i, m)$ could be calculated, $\Delta R_F(x_{ij}, y_{ij}, m)$ will be obtained. Therefore, the key issue is to resolve $\Delta R_C(x_i, y_i, m)$.

The LST in an urban area has strong spatial continuity compared with the surface reflectance image, and the spatial distribution of LST presents an irregular pattern because of the proximity effect of LST in urban areas and the distribution pattern of human activities. Consequently, assuming the changes for each class is the same among all coarse-resolution pixels from T_B to T_P in FSDAF that is unreasonable and not suitable for the LST fusion in an urban area. According to the first law of geography [78] that the near things are more relevant than the far things, and it is reasonable that $\Delta R_C(x_i, y_i, m)$ are the same in a small area. A sliding window is introduced to establish a linear model to perform the spatial unmixing process, after our test in two study areas, the size of a sliding window is designed 5 × 5 MODIS LST pixels, which can gain the optimal fusion result in urban areas, and it ensures that the linear equation of spatial unmixing is minimally influenced by collinearity and land cove type change. The unmixing equation is as follows: where *n* is the number of the coarse-resolution LST pixels in the sliding window. $\Delta R_C(c = 1 \ldots m)$ can be calculated by the inversion equation (4) shown at the bottom of the next page, and the changes of endmembers for coarse-resolution pixels can be obtained.

Because step 1 of CFSDAF has corrected the differences between coarse-resolution LST data and fine-resolution LST data, it is reasonable to assign $\Delta R_C(x_i, y_i, m)$ to the corresponding $\Delta R_F(x_{ij}, y_{ij}, m)$. The temporal increment for one fine-resolution pixel is the linear mixture with endmembers' abundances and the endmembers changes of the fine resolution, and the temporal increments $\Delta_F^t(x_{ij}, y_{ij})$ are calculated through (5). However, in FSDAF, $\Delta R_C(x_{ij}, y_{ij}, m)$ are directly assigned to the fine-resolution LST pixels of each class, which will produce the same change values for each class from T_B to T_P in the whole image, and it neglects the within-class variability of LST in the same land cover type and results in the spatial discontinuity of LST

$$\Delta_F^t\ (x_{ij}, y_{ij}) = A_F^B\ (x_{ij}, y_{ij}) \times \Delta R_F\,(x_{ij}, y_{ij}, m)\,. \tag{5}$$

4) Get the Spatial Increments by IDW Interpolation: Generally speaking, the predicted fine-resolution LST image on T_P can be calculated by combining the temporal increments and the fine-resolution LST image on T_B if there have no land cover changes between T_B and T_P. However, the land cover types often change from T_B to T_P in many cases, and the mutative signals of the land cover types can be gained from the coarse-resolution LST image on T_P. In CFSDAF, the IDW interpolation is introduced to downscale the coarse-resolution LST image [79]. The general idea of IDW interpolation supposes that the attribute value of an unsampled point is the weighted average of known values within the neighborhood, and the weights are inversely related to the distances between the predicted locations and the known locations [72]. In CFSDAF, the neighborhoods of IDW interpolation are set to a window size that is similar to spatial unmixing, which can ensure the interpolation accuracy and simultaneously reduce input parameters. IDW interpolation is applied to downscale the coarse-resolution LST image on T_B and T_P, respectively, and to gain the spatial increments from T_B to T_P, which are marked as $\Delta_F^s(x_{ij}, y_{ij})$. But in FSDAF, the TPS interpolation is only used for the coarse-resolution on T_P to distribute residual, which may underestimate the contribution of TPS interpolation [80].

In this article, IDW interpolation is used to obtain spatial increments instead of the TPS interpolation because of two main reasons. One reason is that the computational time of TPS interpolation will increase substantially when the TPS interpolation is applied to the large areas or long-term studies. Another reason is that the accuracy of IDW interpolation has no significant reduction compared with the TPS interpolation, and the differences in results by IDW interpolation or TPS interpolation are discussed in Section V-C.

5) Integrate the Spatial Increments and the Temporal Increments: The temporal increments and spatial increments are known as two independent predictions. The former is based on the spatial unmixing theory that depends on the temporal changes of LST, and the latter mainly relies on spatial dependence of LST based on IDW interpolation. The predicted results of temporal increments can reserve the spatial details and spatial continuity of LST but cannot capture the land cover type change, and the predicted results of spatial increments can obtain the information of land cover type changes but cannot retain the spatial details. Therefore, a reasonable integration of the above two increments can reserve the spatial details and monitor the land cover type change simultaneously.

An objective function of weighted increments is introduced to integrate the two increments, which was used in the IFSDAF method [80]. The main idea of the weighted method is summing the weighted spatial and temporal increments, and the combination is close to the real change of LST as possible. The mathematical formula of the objective function can be expressed as

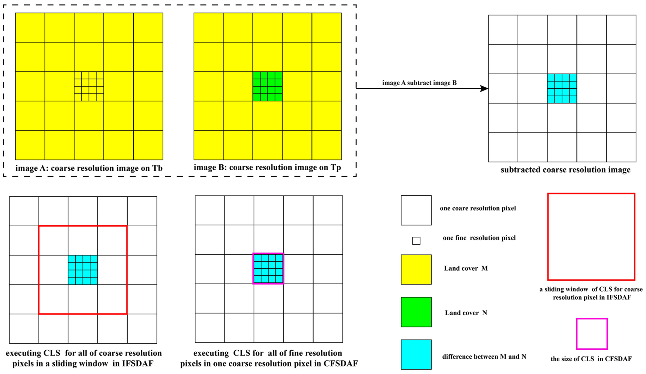


image A: coarse resolution image on Tb
image B: coarse resolution image on Tp
image A subtract image B
subtracted coarse resolution image
executing CLS for all of coarse resolution pixels in a sliding window in IFSDAF
executing CLS for all of fine resolution pixels in one coarse resolution pixel in CFSDAF
one coare resolution pixel
one fine resolution pixel
Land cover M
Land cover N
difference between M and N
a sliding window of CLS for coarse resolution pixel in IFSDAF
the size of CLS in CFSDAF

Fig. 2. Difference of CLS process between IFSDAF and CFSDAF.

$$\left(\widehat{w_t}, \widehat{w_s}\right) = \arg\min\left(w_t \Delta_F^t\left(x_{ij}, y_{ij}\right) + w_s \Delta_F^s\left(x_{ij}, y_{ij}\right) - \Delta F\left(x_{ij}, y_{ij}\right)\right)$$

$$\text{with } w_t, w_s \in (0,1)\ w_t + w_s = 1 \tag{6}$$

where w_t and w_s are the weight of temporal increments and spatial increments, respectively. $\Delta F(x_{ij}, y_{ij})$are the up-sampled results of coarse-resolution LST through nearest-neighbor interpolation, which can be regarded as the real changes of LST on a fine-resolution scale to some extent. Equation (6) can be calculated by the CLS method, and w_t and w_s for each fine-resolution pixel can be gained. In IFSDAF, the CLS is executed in a sliding window of the coarse resolution, which leads to the errors of CLS unmixing when the abrupt events occur in a small region as follows in Fig. 2, and the CLS of CFSDAF is implemented for fine-resolution pixels of one coarse-resolution pixel that ensures the accuracy of CLS. The final increments $\Delta F^{\text{int}}(x_{ij}, y_{ij})$ from T_B to T_P for fine-resolution LST pixels can be calculated as follows:

$$\Delta F^{\text{int}}\left(x_{ij}, y_{ij}\right) = w_t\, \Delta_F^t\left(x_{ij}, y_{ij}\right) + w_s \Delta_F^s\left(x_{ij}, y_{ij}\right) \tag{7}$$

where $\Delta F^{\text{int}}(x_{ij}, y_{ij})$ is the predicted increments by integrating the temporal increments $\Delta_F^t(x_{ij}, y_{ij})$ and the spatial increments $\Delta_F^t(x_{ij}, y_{ij})$ for fine-resolution pixels.

Although the predicted increments of fine-resolution LST by (6) and (7) can be regarded as the optimal results, they are not equal to real increments [73]. It has some residuals between predicted increments and real increments, and the residuals can be expressed as

$$r\left(x_i, y_i\right) = \Delta C\left(x_i, y_i\right) - \frac{1}{N}\left(\sum_{j\,=\,1}^{N} \Delta F^{\text{int}}\left(x_{ij}, y_{ij}\right)\right) \tag{8}$$

$r(x_i, y_i)$ is the residual for one coarse-resolution pixel. *N* is the number of fine-resolution pixels in one coarse-resolution pixel, and *j* is the *j*th fine-resolution pixel in one coarse-resolution pixel. $r(x_i, y_i)$ for one coarse-resolution pixel can be obtained

$$\begin{bmatrix} \Delta C\left(x_1, y_1\right) \\ \vdots \\ \Delta C\left(x_i, y_j\right) \\ \vdots \\ \Delta C\left(x_n, y_n\right) \end{bmatrix} = \begin{bmatrix} A_C^B\left(x_1, y_1, 1\right) & A_C^B\left(x_1, y_1, 2\right) & \cdots & A_C^B\left(x_1, y_1, m\right) \\ \vdots & \vdots & & \vdots \\ A_C^B\left(x_i, y_j, 1\right) & A_C^B\left(x_i, y_j, 2\right) & \cdots & A_C^B\left(x_i, y_j, m\right) \\ \vdots & \vdots & & \vdots \\ A_C^B\left(x_n, y_n, 1\right) & A_C^B\left(x_n, y_n, 2\right) & \cdots & A_C^B\left(x_n, y_n, m\right) \end{bmatrix} \begin{bmatrix} \Delta R_C\left(x_i, y_i, 1\right) \\ \vdots \\ \Delta R_C\left(x_i, y_i, 2\right) \\ \vdots \\ \Delta R_C\left(x_i, y_i, m\right) \end{bmatrix} \tag{4}$$

by (8), and the final residuals for all fine-resolution pixels in one coarse-resolution pixel are equal. The final increments from T_B to T_P can be calculated by

$$\Delta F^{\text{fin}}\ (x_{ij}, y_{ij}) = \ \Delta F^{\text{int}}\,(x_{ij}, y_{ij}) + R\,(x_{ij}, y_{ij}) \qquad (9)$$

where $\Delta F^{\text{fin}}(x_{ij}, y_{ij})$ is the final increment. $R(x_{ij}, y_{ij})$ is the residual for one fine-resolution pixel, and the $R(x_{ij}, y_{ij})$ is equal to $r(x_i, y_i)$ in one coarse-resolution pixel.

