A Rigorously-Incremental Spatiotemporal Data Fusion Method for Fusing Remote Sensing Images

Weipeng Jing, *Member, IEEE*, Tongtong Lou, Zeyu Wang, Weitao Zou, Zekun Xu, Linda Mohaisen, Chao Li, and Jian Wang

***Abstract*—The spatiotemporal remote sensing images have significant importance in forest ecological monitoring, forest carbon management, and other related fields. Spatiotemporal data fusion technology of remote sensing images combines high spatiotemporal and high temporal resolution images to address the current limitation of single sensors in obtaining high spatiotemporal resolution. This technology has gained widespread attention in recent years. However, the current models still exhibit some shortcomings in dealing with land cover changes, such as poor clustering results, inaccurate incremental spatiotemporal calculations, and sensor differences. In this article, we propose a rigorously-incremental spatiotemporal data fusion method for fusing remote sensing images with different resolutions to address the aforementioned problems. The proposed method utilizes the particle swarm optimization Gaussian mixture model to extract endmembers and establishes a linear relationship between sensors to obtain accurate time increments. Furthermore, bicubic interpolation is used instead of thin plate spline interpolation for spatial interpolation, and also support vector regression is used to calculate weights for obtaining a weighted sum of temporal and spatial increments. In addition, sensor errors are allocated to the calculation of residuals. The experimental results show the efficacy of the proposed algorithm for fusing fine image Landsat with coarse image MODIS data and conclude that the proposed algorithm presents a better solution for heterogeneous data with strong phenological changes and regions with changes in surface types, which provides a better solution for remote sensing image fusion and, hence, improves the accuracy, stability, and robustness of data fusion.**

***Index Terms*—Registration error, rigorously-incremental spatiotemporal data fusion (RISDAF), spatiotemporal fusion, spatiotemporal increment, spectral unmixing.**

I. Introduction

FORESTS, as the largest carbon pool in terrestrial ecosystems, have the characteristics of wide distribution and diverse types [1]. Therefore, accurate and comprehensive forest resource monitoring is significant for maintaining the ecosystem's capacity, but it also faces significant technical challenges [2], [3], [4]. With the high development of satellite sensors in the last decades, remote sensing images have provided a powerful technical support for the dynamic monitoring of large-scale forest resources and accurate forest observation data [5] and have introduced many advantages such as wide coverage and high precision [6], [7]. However, due to hardware and technological limitations of sensors, a single-satellite product cannot obtain remote sensing images with both high temporal and high spatial resolution at the same time [8]. The emergence of spatiotemporal fusion technology provides an effective solution for processing remote sensing images [9].

The spatiotemporal fusion algorithm for remote sensing images has been developed since the 1990s and has made a significant progress in the past decade [10]. The core idea behind spatiotemporal fusion is to fuse remote sensing images from multiple sensors to compensate their respective shortcomings and generate high spatiotemporal resolution remote sensing images [11], as illustrated in Fig. 1. For example, the reflectivity images obtained by Landsat series, advanced land observation satellite (ALOS), and GF series satellites have a good spatial resolution of 3–30 m [12]. However, these images require long satellite revisit cycles. The natural obscuration of clouds and complex terrain limits the variety of high spatial resolution images in terms of fast surface features (these images are referred as "fine images" in this article). While satellites such as moderate-resolution imaging spectroradiometer (MODIS) have a revisit time of one day (these images are referred as "coarse images" in this article), and their spatial resolution is low, from 250 m to 1 km, which cannot capture spatial details [13]. Therefore, generating a remote sensing image with a resolution of 30 m and a revisit period of 1 day by fusing fine and coarse images can improve data accuracy, completeness, and timeliness. This provides a reliable data support for accurately describing and simulating changes on the Earth's surface [14].