6) Get the Final Prediction by Using the Information of Neighborhood: After calculating the residuals, the final prediction values on T_P can be calculated by the following equation:

$$\widehat{F_P}\ (x_{ij}, y_{ij}) = F_B\ (x_{ij}, y_{ij}) + \Delta F^{\text{fin}}\,(x_{ij}, y_{ij}) \qquad (10)$$

where $\widehat{F_P}(x_{ij}, y_{ij})$ is the final prediction value. However, due to the calculation complexity of CFSDAF and calculation ways of pixel by pixel, and it inevitably causes some calculation errors. Moreover, step (5) for distributing the residuals is based the coarse-resolution pixel, which produces the "block effects." To solve the above problems, this article adopts the information of neighborhood to smooth the above predicted results that can reserve more spatial details, and a specific calculation process can be found in the STARFM or FSDAF method [25], [63]. The size of the neighborhood is the same as the spatial unmixing in step (2), which can reduce the input parameters. The final prediction of the fine-resolution LST image on T_P can be calculated by

$$F_P^{\text{fin}}\,(x_{ij}, y_{ij}) = F_B\,(x_{ij}, y_{ij}) + \sum_{k=1}^{n} w_k \Delta F^{\text{fin}}\,(x_{ij}, y_{ij}) \quad (11)$$

where $F_P^{\text{fin}}(x_{ij}, y_{ij})$ is the predicted fine-resolution LST image on T_P. *n* is the number of similar pixels for central pixel in a sliding window, and *k* is the *k*th similar pixel. w_k is the weight for *k*th similar pixel.

III. Testing Data

A. Study Area and Data

Beijing and Shenzhen in China were selected as study areas to test the performance of CFSDAF. Most cities in the world are located at the low- or mid-latitude regions where large urban populations are living, and Beijing and Shenzhen are typical mega cities and seated in the north and south of China, which have a certain representativeness for assessing the CFSDAF. The location of two study areas is shown in Fig. 3.

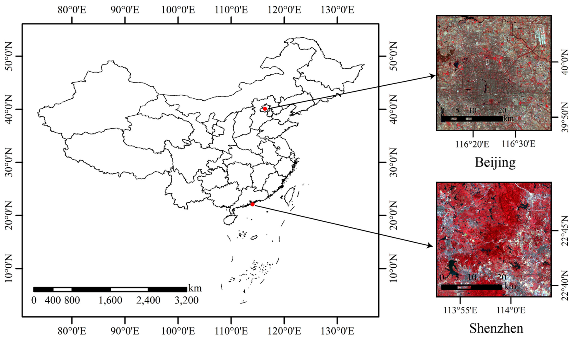


Fig. 3. Location of two study areas.

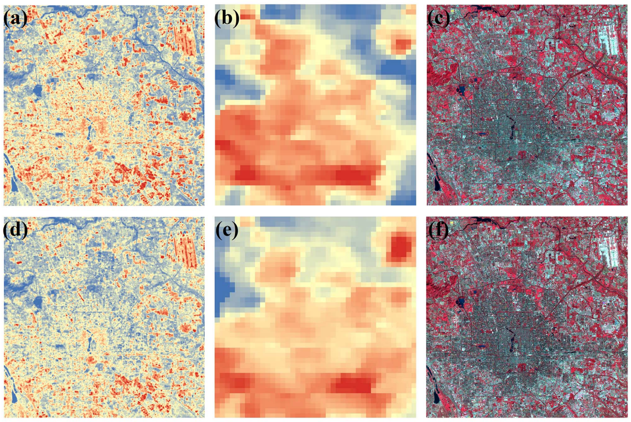


Fig. 4. Experimental data in a heterogeneous area. (a)–(c) Landsat LST image, MODIS LST image, and Landsat visible and near-infrared image on September 4, 2014; (d)–(f) Landsat LST image, MODIS LST image, and Landsat visible and near-infrared image on October 6, 2014, in Beijing area, respectively.

Beijing, as the capital of China, is one of the most populations in China, with a population of more than 20 million, where the landscape is highly heterogeneous, and its latitude is about 40° north. In this study, two pairs of MODIS LST and Landsat LST were selected to test the CFSDAF, and the dates of selected images are September 4, 2014, and October 6, 2014. We designed two groups' experiments to test the performance of CFSDAF in Beijing. When the base date is on September 4, 2014, and the prediction date is on October 6, 2014, we called it a forward prediction, and it can be regarded as a backward prediction from October 6, 2014, to September 4, 2014. The MODIS LST images are the MODIS daily surface temperature product data (MOD11A1), and the resolution is 1000 m, which can be downloaded from NASA. The Landsat LST images are a secondary product and downloaded from USGS, which have been resampled to 30-m resolution. The range of the study area in Beijing is 40 km × 40 km that contains the main urban area, and the MODIS LST images and Landsat LST images are listed in Fig. 4.

Shenzhen, as the demonstration pilot zone of China, its latitude is about 22° north, which was developed from a small fishing village to international big city in the past 40 years. The land cover type has changed significantly in Shenzhen in the past, and Shenzhen was adopted to test the performance of CFSDAF in abrupt areas. Shenzhen experienced rapid urban expansion from 2000 to 2003, and the dates of input images are September 14, 2000, January 10, 2003. To better present the signals of land cover type change, the simulated MODIS-like images were used as coarse-resolution LST images because some information of land cover change was inconsistent between MODIS image and Landsat image, and similar treatments have been adopted in past fusion methods [49], [63]. Specifically, the 30-m Landsat LST image was aggregated to 240 m, which can reserve information of land cover change. Second, the aggregated image added some certain values, which considered the differences between MODIS LST image and Landsat LST image, and the simulated

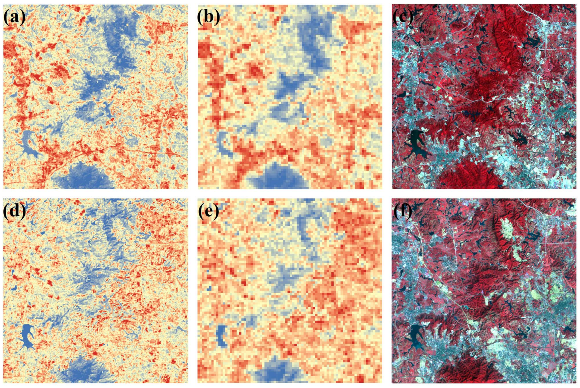


Fig. 5. Experimental data in an abrupt area. (a)–(c) Landsat LST image, MODIS-like LST image, and Landsat visible and near-infrared image on September 14, 2000; (d)–(f) Landsat LST image, MODIS-like LST image, and Landsat visible and near-infrared image on January 10, 2003, in Shenzhen area, respectively.

MODIS-like LST can be obtained. The study data of Shenzhen are listed in Fig. 5.

There have six images for Beijing area or Shenzhen area, which are two pairs of MODIS LST images, Landsat LST images, and Landsat visible and near-infrared images on T_B and T_P. The Landsat LST image on T_P was selected to compare the fusion results and assess the fusion accuracy. Due to the heterogeneous landscape in urban areas, the LST of the predicted image at 30-m resolution is still a mixed pixel. In-situ LST data, which represent a single landscape on a point scale, was selected to assess the predicted image in an urban area that has more than one land cover type is not appropriate. All input images were corrected and geographically co-registered and clipped the same size. The scale ratio between MODIS LST image and Landsat LST image is 32 for Beijing area, and 8 for Shenzhen area. After the above processing, all input images are ready for CFSDAF fusion.

B. Comparison and Evaluation

The fusion results obtained from CFSDAF were compared visually and quantitatively with two popular methods, which are STARFM and FSDAF, and the above three methods had the same input images. To further ensure the accuracy and fairness of the evaluation process, the parameters of STARFM and FSDAF are consistent with CFSDAF. The average difference (AD), root-mean-square error (RMSE), correlation coefficient (CC), Robert's edge (Edge), and local binary pattern (LBP) were selected to evaluate the accuracies between the synthetic Landsat-like LST image and the reference Landsat LST image on T_P. Edge and LBP are considered as the optimal accuracy index to evaluate the spatial accuracy of fusion results [81].

The closer value of AD or RMSE is 0, and the closer value of CC is 1. The predicted image is closer to the real image. The range of Edge or LBP between -1 and 1, and 0 represents the best fusion result. More negative values indicate edge features or textural features in the fused image over smoothed; more positive values indicate edge features in the fused image over sharpened.

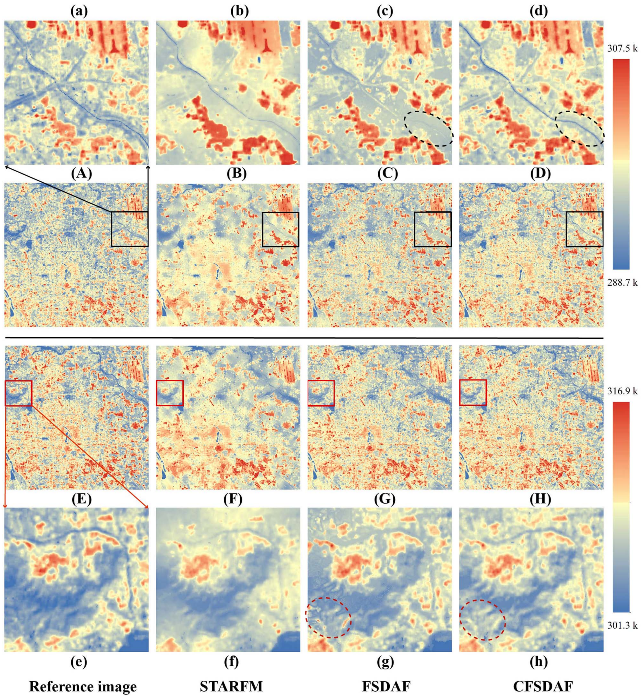


Fig. 6. Visual comparison between reference images and the fusion images by STARFM, FSDAF, and CFSDAF for the heterogeneous area in Beijing (the upper part of the black line is called forward prediction and the lower part is backward prediction). (A) Original Landsat LST image of October 6, 2014, and the predicted LST image by STARFM (B), FSDAF (C), and CFSDAF (D), the corresponding lowercase (a)–(d) are the enlarged black box in the figure. (E) Original Landsat LST image of September 4, 2014, and the predicted LST images by STARFM (F), FSDAF (G), and CFSDAF (H), the corresponding lowercase (e)–(f) are the enlarge red box in the figure.

IV. Results

A. Visual Comparison Between the Predicted Images and Reference Images in Two Experiment Regions

The first experiment area was selected to evaluate the performance of CFSDAF for the heterogeneous area in Beijing. Fig. 6 exhibits the fusion images of two groups' experiments (forward prediction and backward prediction) by STARFM, FSDAF, and CFSDAF. The visual comparisons between the predicted images and reference images were adopted to judge the quality of fusion results through three methods. It is apparent that the predicted images by STARFM [see (B) and (F) in Fig. 6] were inferior to the predicted images by FSDAF [see (C) and (G) in Fig. 6] and CFSDAF [see (D) and (H) in Fig. 6] compared with the reference images [see (A) and (E) in Fig. 6], and the predicted images by STARFM cannot reserve the spatial details and smooth the boundaries of different land cover type. Spatial distributions of the predicted images by FSDAF and CFSDAF for two groups of experiments are closer to the reference images. To better present the differences between the three fusion methods in the heterogeneous area, we enlarged the black box for forward prediction and the red box for backward prediction to exhibit more spatial details. Compared Fig. 6(a) with Fig. 6(b)–(d), it is obvious that spatial details of ground objects cannot be preserved using STARFM [see Fig. 6(b)], and the predicted image by FSDAF [see Fig. 6(c)] cannot obtain the puny river but CFSDAF can capture the puny river and more spatial details,

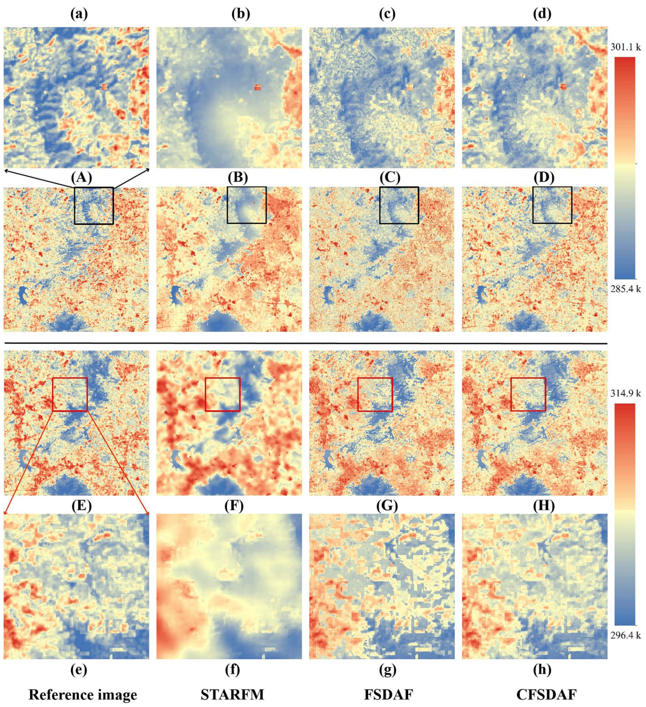


Fig. 7. Visual comparison between reference images and the fusion images by STARFM, FSDAF, and CFSDAF for the abrupt area in Shenzhen (the upper part of black line is called forward prediction and the lower part is backward prediction). (A) Original Landsat LST image of January 10, 2003, and the predicted LST images by STARFM (B), FSDAF (C), and CFSDAF (D), the corresponding lowercase (a)–(d) are the enlarged black box in the figure. (E) Original Landsat LST image of September 14, 2000, and the predicted LST images by STARFM (F), FSDAF (G), and CFSDAF (H), the corresponding lowercase (e)–(f) are the enlarged red box in the figure.

TABLE I
ACCURACY ASSESSMENT OF THE STARFM, FSDAF, AND CFSDAF IN TWO EXPERIMENTAL AREAS

Method			AD	RMSE	CC	Edge	LBP
STARFM	Beijing	forward	1.007	1.260	0.867	-0.3803	0.0063
		backward	1.416	1.794	0.807	-0.4215	-0.0143
	Shenzhen	forward	1.230	1.482	0.817	-0.5288	0.1077
		backward	1.348	1.714	0.892	-0.6726	0.0861
FSDAF	Beijing	forward	0.997	1.263	0.897	-0.1124	-0.0068
		backward	1.117	1.399	0.893	-0.2983	-0.0123
	Shenzhen	forward	1.383	1.671	0.725	-0.2893	**0.0749**
		backward	1.424	1.816	0.873	**-0.2268**	**0.0450**
CFSDAF	Beijing	forward	**0.915**	**1.177**	**0.904**	**-0.0630**	**0.0052**
		backward	**0.920**	**1.176**	**0.902**	**-0.2253**	**-0.0056**
	Shenzhen	forward	**1.025**	**1.290**	**0.837**	**-0.2489**	0.0788
		backward	**1.184**	**1.535**	**0.893**	-0.2465	0.0456

The best results are marked in bold.

and spatial continuity of LST. From the dashed black ellipse of Fig. 6(c) and (d), there is no obvious change from the puny river to side in FSDAF, mainly due to the FSDAF does not consider the intraclass variability. And CFSDAF can retain the spatial details and spatial continuity. Compared Fig. 6(e) with Fig. 6(f)–(h), we find the same spatial distribution pattern with Fig. 6(a)–(d), and FSDAF cannot capture the boundary of the puny river and spatial details. From the dashed red ellipse of Fig. 6(g) and (h), FSDAF overestimates LST of the ground object, and CFSDAF is closer to the reference image. As a result, the CFSDAF is better than FSDAF and STARFM in heterogeneous areas from forward prediction and backward prediction, and the fusion images by CFSDAF can retain more spatial details, distinguish the boundaries of different land cover type, reserve the puny spatial features.

The second experiment area in Shenzhen was adopted to test the performance of CFSDAF for an abrupt event, where experienced rapid urbanization and the land cover types have changed from vegetation to built-up areas. Fig. 7 shows the fusion results by STARFM, FSDAF, and CFSDAF. From the visual comparison, the two experiments including forward prediction and backward prediction exhibit that STARFM [see (B) and (F) in Fig. 7] cannot reserve more spatial details compared with FSDAF [see (C) and (G) in Fig. 7] and CFSDAF [see (D) and (H) in Fig. 7], and the predicted images by FSDAF and CFSDAF are closer to the actual images. Meanwhile, the predicted image [see (C) in Fig. 7] by FSDAF cannot reserve spatial details and have much "noise point," which is mainly caused by the lack of intraclass variability in FSDAF. The black box from [see (A)–(D) in Fig. 7] is enlarged to exhibit more spatial details, where land cover types have changed. From Fig. 7(a) to (d), the predicted image by STARFM [see Fig. 7(b)] cannot distinguish the boundaries of different land cover types and cannot accurately retain the abrupt area, and FSDAF and CFSDAF can capture abrupt information to some extent, which is mainly due to the interpolation process for the coarse-resolution image on T_P. From the enlarged area of Fig. 7(e)–(h), the predicted image by STARFM cannot capture the spatial details and FSDAF cannot reserve spatial continuity, and the predicted image by CFSDAF is basically consistent with the reference image in terms of spatial distribution.

B. Quantitative Evaluation Between the Predicted Images and Reference Images

The quantitative evaluation of fusion images in two study areas is listed in Table I, each study area has two pair experiments including forward prediction and backward prediction. Five assessment indices including AD, RMSE, CC, Edge, and LBP are selected to compare the fusion images by STARFM, FSDAF, and CFSDAF separately.