Manuscript received 25 April 2023; revised 13 June 2023; accepted 12 July 2023. Date of publication 18 July 2023; date of current version 28 July 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 32171777 and Grant 32271865, in part by the Central University Basic Research Fund of China under Grant 2572022AW59, in part by the Science and Technology Program of Inner Mongolia Autonomous Region under Grant 2022YFSJ0037, and in part by the National Earth Observation Data Center under Grant NODAOP2022001. *(Corresponding author: Chao Li.)*

Weipeng Jing, Tongtong Lou, Weitao Zou, Zekun Xu, and Chao Li are with the College of Computer and Control Engineering, Northeast Forestry University, Harbin 150040, China (e-mail: jwp@nefu.edu.cn; ltt0221@nefu.edu.cn; zouweitao1996@nefu.edu.cn; xuzekun@nefu.edu.cn; lichaonefuzyz@nefu.edu.cn).

Zeyu Wang is with the College of Computer and Control Engineering, Northeast Forestry University, Harbin 150040, China, and also with the School of Information and Intelligence Engineering, University of Sanya, Sanya 572000, China (e-mail: wzy@nefu.edu.cn).

Linda Mohaisen is with the Department of Information Technology, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia (e-mail: Lmohaisen@kau.edu.sa).

Jian Wang is with the Aerospace Information Research Institute, Chinese Academy of Sciences (CAS), Beijing 100094, China (e-mail: wangjian@radi.ac.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3296468

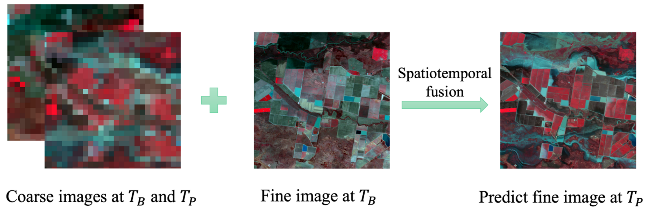

Fig. 1. Schematic diagram of spatiotemporal fusion. T_B stands for the base date, and T_P stands for the prediction date.

The traditional spatiotemporal fusion algorithms predominantly fall into three categories: 1) weight-based, 2) spectral-unmixing-based, and 3) learning-based. The weight-based and spectral-unmixing-based are two categories of the first spatiotemporal fusion algorithm [15]. The weight-based methods calculate changes in surface emissivity by weighing the pixel values of remote sensing images to predict corresponding time images [16]. Typical-weight-based methods include the SpatioTemporal Adaptive Reflectance Fusion Model (STARFM) [17], Enhanced STARFM (ESTARFM) [18], Fit-FC [19], among others. These methods improve the accuracy of the weighted spatiotemporal fusion algorithm, and they work efficiently in homogeneous areas without requiring external data support while also having high fusion efficiency. However, they do not perform well in heterogeneous areas and have poor details in reconstructed images [20], [21]. Whereas the spectral unmixing based on mixed pixel decomposition is proposed to predict unknown fine images by spectral separation and endmember abundance calculation of coarse images [22], [23], [24]. However, a significant disadvantage of this method is that it is carried out under the assumption that the land type does not change, which cannot meet the occurrence of sudden events. On the other hand, learning-based spatiotemporal fusion algorithms are trained using existing datasets and treat the prediction of high-resolution images as the generation of supervised superresolution images [25], [26], [27]. Deep-learning-based methods have powerful feature extraction capabilities, and models such as the deep convolutional spatiotemporal fusion network [28] and depth networks based on spatiotemporal data fusion [29] have improved the accuracy of spatiotemporal fusion by optimizing the network and loss function. However, most deep-learning-based models are trained based on idealized datasets, which require at least three pairs of images as input [30]. This ignores the difficulty of obtaining ideal data in realistic studies due to weather and cloudiness. Furthermore, in impact fusion, the architecture and loss functions should be fully applied in the algorithm, which is computationally intensive [31].

To address the deficiencies mentioned above in spatiotemporal fusion algorithms, researchers have started to combine and optimize existing algorithms to enhance their generality by integrating the advantages of two or more models. Better results have been achieved in heterogeneous regions and regions with sudden changes in land type [32]. Zhu et al. [33] proposed a Flexible Spatiotemporal DAta Fusion (FSDAF) algorithm. FSDAF combines STARFM [17] and the unmixing-based data fusion [34] algorithms and integrates thin plate spline (TPS). Compared to other spatiotemporal fusion algorithms, FSDAF only requires the input of one pair of fine and coarse images at T_B and the coarse images at the moment of T_P, which reduces the input data needed [35]. FSDAF also performs well in heterogeneous data, as it can capture more information on physical changes in coarse images and has a high fusion accuracy. Recently, several improved models based on FSDAF have been developed. For instance, Improved Flexible Spatiotemporal DAta Fusion (IFSDAF) [36], which predicts the normalized difference vegetation index (NDVI); subpixel class fraction-based flexible spatiotemporal data fusion [37], which extracts endmember abundance based on subpixel information that consequently improves the accuracy of heterogeneous data prediction; FSDAF 2.0 [38], which solves the problem of boundary pixel mixing and effectively restores land cover changes; and object-based spatiotemporal fusion model [39] which combines nonpixel-based image segmentation with a weighting function that achieves good results in homogeneous physical changes. Like FSDAF, all the aforementioned algorithms assume that there is no change in land cover type during the temporal prediction phase and calculate the change in fine images directly from the coarse image change. However, this approach leads to bias in the prediction results [40], [41].

In conclusion, the hybrid spatiotemporal data fusion model has proven to be effective in dealing with land cover changes and has become the mainstream approach for remote sensing image fusion. It has been successfully applied in various fields such as land surface temperature monitoring [42], [43], vegetation coverage detection [44], [45], and forest resource change monitoring [46]. These applications demonstrate the potential of hybrid algorithms in addressing practical problems in remote sensing.

Currently, the hybrid spatiotemporal fusion algorithm, represented by FSDAF, still exhibits some shortcomings, which can be classified into the following aspects.

1) *In terms of temporal change dimension:* Spectral unmixing can roughly preserve the surface cover structure. However, poor clustering results due to randomly selected initial values in the clustering algorithm can seriously affect the accuracy of mixed image element decomposition, leading to reduced prediction accuracy and stability of the model. Additionally, this method tends to ignore the details inside the image. The predicted time of FSDAF assumes that there is no land type change occurring between T_B and T_P, and also the change of each end element is the same in the coarse resolution image. However, this assumption can introduce a great uncertainty to the time prediction.
2) *In terms of spatial variation dimension:* The rapid growth of forest crops and human activities can cause significant spectral changes on the surface. TPS interpolators are commonly used in FSDAF and some existing algorithms can perform well in homogeneous regions. However, their interpolation results are often too smooth in heterogeneous regions, which ignores important spatial details. Additionally, the assumption that fusion errors come from homogeneous landscapes in FSDAF lacks the theoretical basis.

3) *In terms of sensor differences:* Existing spatiotemporal fusion algorithms do not fully consider the differences between sensors in fine and coarse images and their impact on fusion. This issue has attracted many scholars' attention, and a linear model has been developed to address the problem of sensor differences between two sensors [47]. However, this solution does not completely solve the problem of alignment errors.

To deal with the above difficulties and problems, this article proposes a rigorously-incremental spatiotemporal data fusion (RISDAF) method for fusing remote sensing images with different resolutions, which is verified by the fusion of fine images and coarse image. The proposed algorithm provides a better solution for heterogeneous data with strong phenological changes and areas with changes in surface types. Moreover, the proposed algorithm exhibits good stability and robustness.

In this article, Landsat and MODIS data are adopted as fine and coarse images, respectively. The main contributions of this work are concluded as follows.

1) To address the issue of inaccurate prediction of temporal changes, we use a particle swarm optimized Gaussian mixture model for end element extraction called particle swarm optimization Gaussian mixture model (PSO-GMM). This method overcomes problems associated with poor clusterings, such as unbalanced samples and noncylindrical data, and improves the accuracy and adaptability of mixed image element decomposition. Additionally, the linear regression algorithm is used to correct sensor errors. Furthermore, the difference between fine and coarse pixels is standardized when allocating time changes to fine pixels.

2 To overcome the issue of inaccurate prediction of spatial changes, we use bicubic interpolation for spatial interpolation instead of TPS interpolation, which is commonly used for spatial prediction. This method preserves the connectedness of spatial increments and improves computational speed, scalability, and smoothness. Furthermore, the article does not incorporate temporal and spatial prediction results directly into the computational process. Rather, a weighting algorithm is employed to combine the weights of temporal and spatial increments. These weights are calculated via the support vector regression (SVR) algorithm, which enhances the robustness and accuracy of data fusion.