For the assessment indices in Beijing area, all assessment indices suggest that CFSDAF is superior to STARFM and FSDAF. Specifically, CFSDAF has the smaller AD, RMSE, and higher CC compared with STARFM and FSDAF. For forward prediction, the CFSDAF is slightly better than STARFM and FSDAF. But for backward prediction, there have obvious differences for STARFM, FSDAF, and CFSDAF (AD 1.416 versus 1.117 and 0.920, RMSE 1.794 versus 1.399 and 1.176, CC 0.807 versus 0.893 and 0.902). The metrics of Edge and LBP also clearly demonstrate that CFSDAF is much better than FSDAF and STARFM regarding the spatial details (Edge −0.063 versus −0.1124 versus −0.3803 for forward prediction and −0.2253 versus −0.2928 versus −0.4215 for backward prediction). For Shenzhen area, whether forward prediction or backward prediction, four indices suggest that CFSDAF is superior to STARFM and FSDAF from assessment metrics of AD, RMSE, and CC. Moreover, we find the fusion accuracies of FSDAF are lower

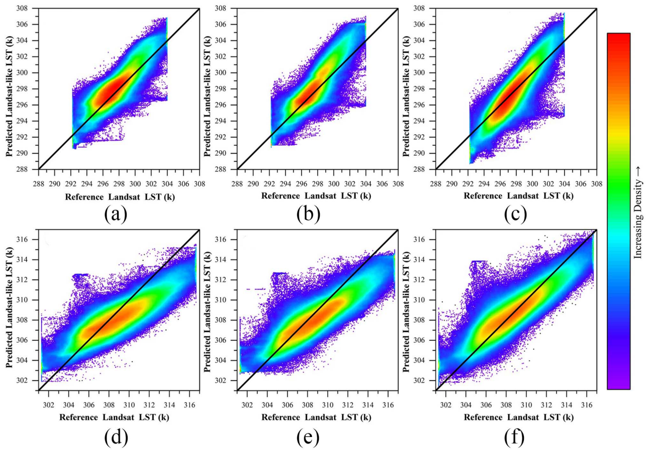


Fig. 8. Scatter plots of predicted Landsat-like LST and reference Landsat LST in Beijing area. (a)–(c) Forward prediction. (d)–(f) Backward prediction. Results by (a) and (d) STARFM, (b) and (e) FSDAF, and (c) and (f) CFSDAF.

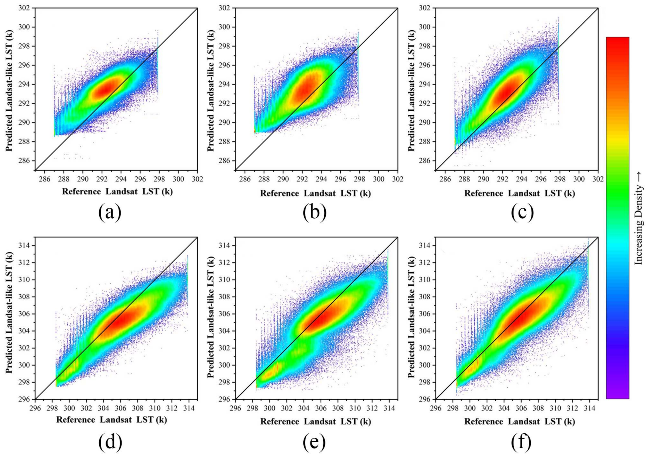


Fig. 9. Scatter plots of predicted Landsat-like LST and reference Landsat LST in Shenzhen area. (a)–(c) Forward prediction. (d)–(f) Backward prediction. Results by (a) and (d) STARFM, (b) and (e) FSDAF, and (c) and (f) CFSDAF.

than STARFM, which is mainly due to the FSDAF lack of intraclass variability and in turn produces much "noise." For AD, the CFSDAF has an approximate 0.36 K reduction over FSDAF and 0.2 K reduction over STARFM for forward prediction, and 0.24 K reduction and 0.16 K reduction for backward prediction. The obvious improvements are also available from RMSE (1.482 versus 1.671 and 1.290 for forward prediction, 1.714 versus 1.816 and 1.535 for backward prediction), and CC (0.817 versus 0.725 and 0.837, 0.892 versus 0.873 and 0.893). Meanwhile, we find that Edge for CFSDAF is better than FSDAF from the forward prediction (−0.2489 versus −0.2893) and is slightly worse than FSDAF (0.2465 versus −0.2268), but LBP for CFSDAF is almost identical with FSDAF from forward and backward predictions. From the spatial accuracy (Edge and LBP), CFSDAF is better than FSDAF and far better than STARFM in Beijing, but the CFSDAF and FSDAF are basically the same in Shenzhen, and both are better than STARFM. The above differences between the two regions are mainly since the heterogeneity of Beijing is stronger than that of Shenzhen, which further indicates that the CFSDAF can better retain spatial details in regions with strong heterogeneity.

Figs. 8 and 9 show the Scatter plots of predicted Landsat-like LST and reference Landsat LST in Beijing area and Shenzhen area, separately, and (a)–(c) are the forward prediction and (d)–(f) are the backward prediction. It is distinctly seen that the predicted results by CFSDAF are better than FSDAF and STARFM whenever forward prediction or backward prediction in Beijing or Shenzhen area, where more values are concentrated on the side of the black line (also called a balanced line) from Figs. 8 and 9.

C. Spatial Distribution of Errors for Predicted LST Images

Figs. 10 and 11 present the distribution errors in Beijing and Shenzhen area separately. (a)–(d) represent the distribution errors of forward prediction, and (e)–(f) are backward predictions for Figs. 10 and 11. The distribution errors are regarded as

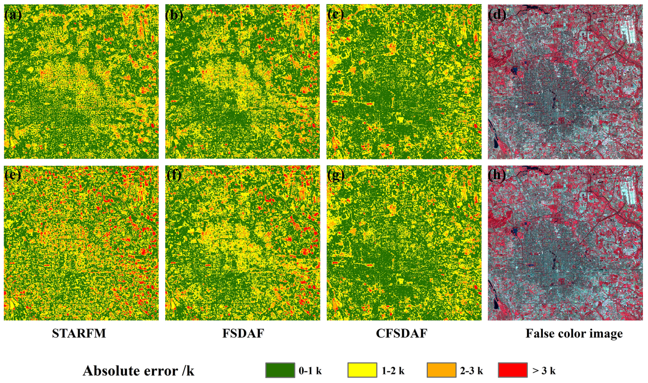


Fig. 10. Distribution errors of the predicted images for three fusion methods in Beijing area. (a)–(d) Forward prediction. (e)–(f) Backward prediction. From left to right in the figure, there are distribution errors of (a) and (e) STARFM, (b) and (f) FSDAF, and (c) and (g) CFSDAF, and (d) and (h) the corresponding of False color image.

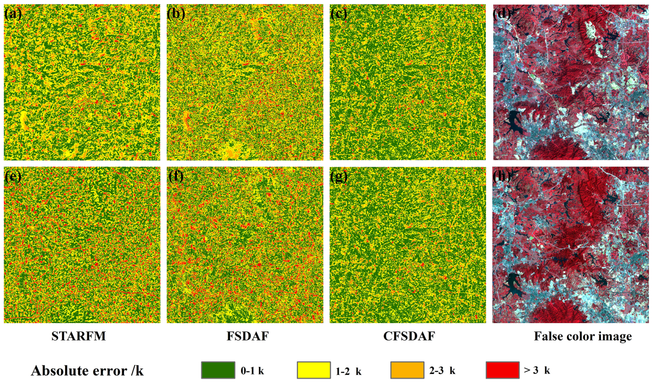


Fig. 11. Distribution errors of the predicted images for three fusion methods in Shenzhen area. (a)–(d) Forward prediction. (e)–(f) Backward prediction. From left to right in the figure, there are distribution errors of (a) and (e) STARFM, (b) and (f) FSDAF, and (c) and (g) CFSDAF, and (d) and (h) the corresponding of False color image.

TABLE II
AREA PERCENTAGE OF DISTRIBUTION ERRORS FOR THREE PREDICTED LST IMAGES BY STARFM, FSDAF, AND CFSDAF AT FOUR ERROR LEVELS IN BEIJING AREA

Beijing	Error level(K)	STARFM	FSDAF	CFSDAF
Forward prediction	0–1	56.7	58.2	62.8
	1–2	32.4	30.7	28.8
	2–3	9.2	9.1	7.0
	>3	1.7	2.0	1.4
Backward prediction	0–1	47.5	51.9	62.2
	1–2	32.5	33.9	29.5
	2–3	14.3	10.8	6.9
	>3	5.7	3.5	1.4

TABLE III
AREA PERCENTAGE OF DISTRIBUTION ERRORS FOR THREE PREDICTED LST IMAGES BY STARFM, FSDAF, AND CFSDAF AT FOUR ERROR LEVELS IN SHENZHEN AREA

Shenzhen	Error level(K)	STARFM	FSDAF	CFSDAF
Forward prediction	0–1	44.3	39.6	56.4
	1–2	38.6	36.2	32.4
	2–3	14.1	18.2	9.1
	>3	3	6	2.1
Backward prediction	0–1	46.1	43.3	52.4
	1–2	30.5	31.3	29.6
	2–3	15.1	15.7	12.4
	>3	8.3	9.7	5.6

the absolute errors between the predicted images and reference images on T_P, which can reveal the spatial distribution of errors for different land cover types, and it is useful to study spatiotemporal variations of UHI. The distribution errors of predicted LST images are classified into four levels: 1) 0–1K, 2) 1–2K, 3) 2–3K, and 4) $>$3K.