3) To resolve the issue of sensor errors, this article introduces sensor errors into the residuals and assigns them to each fine pixel. This correction improves the spatial distribution of image fusion results for reliability and reduces the impact of sensor errors on image fusion.

The rest of this article is organized as follows. Section II presents the specific architecture and implementation process of the proposed algorithm RISDAF. Section III describes the dataset and experimental settings, while Section IV presents the experimental results and analysis. Section V provides a discussion and shows some necessary intermediate experimental results during the experiments. Finally, Section V concludes this article.

II. Methods

The proposed RISDAF algorithm takes the coarse and fine images at the T_B moment and the coarse image at the T_P time as inputs to predict the fine image at the T_P time. Here, T_B and T_P represent the base date and predicted date, respectively. The proposed model aims to address the following problems: inaccurate endmember division in unsupervised classification during the spectral unmixing process, the accuracy deviation of fusion results caused by sensor errors, and the difference in spectral changes caused by strong spatial changes. The overall idea can be summarized as shown in (1). The fine image pixel value F_P at the moment T_P equals the sum of the fine image pixel value F_B at the moment T_B, the increment ΔF^{ST} and the residual ε is

$$F_P = F_B + \Delta F^{ST} + \varepsilon. \tag{1}$$

The proposed model can be divided into four main steps as follows. The first step is the temporal prediction based on spectral unmixing. The second step is the spatial variation based on land cover combined with the temporal prediction results by a weighting algorithm. The third step is the residual correction that enhances the fusion accuracy by introducing the residual r_i and sensor error r_e. Finally, the last step is the enhanced neighborhood prediction by spatial filtering. Fig. 2 shows the flowchart of the proposed RISDAF.

A. Temporal Increment Prediction Based on Spectral Unmixing

1) Mixed Pixel Decomposition Based on PSO-GMM: The spectral unmixing of remote sensing images acquired by land satellites is a challenging task due to multiple surface heterogeneous coverage types within a single pixel. In this regard, we perform endmember determination and image boundary extraction on the fine image at T_B prior to spectral unmixing. However, traditional clustering algorithms used in spectral unmixing suffer from poor stability and sensitivity to cluster center selection. To address this issue, the proposed approach utilizes a PSO-GMM to extract endmembers. The PSO algorithm optimizes the objective function in GMM clustering while adjusting the position $pbest_i = (p_{i1}, p_{i2}, \ldots, p_{iD})$ and velocity $\nu_i = (\nu_{i1}, \nu_{i2}, \ldots, \nu_{iD})$ of each particle by t iterations to ensure convergence of the algorithm within a certain range. PSO-GMM addresses the problem of poor clustering for unbalanced samples and noncylindrical data, which are the most challenging tasks for the traditional clustering algorithms such as ISODATA and *K*-means algorithms. Using PSO-GMM to decompose the surface reflectance, the endmember hard classification map is generated, and the abundance value of each endmember is calculated for each fine pixel. This method reduces the problem of poor clustering results caused by the random selection of initial values in the GMM clustering algorithm and improves the accuracy of endmember extraction. The abundance value of each endmember is expressed as shown in

$$A_C^B(x_i, y_i) = N_C(x_i, y_i)/k \tag{2}$$

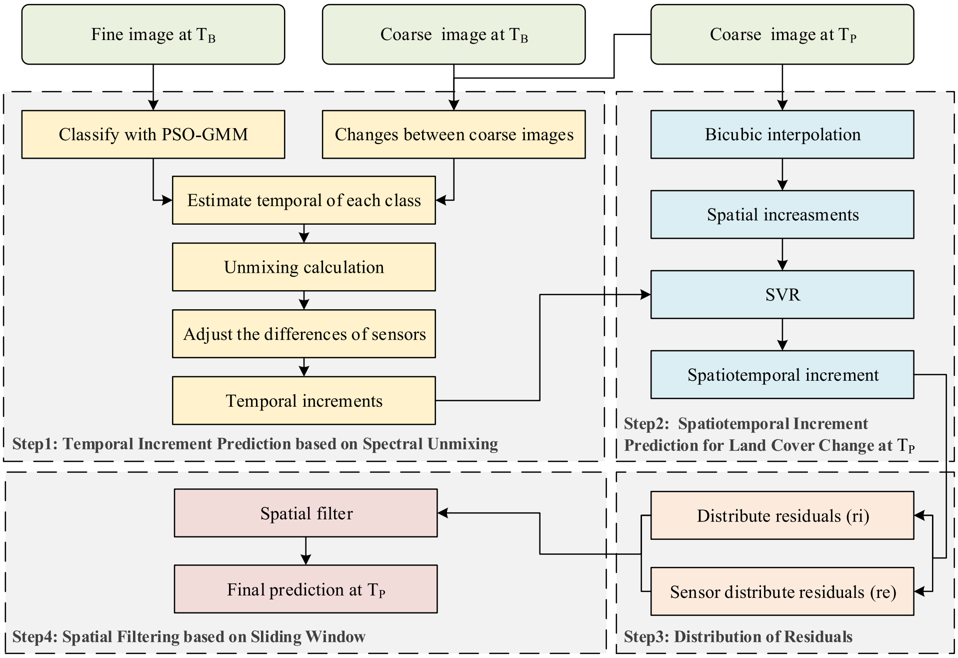
Fine image at T_B
Coarse image at T_B
Coarse image at T_P
Classify with PSO-GMM
Changes between coarse images
Estimate temporal of each class
Unmixing calculation
Adjust the differences of sensors
Temporal increments
Step1: Temporal Increment Prediction based on Spectral Unmixing
Bicubic interpolation
Spatial increasments
SVR
Spatiotemporal increment
Step2: Spatiotemporal Increment Prediction for Land Cover Change at T_P
Spatial filter
Final prediction at T_P
Step4: Spatial Filtering based on Sliding Window
Distribute residuals (ri)
Sensor distribute residuals (re)
Step3: Distribution of Residuals

Fig. 2. Flowchart of the proposed RISDAF method.

where $N_C(x_i, y_i)$ is the number of fine pixels belonging to m class at the coarse pixel (x_i, y_i), and k is the number of fine pixels in a coarse pixel.

The RISDAF algorithm assumes that the land type remains unchangeable during the time prediction period. Based on the mixed pixel theory, we assume that the pixel value of the image is a linear combination of the endmember values and their corresponding abundances. Therefore, the pixel values of Landsat and MODIS at T_B can be expressed as

$$F_B(x_{ij}, y_{ij}, b) = \sum_{m=1}^{n} A_F^B(x_{ij}, y_{ij}, m) \times E_F^B(m, b) + \varphi \tag{3}$$

$$C_B(x_i, y_{i,}, b) = \frac{1}{k}\sum_{i=1}^{1} F_B(x_{ij}, y_{ij}, b) \tag{4}$$

where b represents the band b, n represents the endmembers of the fine image at T_P, m represents the mth endmember according to the linear mixed model. A represents the abundance value of the endmember, E represents the reflectivity of each endmember, and k represents a coarse-resolution pixel containing k fine-resolution pixels. As shown in the following equation, in the mixed pixel decomposition, the time change of the coarse pixel is the weighted sum of all the category changes that it contains:

$$C_B = \frac{1}{k}\sum_{m=1}^{n} A_F^B(x_{ij}, y_{ij}, b) \times E_F^B(m, b) + \varphi. \tag{5}$$