For the distribution errors map of Beijing area in Fig. 10, it can be seen intuitively that there have significant differences in the spatial distribution of errors between three fusion methods from forward prediction. The prediction accuracy of CFSDAF in urban regions is significantly better than that of STARFM and FSDAF, which have more green pixels. For backward prediction, the proportion of absolute errors with less than 1K for CFSDAF is significantly larger than that for STARFM and FSDAF, especially in urban region. In addition, the proportion of errors greater than 2K for STARFM is greater than that for FSDAF and CFSDAF, and the proportion of errors greater than 3K for CFSDAF is the lowest. Table II shows the area percentage of distribution errors for three predicted LST images in Beijing area. There is about 62.8% area under 1K for CFSDAF from the forward prediction, which has some improvements as compared to FSDAF with 4.6% and STARFM with 6.1%. similarly, the proportion of less than 1K is significantly different among the three methods (47.5% for STARFM versus 51.9% for FSDAF and 62.2% for CFSDAF) from backward prediction.

For the distribution errors map of Shenzhen area in Fig. 11, whether for forward prediction or backward prediction, the proportion of absolute errors with less than 1K for CFSDAF is higher than STARFM and FSDAF, and the spatial distribution of that is discretely distributed. For forward prediction, the proportion of less than 1K for STARFM is greater than that in FSDAF, and the percentage greater than 2K in FSDAF is significantly more than that in STARFM. For backward prediction, there has more proportion of greater than 2K in FSDAF and STARFM than in CFSDAF, and the absolute errors of CFSDAF with less than 1K is the largest for the three fusion methods. Table III shows the area percentage of distribution errors in Shenzhen area. Intuitively, the proportion of less than 1K for CFSDAF is the highest (56.4% and 52.4%), followed by STARFM (44.3% and 46.1%) and FSDAF (39.6% and 43.3%). It can be seen from Table III that the percentage of larger than 3K in STARFM (3% and 8.3%) and FSDAF (6% and 9.7%) are higher than CFSDAF (2.1% and 5.6%).

V. DISCUSSION

A. *Improvements of CFSDAF Compared With FSDAF*

The experiment results of two study areas in Section IV show that the CFSDAF outperforms FSDAF for LST fusion, which is mainly due to the following reasons.

First, the differences between coarse-resolution LST images and fine-resolution LST images were adjusted by introducing a simple linear model at a coarse-resolution scale. Compared with the surface reflectance data, the LST image has two unique characteristics. One characteristic is that the LST will change significantly over time, which leads to obvious differences in LST images obtained by different sensors at different times. Another is that the LST image is retrieved from thermal infrared data, and different inversion methods for the LST image will generate diverse results. Therefore, correcting the differences between coarse-resolution LST and fine-resolution LST is the first step and key step. The main reasons for choosing a lineal model rather than a nonlinear model in this study are as follows. One reason is that an appropriate nonlinear model is difficult to choose to fit the differences in LST with different resolutions. Another reason is that even if the nonlinear model can achieve a good fitting effect on T_B, applying the model to the adjustment between coarse-resolution LST and fine-resolution LST on T_P, which will bring about greater errors. Due to the LST of fine-resolution pixels in an urban area is the mixed result with different land cover types, previous study that adjusting the differences based on land cover classification map at fine resolution [39], which may lead to discontinuity of LST at the boundaries of different land cover type. Meanwhile, if the land cover map changes or is misclassified, and some errors will be caused.

Second, the visible and near-infrared image of the fine resolution was introduced to perform soft classification, which considered the mixed pixel of the fine-resolution LST image in an urban area, and the predicted LST images can reserve more spatial details and spatial continuity. Specifically, the visible and near-infrared image was used to extract endmembers and execute spectral unmixing, to gain the abundance of fine-resolution image. Then, the abundances of fine-resolution image were aggregated to generate abundances of coarse-resolution LST data. At last, the CLS was carried out for spatial unmixing in a moving window to get the temporal increments because the LST has the spatial continuity. However, the FSDAF considers fine pixel as pure pixel and execute hard classification, which ignore the mixed pixel in the heterogeneous area and does not consider the within-class variability, and the fusion results of FSDAF will be further away from the actual results when the classification map has some errors. Hence, the fusion images by FSDAF have much "noise" or "patch," which cannot retain spatial continuity and spatial details of LST. Moreover, for the unmixing process of FSDAF, the k-purest coarse pixels in the whole image were selected to perform spatial unmixing, where the way of selecting the purest pixels is empirical and not strict [49], [82], and it ignores the spatial variability of LST.

Third, IDW interpolation was introduced to replace TPS interpolation. Although the TPS interpolation has higher interpolation accuracy than IDW interpolation [63], the computational efficiency of IDW interpolation is faster than TPS interpolation, especially for large areas or long-term studies. The time complexity of IDW interpolation is O (n3), and the TPS interpolation is O(n) [50]. Therefore, there has a tradeoff between accuracy and efficiency for the actual application. Moreover, the TPS interpolation for FSDAF was only used to interpolate coarse-resolution image on T_P, and the interpolated result was selected to guide the distribution of residuals, which underestimated the contribution of interpolation results for the final fusion results [80]. In CFSDAF, the difference values of coarse-resolution images from T_B to T_P were interpolated to gain the spatial increments, the signals of land cover changes can be captured when there have the land cover changes between T_B and T_P.

Fourth, the CLS method was selected to combine the temporal increments and spatial increments. Using the CLS method to integrate the two increments can preserve the spatial details and signal of land cover change simultaneously. Furthermore, it has been found that the contribution of spatial increments for final fusion results is more than the temporal increments to some extent [80]. Accordingly, the scale between coarse-resolution image and fine-resolution image should be relatively small, which can ensure the interpolation result with more accuracy and, in turn, improve the final fusion precision. However, in FSDAF, a homogeneity index HI was introduced to distribute residuals, which is based on the map classification on T_B. When there is the land cover change or misclassification, more errors will be introduced, and the misclassification map for LST is very common because the classification map is based on one band of LST.

B. Influences of Scale Between MODIS LST and Landsat LST

There has a common phenomenon that the resolution of the thermal infrared sensor is lower than the visible and near-infrared sensor for the same satellite (e.g., MODIS LST pixel is 1000 m, surface reflectance of MODIS pixel is 250 or 500 m). In many cases, the urban surface changes happen in a small area, which is difficult to be captured within the resolution of 1000 m. As shown in Fig. 12, A and B are two MODIS pixels with 1000-m resolution, and A is a pixel with land cover change, but B is a pixel without land cover change. Nevertheless, when the resolution of MODIS pixel with1000 m is downscaled to 250 m, A1 becomes the pixel without land cover type change and the B4 is a changed pixel. Accordingly, the downscaled MODIS image can accurately reserve more spatial information and monitor the small changes, especially for the boundary of different land cover types. Some scholars have discussed the differences in fusion results between MODIS LST pixels without downscaling and downscaled MODIS LST pixels [83], the results show that the latter has more accuracy than the former.

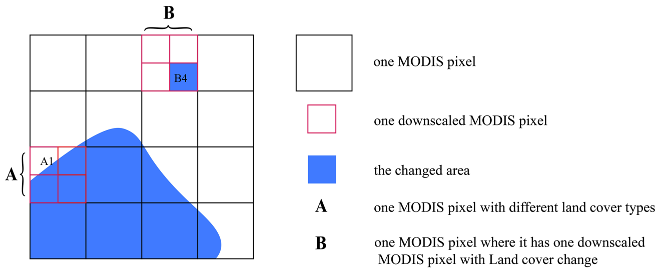

B
B4
A
A1
one MODIS pixel
one downscaled MODIS pixel
the changed area
A
one MODIS pixel with different land cover types
B
one MODIS pixel where it has one downscaled MODIS pixel with Land cover change

Fig. 12. Land cover type differences in MODIS pixels at different resolutions.

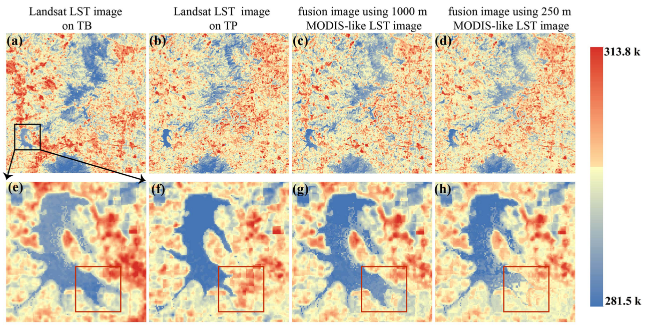

Landsat LST image on TB
Landsat LST image on TP
fusion image using 1000 m MODIS-like LST image
fusion image using 250 m MODIS-like LST image
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
313.8 k
281.5 k

Fig. 13. Differences in fusion images by using the 1000-m MODIS-like LST image and the 250-m MODIS-like LST image. (a) and (b) Landsat LST image on T_B (September 14, 2000) and on T_P (January 10, 2003); (c) and (d) fusion image by using the 1000-m MODIS-like LST image and the 250-m MODIS-like LST image separately. (e) and (f) Zoom-in areas of black block in (a) and (d).