2) Adjust the Differences of Sensors: Sensor errors can arise from various sources, including the specific design of the sensor, such as its bandwidth, imaging angle, and spectral response function, among others. These errors can lead to nonuniform reception of images that result in varying capabilities to capture land surface information. Due to the inherent of sensor differences, errors are inevitable during the imaging stage. Therefore, correcting the differences between the imaging of the two types of sensors is crucial. In this article, we propose a linear model to adjust the relationship between the two types of sensors, which standardizes the differences and corrects the sensor errors, as follows:

$$\Delta E_F(m, b) = a \times \Delta E_C(m, b). \tag{6}$$

3) Temporal Increment Prediction: In FSDAF, the time change increment from T_B to T_P is assumed to be calculated directly as the difference between the coarse images at the two-time points, which can lead to significant uncertainty in the time prediction results. To address this issue, we propose to employ a subpixel-based approach for predicting the time of fine images at T_P. The proposed RISDAF algorithm assumes that there is no change in land cover during the spatiotemporal fusion process, which means that endmembers and abundances of Landsat pixels remain constant. The time change from T_B to T_P can then be calculated as follows:

$$\Delta F^T = \sum_{m=1}^{n} A_F^B(x_{ij}, y_{ij}, m) \times \Delta E_F(m, b). \tag{7}$$

Since A is only known in (7), the change in time cannot be calculated. However, the time change of the coarse image can be expressed as follows:

$$\Delta C^T = \sum_{m=1}^{n} A_C^B(x_{ij}, y_{ij}, m) \times \Delta E_C(m, b). \tag{8}$$

In this article, the SVR algorithm is selected instead of the traditional least squares method to solve for ΔE_C, which improves the robustness of abnormal data. The preliminary results of the time increment prediction can be shown as in

$$\Delta T = \Delta F^{T} = \sum_{m=1}^{n} A_F^B \left(x_{ij}, y_{ij}, m\right) \times a \times \Delta E_C. \quad (9)$$

B. Spatiotemporal Increment Prediction for Land Cover Change at T_P

If the land cover category within the coarse pixel scale undergoes significant changes, it is demonstrated that the coarse image contains information that can provide insight into the changes in vegetation cover within the image. The incremental value of spatial dependence can be estimated by interpolation as follows:

$$\Delta S \left(x_{ij}, y_{ij}, b\right) = F_P^I \left(x_{ij}, y_{ij}, b\right) - F_B^I \left(x_{ij}, y_{ij}, b\right). \quad (10)$$

Therefore, the change information can be directly obtained. However, previous studies have shown that the TPS interpolator performs better in regions with similar characteristics but results in oversmoothed interpolations in heterogeneous regions, disregarding important spatial details. To address this issue, we employ the bicubic interpolation method to interpolate the coarse image at time T_P to the fine scale, and hence obtaining spatial prediction results for the time of T_P. This method enhances the information contained in the transition resolution image at T_P and also improves the accuracy of the spatial prediction results.

Since the temporal and spatial predictions are two separate parts. The temporal prediction maximizes the spatial detail and accuracy of remote sensing images but fails to capture the overlay changes during the spectral unmixing. On the other hand, the spatial prediction captures the overlay changes but ignores the spatial details of the image. Therefore, the spatial prediction results and temporal prediction results can be combined together by a weighting algorithm to improve the robustness and accuracy of the fusion algorithm. In this article, the SVR algorithm is used to solve the optimal temporal and spatial incremental weights, and the objective function of the weight increments can be expressed as

$$(\hat{\omega}_s, \hat{\omega}_t) = \operatorname{argmin} \frac{1}{2}\|W\|^2 + Q \sum_{n=1}^{k} \ell_\epsilon \left(\Delta C_h - \Delta\hat{C}_h\right) \quad (11)$$

in which

$$\Delta C_h = \omega_s \times \Delta C_h^S + \omega_t \times \Delta C_h^T \quad (12)$$

$$\omega_t = 1 - \omega_s \quad (13)$$

where W represents the weights of the features, Q is the regularization constant, and ℓ_ϵ is the sitar insensitive loss function. ω_s and ω_t represent the weights of the spatial and temporal increments, respectively. Theoretically, the sum of these two weights is 1. ΔC_h^S, ΔC_h^T, and $\Delta\hat{C}_h$ are the spatially relevant increment, the temporally relevant increment and the true increment of the h_{th} coarse pixel, respectively. The final incremental weighted sum of spatial and temporal increments is shown as

$$\Delta F^{ST} \left(x_{ij}, y_{ij}, b\right) = \omega_s \times \Delta S + \omega_t \times \Delta T. \quad (14)$$

C. Distribution of Residuals

Although the incremental combination ΔF^{ST} after the weight distribution combination can capture the fine change increment, there are still some errors. This calculation process is similar to FSDAF, we introduce a residual between T_B and the predicted value T_P, assuming that the residual R is related to the heterogeneity between the images, and calculates the spatial distribution of the pixel residual of the fine image between the predicted value and the real pixel value. The specific calculation process can be referred to (14)–(19) in FSDAF [33] and assign the residual distribution to each fine pixel in the image to obtain $r_i(x_{ij}, y_{ij}, b)$.

Strong temporal variations and errors between sensors can also lead to differences in data during the fusion process. TPS interpolation is used in FSDAF to guide the residual distribution by applying the equilibrium index HI from the classification map of fine images at T_B, but when the land type at T_P changes, there will be a large error in the estimation of residuals. In this article, we propose new residuals by combining the differences that exist between sensors in the fusion process to address the issue of strong temporal variations and errors between sensors. In the calculation of time prediction results, the reflectivity of the *m*-terminal element in the *b*-band at the T_B moment, denoted as $E_F^B(m, b)$, can be calculated from (15). However, the calculated result is not the pixel value at the real T_B moment and it is instead denoted by $E_F^{B'}(m, b)$. A linear model is constructed between these two values to calculate the error value formula of each pixel as

$$E_F^B(m, b) = d \times E_F^{B\prime}(m, b) + \varepsilon \quad (15)$$

$$\Delta\sigma = d \times E_F^{B\prime}(m, b) + \varepsilon - E_F^{B\prime}(m, b). \quad (16)$$

Since σ only calculates the difference between the two, it does not consider the impact of its difference on the spatiotemporal fusion results. Therefore, on this basis, this article proposes the reliability distribution residual $r_q(x_i, y_i, b)$ under the Gaussian distribution to normalize it as

$$r_q \left(x_i, y_i, b\right) = 1 - \frac{\Delta\sigma \times \text{mean}\, \Delta\sigma}{2 \times \text{stddv}\, \Delta\sigma \times \text{mean}\, \Delta C} \quad (17)$$

where $mean$ represents the standard deviation of the data and $stddv$ represents the standard deviation, to ensure that the residual is within a reasonable range. Assigning it to the fine pixels results in $r_e(x_{ij}, y_{ij}, b)$. Ultimately, the overall residuals assigned to each fine pixel are expressed by $R(x_{ij}, y_{ij}, b)$ as

$$R \left(x_{ij}, y_{ij}, b\right) = r_i \left(x_{ij}, y_{ij}, b\right) + r_e \left(x_{ij}, y_{ij}, b\right). \quad (18)$$

D. Spatial Filtering Based on Sliding Window

Due to spectral discontinuity at the boundary of low spatial resolution pixels, there is a blocking effect when the image is mixed, resulting in a loss of spatial details. Therefore, spatial filtering is used to mitigate this problem and achieve fine image prediction at T_P. In spatial filtering, pixels with similar spectral and land cover information are considered to be similar pixels. In the sliding window, up to v adjacent pixels with similar spectra are selected for each target fine pixel based on spectral distance.