In this section, the simulative MODIS-like LST image with 1000 m by upscaling the Landsat LST image to execute CFSDAF with 30-m Landsat LST image and to compare the predicted image by using a 250-m MODIS-like LST image and then evaluate the impact of downscaling the coarse-resolution LST image on fusion result. Fig. 13(a) and (b) shows the Landsat LST images on T_B (September 14, 2000) and T_P (January 10, 2003), and Fig. 13(c) and (d) shows the fusion images by using a 1000-m MODIS-like LST image and a 250-m MODIS-like LST image separately. Intuitively, there is no obvious difference between Fig. 13(c) and (d). When we enlarge the black block from Fig. 13(a)–(d), it can be seen from the red block of zoomed area in Fig. 13(e)–(f) that the area of lake decreases from T_B to T_P, and the fusion results indicate that only using a 250-m MODIS LST image for fusion can capture this change, which mainly due to the MODIS-like LST with 250-m resolution can reserve the more spatial details that are the changed lake area compared with 1000-m resolution MODIS-like LST.

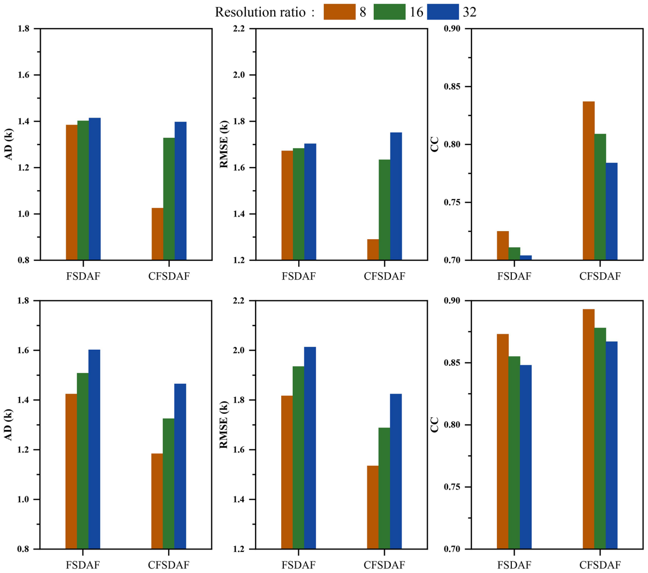


Fig. 14. Quantitative comparison of the LST fusion results under different levels of resolution ratio between the MODIS LST image and the Landsat LST image in Shenzhen area. There are AD, RMSE and CC from left to right (above for forward prediction, below for backward prediction).

Furthermore, downscaling the 1000-m MODIS LST image to the 250-m resolution can narrow the resolution ratio between MODIS LST pixel and the Landsat LST pixel (the resolution ratio from 32 to 8), which can improve the accuracy of IDW interpolation and further enhance the final predicted result. In this section, we upscale the Landsat LST image to generate the MODIS-like LST image with 250-, 500-, 1000-m resolution, where the resolution ratios between MODIS-like LST pixel and Landsat LST pixel are 8, 16, and 32, separately, to analysis the sensitivity of CFSDAF to the resolution ratio. In Fig. 14, it can be concluded that the fusion accuracies (from AD, RMSE, and CC) of CFSDAF are better than FSDAF in different resolution scales, whether there is the forward prediction or backward prediction. Meanwhile, with the improvements of the resolution ratio (from 32 to 8), the change ranges of the four indices in CFSDAF are greater than that in FSDAF, and when the resolution ratio goes from 32 to 8, the differences between CFSDAF and FSDAF gradually increases from AD and RMSE, which shows CFSDAF is more sensitive to the resolution scale. Consequently, the downscaling or superresolution technical of LST can be used to improve the fusion accuracy of LST before executing the CFSDAF [16], [84].

C. Differences Between CFSDAF-IDW and CFSDAF-TPS

In this study, IDW interpolation was adopted for the CFSDAF method instead of TPS interpolation. It is necessary to evaluate the differences between the CFSDAF-based IDW interpolation (CFSDAF-IDW) and the CFSDAF-based TPS interpolation (CFSDAF-TPS) in fusion accuracy and the computational efficiency. Table IV shows the fusion accuracies of four indices for CFSDAF-IDW and CFSDAF-TPS in Beijing and Shenzhen. The fusion accuracies of CFSDAF-IDW are almost the same as the CFSDAF-TPS in Beijing area, and the fusion accuracies of CFSDAF-IDW are slightly lower than the CFSDAF-TPS in Shenzhen area. Therefore, the puny errors can be ignored between CFSDAF-IDW and CFSDAF-TPS, but the computational efficiency of CFSDAF-IDW has been significantly improved. From Fig. 15, the computation time of CFSDAF-TPS significantly increases as the numbers of coarse pixels increase gradually. Accordingly, it will be a better choice to choose the CFSDAF-IDW when handling the data for the large areas or long-term studies. To facilitate the users' choice, the IDW interpolation and TPS interpolation are provided as an option in the CFSDAF program, but the default is IDW interpolation.

TABLE IV
COMPARISON OF FUSION ACCURACIES BETWEEN CFSDAF-IDW AND CFSDAF-TPS IN BEIJING AREA AND SHENZHEN AREA

			AD	RMSE	CC
CFSDAF-IDW	Beijing	forward	0.915	1.177	0.904
		backward	0.920	1.176	0.902
	Shenzhen	forward	1.025	1.290	0.837
		backward	1.185	1.535	0.892
CFSDAF-TPS	Beijing	forward	0.915	1.176	0.904
		backward	0.919	1.175	0.902
	Shenzhen	forward	1.015	1.274	0.840
		backward	1.184	1.533	0.893

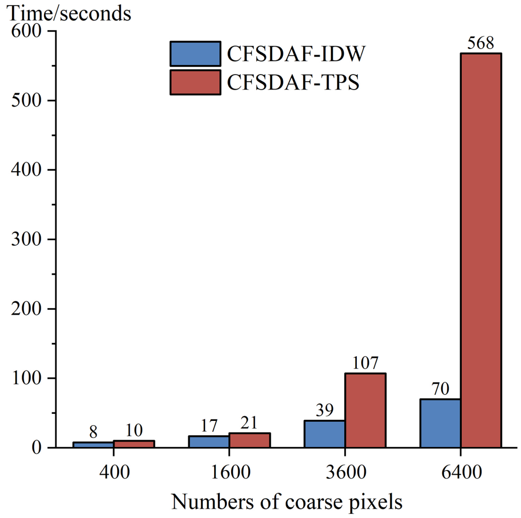


Fig. 15. Comparation of computational time for CFSDAF-IDW and CFSDAF-TPS.

D. Further Improvements of CFSDAF

Although the CFSDAF can reserve spatial details and spatial continuity of LST and capture the information of Land cover change for LST fusion, there have some limitations. First, the input images for CFSDAF are cloud-free and have good quality, which is difficult to be obtained for most of the areas at low latitudes. Combining the spatiotemporal reconstruction method and the spatiotemporal fusion method could be an effective way to generate all-weather data for most areas of the world [29]. Second, due to the fusion accuracy of CFSDAF affected by spatial increments to a large extent, the resolution ratio between the coarse-resolution LST and fine-resolution LST should not be large. Considering the large resolution scale between the MODIS LST image and the Landsat LST image, we suggest integrating the downscaling of the MODIS LST image and spatiotemporal fusion to achieve higher fusion accuracies for LST. Third, the linear model used to adjust the difference of coarse-resolution LST and fine-resolution LST may bring in large errors if the time interval between base date and prediction date is too large. The code of CFSDAF can be found on the URL: https://github.com/max19951001/CFSDAF.

VI. Conclusion

This study proposed a CFSDAF method for LST fusion in an urban area, which first adjusted the differences between coarse-resolution LST and fine-resolution LST at a coarse-resolution scale. Then, it considered the mixed pixel of the fine-resolution LST image (e.g., Landsat image) by introducing surface reflectance data of the Landsat image for performing soft classification, and the predicted images by CFSDAF can restore more spatial details and spatial continuous of LST. IDW interpolation was adopted to replace TPS interpolation in FSDAF, which greatly improved the computational time of CFSDAF for large areas or long-term studies and ensured the fusion precision as well. Moreover, the CLS method was selected to combine the temporal increments and spatial increments at the fine-resolution scale to take advantage of both increments, and the final fusion results can reserve the spatial details of LST and monitor the LST in abrupt areas simultaneously.

Beijing and Shenzhen are selected to test the performance of CFSDAF for LST fusion in heterogeneous and abrupt area, and the experiments of two areas show that the LST images by CFSDAF are more accurate than the other two fusion methods (STARFM and FSDAF) from the visual comparison and quantitative assessment. Moreover, the computational efficiency of CFSDAF is better than FSDAF. In addition, the accuracies of CFSDAF are influenced by the resolution ratio between coarse-resolution LST and fine-resolution LST, and we recommend shrinking the resolution ratio of two sensors before executing spatiotemporal fusion to improve the accuracy of the predicted image.

Although the CFSDAF is originally developed for LST fusion, it has the potential to fusion other products such as surface reflectance or vegetation index, and we also call for more testing of CFSDAF by using other sensors such as VIIRS LST and SLSTR LST.

Acknowledgment

The authors would like to thank Prof. Jin Chen and Prof. Xiaolin Zhu for providing the fusion IDL code of STARFM and FSDAF.

References

[1] N. H. Koroso, M. Lengoiboni, and J. A. Zevenbergen, "Urbanization and urban land use efficiency: Evidence from regional and Addis Ababa satellite cities, Ethiopia," *Habitat Int.*, vol. 117, 2021, Art. no. 102437.

[2] UN Department of Economic and Social Affairs, "World urbanization prospects: The 2018 revision," New York, NY, USA, Tech. Rep. ST/ESA/SER.A/420, 2018. [Online]. Available: https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf

[3] H. H. Kim, "Urban heat island," *Int. J. Remote Sens.*, vol. 13, no. 12, pp. 2319–2336, 1992.

[4] T. R. Oke, "City size and the urban heat island," *Atmospheric Environ.*, vol. 7, no. 8, pp. 769–779, 1973.

[5] H. E. Landsberg, "The urban climate," *Int. Geophys. Ser.*, vol. 28, 1981, Art. no. 275.

[6] T. R. Oke *Boundary Layer Climates*, 2nd ed. London, U.K.: Routledge, 1987.

[7] L. Howard, *The Climate of London: Deduced From Meteorological Observations Made in the Metropolis and at Various Places Around It, Harvey and Darton*. New York, NY, USA: New York Public Library, 1833.

[8] B. Huang, J. Wang, H. Song, D. Fu, and K. Wong, "Generating high spatiotemporal resolution land surface temperature for urban heat island monitoring," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1011–1015, Sep. 2013.

[9] S. Wang et al., "Downscaling land surface temperature based on nonlinear geographically weighted regressive model over urban areas," *Remote Sens.*, vol. 13, no. 8, 2021, Art. no. 1580.

[10] J. Nichol, "Remote sensing of urban heat islands by day and night," *Photogrammetric Eng. Remote Sens.*, vol. 71, no. 5, pp. 613–621, 2005.

[11] D. R. Streutker, "A remote sensing study of the urban heat island of Houston, Texas," *Int. J. Remote Sens.*, vol. 23, no. 13, pp. 2595–2608, 2002.

[12] Z.-L. Li et al., "Satellite-derived land surface temperature: Current status and perspectives," *Remote Sens. Environ.*, vol. 131, pp. 14–37, 2013.

[13] Y. Zhang and J. Cheng, "Spatiotemporal analysis of urban heat island using multisource remote sensing data: A case study in Hangzhou, China," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 9, pp. 3317–3326, Sep. 2019.

[14] M. C. Anderson et al., "Mapping daily evapotranspiration at field to continental scales using geostationary and polar orbiting satellite imagery," *Hydrol. Earth Syst. Sci.*, vol. 15, no. 1, pp. 223–239, 2011.

[15] Y. Bai, M. S. Wong, W.-Z. Shi, L.-X. Wu, and K. Qin, "Advancing of land surface temperature retrieval using extreme learning machine and spatio-temporal adaptive data fusion algorithm," *Remote Sens.*, vol. 7, no. 4, pp. 4424–4441, 2015.

[16] W. Zhan et al., "Disaggregation of remotely sensed land surface temperature: A new dynamic methodology," *J. Geophys. Res. Atmos.*, vol. 121, no. 18, pp. 10538–10554, 2016.

[17] X. Zhu, F. Cai, J. Tian, and T. K.-A. Williams, "Spatiotemporal fusion of multisource remote sensing data: Literature survey, taxonomy, principles, applications, and future directions," *Remote Sens.*, vol. 10, no. 4, 2018, Art. no. 527.

[18] J. R. Irons, J. L. Dwyer, and J. A. Barsi, "The next Landsat satellite: The Landsat data continuity mission," *Remote Sens. Environ.*, vol. 122, pp. 11–21, 2012.

[19] C. Justice et al., "An overview of MODIS land data processing and product status," *Remote Sens. Environ.*, vol. 83, no. 1/2, pp. 3–15, 2002.

[20] D. P. Roy et al., "Multi-temporal MODIS–Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data," *Remote Sens. Environ.*, vol. 112, no. 6, pp. 3112–3130, 2008.

[21] J. Ju and D. P. Roy, "The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally," *Remote Sens. Environ.*, vol. 112, no. 3, pp. 1196–1211, 2008.

[22] M. Anderson et al., "Mapping daily evapotranspiration at field to global scales using geostationary and polar orbiting satellite imagery," *Hydrol. Earth Syst. Sci. Discuss.*, vol. 7, pp. 5957–5990, 2010.

[23] P. Bartkowiak, M. Castelli, and C. Notarnicola, "Downscaling land surface temperature from MODIS dataset with random forest approach over alpine vegetated areas," *Remote Sens.*, vol. 11, no. 11, 2019, Art. no. 1319.

[24] H. Bo, W. Juan, S. Huihui, F. Dongjie, and W. KwanKit, "Generating high spatiotemporal resolution land surface temperature for urban heat island monitoring," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1011–1015, Sep. 2013.

[25] F. Gao, J. Masek, M. Schwaller, and F. Hall, "On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2207–2218, Aug. 2006.

[26] W. Zhan et al., "Disaggregation of remotely sensed land surface temperature: Literature survey, taxonomy, issues, and caveats," *Remote Sens. Environ.*, vol. 131, pp. 119–139, 2013.

[27] Q. Mao, J. Peng, and Y. Wang, "Resolution enhancement of remotely sensed land surface temperature: Current status and perspectives," *Remote Sens.*, vol. 13, no. 7, 2021, Art. no. 1306.

[28] Q. Zhang, N. Wang., J. Cheng., and S. Xu., "A stepwise downscaling method for generating high-resolution land surface temperature from AMSR-E data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5669–5681, Sep. 2020.

[29] C. Hutengs and M. Vohland, "Downscaling land surface temperatures at regional scales with random forest regression," *Remote Sens. Environ.*, vol. 178, pp. 127–141, 2016.

[30] H. Ebrahimy and M. Azadbakht, "Downscaling MODIS land surface temperature over a heterogeneous area: An investigation of machine learning techniques, feature selection, and impacts of mixed pixels," *Comput. Geosci.*, vol. 124, pp. 93–102, 2019.

[31] P. Wu et al., "Spatially continuous and high-resolution land surface temperature product generation: A review of reconstruction and spatiotemporal fusion techniques," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 3, pp. 112–137, Sep. 2021.

[32] R. Wang, W. Gao, and W. Peng, "Downscale MODIS land surface temperature based on three different models to analyze surface urban heat island: A case study of Hangzhou," *Remote Sens.*, vol. 12, no. 13, 2020, Art. no. 2134.

[33] C. Yang, Q. Zhan, Y. Lv, and H. Liu, "Downscaling land surface temperature using multiscale geographically weighted regression over heterogeneous landscapes in Wuhan, China," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 5213–5222, Dec. 2019.

[34] N. Agam, W. P. Kustas, M. C. Anderson, F. Li, and C. M. Neale, "A vegetation index based technique for spatial sharpening of thermal imagery," *Remote Sens. Environ.*, vol. 107, no. 4, pp. 545–558, 2007.

[35] V. Bindhu, B. Narasimhan, and K. Sudheer, "Development and verification of a non-linear disaggregation method (NL-DisTrad) to downscale MODIS land surface temperature to the spatial scale of Landsat thermal data to estimate evapotranspiration," *Remote Sens. Environ.*, vol. 135, pp. 118–129, 2013.

[36] W. P. Kustas, J. M. Norman, M. C. Anderson, and A. N. French, "Estimating subpixel surface temperatures and energy fluxes from the vegetation index–radiometric temperature relationship," *Remote Sens. Environ.*, vol. 85, no. 4, pp. 429–440, 2003.

[37] S. Mukherjee, P. Joshi, and R. Garg, "Analysis of urban built-up areas and surface urban heat island using downscaled MODIS derived land surface temperature data," *Geocarto Int.*, vol. 32, no. 8, pp. 900–918, 2017.

[38] H. Govil, S. Guha, A. Dey, and N. Gill, "Seasonal evaluation of downscaled land surface temperature: A case study in a humid tropical city," *Heliyon*, vol. 5, no. 6, 2019, Art. no. e01923.

[39] P. Wu, H. Shen, T. Ai, and Y. Liu, "Land-surface temperature retrieval at high spatial and temporal resolutions based on multi-sensor fusion," *Int. J. Digit. Earth*, vol. 6, no. sup1, pp. 113–133, 2013.

[40] Z. Yin et al., "Spatiotemporal fusion of land surface temperature based on a convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1808–1822, Feb. 2020.

[41] X. Zhu, J. Chen, F. Gao, X. Chen, and J. G. Masek, "An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions," *Remote Sens. Environ.*, vol. 114, no. 11, pp. 2610–2623, 2010.

[42] F. Gao et al., "Fusing Landsat and MODIS data for vegetation monitoring," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 47–60, Mar. 2015.

[43] B. Chen, B. Huang, and B. Xu, "Comparison of spatiotemporal fusion models: A review," *Remote Sens.*, vol. 7, no. 2, pp. 1798–1835, 2015.

[44] B. Huang and H. Song, "Spatiotemporal reflectance fusion via sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3707–3716, Oct. 2012.

[45] Q. Wang and P. M. Atkinson, "Spatio-temporal fusion for daily Sentinel-2 images," *Remote Sens. Environ.*, vol. 204, pp. 31–42, 2018.

[46] Q. Wang, Y. Tang, X. Tong, and P. M. Atkinson, "Virtual image pair-based spatio-temporal fusion," *Remote Sens. Environ.*, vol. 249, 2020, Art. no. 112009.

[47] T. Hilker et al., "A new data fusion model for high spatial-and temporal-resolution mapping of forest disturbance based on Landsat and MODIS," *Remote Sens. Environ.*, vol. 113, no. 8, pp. 1613–1627, 2009.

[48] J. Ma, W. Zhang, A. Marinoni, L. Gao, and B. Zhang, "An improved spatial and temporal reflectance unmixing model to synthesize time series of Landsat-like images," *Remote Sens.*, vol. 10, no. 9, 2018, Art. no. 1388.

[49] D. Guo, W. Shi, M. Hao, and X. Zhu, "FSDAF 2.0: Improving the performance of retrieving land cover changes and preserving spatial details," *Remote Sens. Environ.*, vol. 248, 2020, Art. no. 111973.

[50] H. Gao et al., "cuFSDAF: An enhanced flexible spatiotemporal data fusion algorithm parallelized using graphics processing units," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, May 2022, Art. no. 4403016.

[51] H. Shen, L. Huang, L. Zhang, P. Wu, and C. Zeng, "Long-term and fine-scale satellite monitoring of the urban heat island effect by the fusion of multi-temporal and multi-sensor remote sensed data: A 26-year case study of the city of Wuhan in China," *Remote Sens. Environ.*, vol. 172, pp. 109–125, 2016.

[52] H. Xia, Y. Chen, Y. Li, and J. Quan, "Combining kernel-driven and fusion-based methods to generate daily high-spatial-resolution land surface temperatures," *Remote Sens. Environ.*, vol. 224, pp. 259–274, 2019.

[53] D. Long et al., "Generation of MODIS-like land surface temperatures under all-weather conditions based on a data fusion approach," *Remote Sens. Environ.*, vol. 246, 2020, Art. no. 111863.

[54] Y. Lu, P. Wu, X. Ma, H. Yang, and Y. Wu, "Monitoring seasonal and diurnal surface urban heat islands variations using Landsat-scale data in Hefei, China, 2000–2017," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 6410–6423, 2020.

[55] X. Zhu et al., "A framework for generating high spatiotemporal resolution land surface temperature in heterogeneous areas," *Remote Sens.*, vol. 13, no. 19, 2021, Art. no. 3885.

[56] H. Liu and Q. Weng, "Enhancing temporal resolution of satellite imagery for public health studies: A case study of West Nile virus outbreak in Los Angeles in 2007," *Remote Sens. Environ.*, vol. 117, pp. 57–71, 2012.

[57] Q. Weng, P. Fu, and F. Gao, "Generating daily land surface temperature at Landsat resolution by fusing Landsat and MODIS data," *Remote Sens. Environ.*, vol. 145, pp. 55–67, 2014.

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

[59] J. Quan, W. Zhan, T. Ma, Y. Du, Z. Guo, and B. Qin, "An integrated model for generating hourly Landsat-like land surface temperatures over heterogeneous landscapes," *Remote Sens. Environ.*, vol. 206, pp. 403–423, 2018.

[60] Q. Zhang, J. Cheng, and N. Wang, "Fusion of all-weather land surface temperature from AMSR-E and MODIS data using random forest regression," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2021, Art. no. 2502705.

[61] S. Xu, J. Cheng, and Q. Zhang, "A random forest-based data fusion method for obtaining all-weather land surface temperature with high spatial resolution," *Remote Sens.*, vol. 13, no. 11, 2021, Art. no. 2211.

[62] Y. Zhao, B. Huang, and H. Song, "A robust adaptive spatial and temporal image fusion model for complex land surface changes," *Remote Sens. Environ.*, vol. 208, pp. 42–62, 2018.

[63] X. Zhu, E. H. Helmer, F. Gao, D. Liu, J. Chen, and M. A. Lefsky, "A flexible spatiotemporal method for fusing satellite images with different resolutions," *Remote Sens. Environ.*, vol. 172, pp. 165–177, 2016.

[64] L. Meng, H. Liu, S. L. Ustin, and X. Zhang, "Assessment of FSDAF accuracy on cotton yield estimation using different MODIS products and Landsat based on the mixed degree index with different surroundings," *Sensors*, vol. 21, no. 15, 2021, Art. no. 5184.

[65] M. Kaffash and H. S. Nejad, "Spatio-temporal fusion of Landsat and MODIS land surface temperature data using FSDAF algorithm," *J. Water Soil Sci.*, vol. 25, no. 2, pp. 45–62, 2021.

[66] C. Shi et al., "A comprehensive and automated fusion method: The enhanced flexible spatiotemporal data fusion model for monitoring dynamic changes of land surface," *Appl. Sci.*, vol. 9, no. 18, 2019, Art. no. 3693.

[67] Z. Wan and J. Dozier, "A generalized split-window algorithm for retrieving land-surface temperature from space," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 4, pp. 892–905, Jul. 1996.

[68] Z. Wan, "New refinements and validation of the MODIS land-surface temperature/emissivity products," *Remote Sens. Environ.*, vol. 112, no. 1, pp. 59–74, 2008.

[69] S. Duan et al., "Validation of collection 6 MODIS land surface temperature product using in situ measurements," *Remote Sens. Environ.*, vol. 9, no. 225, pp. 16–29, 2019.

[70] H. Li et al., "Temperature-based and radiance-based validation of the collection 6 MYD11 and MYD21 land surface temperature products over barren surfaces in northwestern China," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1794–1807, Feb. 2021.

[71] J. C. Jiménez-Muñoz and J. A. Sobrino, "A generalized single-channel method for retrieving land surface temperature from remote sensing data," *J. Geophys. Res.: Atmos.*, vol. 108, no. D22, 2003, Art. no. D08112.

[72] X. Yu, X. Guo, and Z. Wu, "Land surface temperature retrieval from Landsat 8 TIRS—Comparison between radiative transfer equation-based method, split window algorithm and single channel method," *Remote Sens.*, vol. 6, no. 10, pp. 9829–9852, 2014.

[73] J. Cheng et al., "Generating the 30-m land surface temperature product over continental China and USA from Landsat 5/7/8 data," *Sci. Remote Sens.*, vol. 4, 2021, Art. no. 100032.

[74] D. Sousa and C. Small, "Global cross-calibration of Landsat spectral mixture models," *Remote Sens. Environ.*, vol. 192, pp. 139–149, 2017.

[75] D. C. Heinz, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 529–545, Mar. 2001.

[76] H. Fujisada, "Design and performance of ASTER instrument," *Proc. Int. Soc. Opt. Eng.*, vol. 2583, pp. 16–25, 1995.

[77] W.-K. Ma et al., "A signal processing perspective on hyperspectral unmixing: Insights from remote sensing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 67–81, Jan. 2014.

[78] W. R. Tobler, "A computer movie simulating urban growth in the Detroit region," *Econ. Geography*, vol. 46, no. sup1, pp. 234–240, 1970.

[79] G. Y. Lu and D. W. Wong, "An adaptive inverse-distance weighting spatial interpolation technique," *Comput. Geosci.*, vol. 34, no. 9, pp. 1044–1055, 2008.

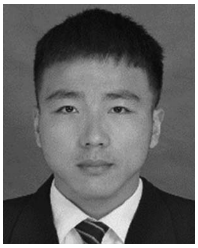
[80] M. Liu et al., "An improved flexible spatiotemporal DAta fusion (IFSDAF) method for producing high spatiotemporal resolution normalized difference vegetation index time series," *Remote Sens. Environ.*, vol. 227, pp. 74–89, 2019.

[81] X. Zhu et al., "A novel framework to assess all-round performances of spatiotemporal fusion models," *Remote Sens. Environ.*, vol. 274, 2022, Art. no. 113002.

[82] X. Li et al., "SFSDAF: An enhanced FSDAF that incorporates sub-pixel class fraction change information for spatio-temporal image fusion," *Remote Sens. Environ.*, vol. 237, 2020, Art. no. 111537.

[83] Q. Wang, Y. Zhang, A. O. Onojeghuo, X. Zhu, and P. M. Atkinson, "Enhancing spatio-temporal fusion of MODIS and Landsat data by incorporating 250 m MODIS data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 9, pp. 4116–4123, Sep. 2017.

[84] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 21–36, May 2003.



Chenlie Shi received the M.S. degree in cartography and geography information system in 2021 from Northwest University, Xi'an, China, where he is currently working toward the Ph.D. degree in geography.

His research interests include the retrieval of land surface temperature, spatiotemporal fusion, downscaling of land surface temperature, and blending of remotely sensed land surface temperature and reanalysis.

Ninglian Wang received the M.S. and Ph.D. degrees from the Cold and Arid Regions Environmental and Engineering Research Institute of Chinese Academy of Sciences (CAS), Beijing, China, in 1991 and 2001, respectively.

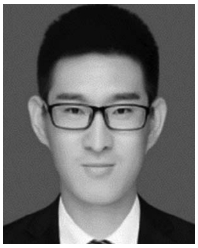
He performed a Postdoctoral research work with Byrd Polar and Climate Research Center, Ohio State University, Columbus, OH, USA, during 2002–2003. He is currently a Professor with Northwest University, Xi'an, China. His research interests include climatic and environmental records in ice core, glacier changes and water resources, cryosphere, and global change.

Quan Zhang received the M.S. degree from Northwest University, Xi'an, China, in 2015, and the Ph.D. degree from Beijing Normal University, Beijing, China, in 2019, both in cartography and geography information system.

He is currently a Postdoctor with the College of Urban and Environmental Science, Northwest University. His research interests include the retrieval, validation, scaling, and blending of remotely sensed land surface temperature and emissivity products.

Zhuang Liu received the M.S. degree in cartography and geography information system from Northwest University, Xi'an, China, in 2019.

He is currently an Engineer with the State Key Laboratory of Geo-Information Engineering, Xi'an.



Xinming Zhu received the M.S. degree from Northwest University, Xi'an, China, in 2018. He is currently working toward the Ph.D. degree with the University of Chinese Academy of Sciences, Beijing, China, both in cartography and geography information system.

His research interests include the retrieval of land surface temperature and atmospheric parameters from satellite data and passive microwave remote sensing.