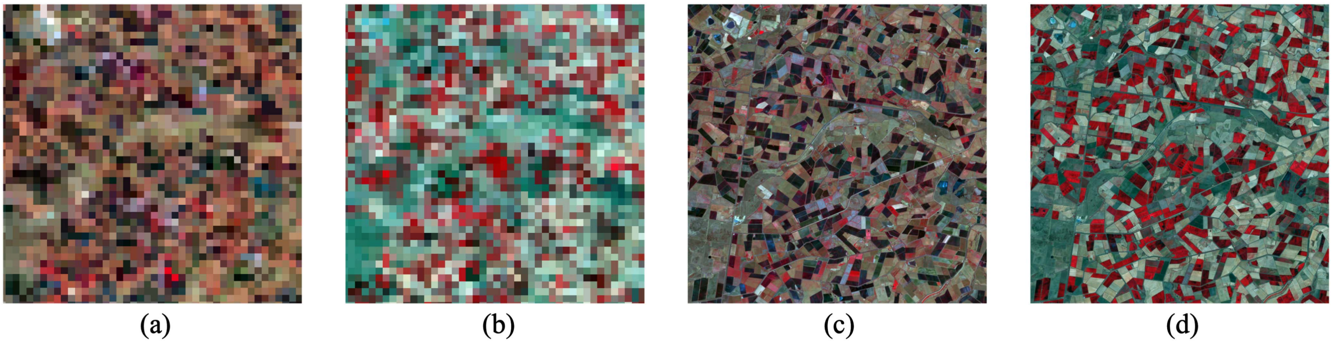

Fig. 3. Test data for Landsat and real MODIS imagery for a heterogeneous landscape. (a) MODIS image acquired on November 25, 2001. (b) MODIS image acquired on January 12, 2002. (c) Landsat image acquired on November 25, 2001. (d) Landsat image acquired on January 12, 2002.

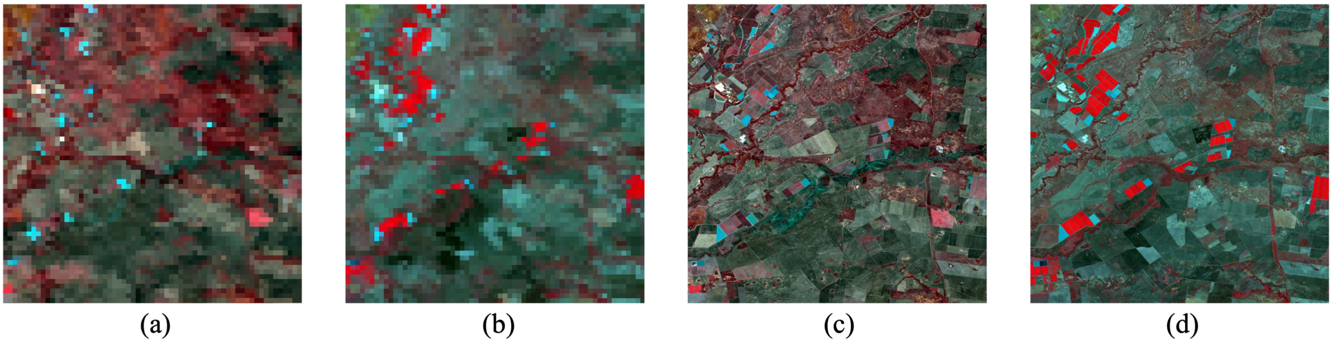

Fig. 4. Test data for Landsat and real MODIS imagery for a homogeneous landscape. (a) MODIS image acquired on November 26, 2004. (b) MODIS image acquired on February 14, 2005. (c) Landsat image acquired on November 26, 2004. (d) Landsat image acquired on February 14, 2005.

The weight of each similar pixel is represented by the Euclidean distance from it to the target fine pixel as follows:

$$D_v = 1 + \sqrt{(x_v - x_{ij})^2 + (y_v - y_{ij})^2}/L \tag{19}$$

where L represents half of the sliding window size (for instance, $L = 10$ represents the window size of 21×21 fine pixels). Therefore, the weight of each fine pixel benefits from the calculation method in ESTARFM [18], which can be calculated as follows:

$$\omega_v = (1/D_v) / \sum_{v=1}^{v} (1/D_v) . \tag{20}$$

In this article, the size of the sliding window is optimized through a series of trial-and-error experiments. The experimental results have led to the decision to set the sliding window size to 41 × 41, in order to balance the prediction accuracy and computational efficiency in the fusion process. The details of the comparison experiment can be found in Section IV. The threshold for similar pixels is set to 10–30. If the number of similar pixels exceeded 30, only the top 30 pixels are used. The resulting high spatial resolution image prediction at the final T_P can be expressed as follows:

$$\Delta F = \sum_{v=1}^{n} \omega_v \left[\Delta F^{ST}(x_{ij}, y_{ij}, b) + R(x_{ij}, y_{ij}, b)\right] \tag{21}$$

$$F'(x_{ij}, y_{ij}, b) = F_B(x_{ij}, y_{ij}, b) + \Delta F. \tag{22}$$

III. Dataset and Design of Experiments

A. Study Areas and Datasets

In the experiments, two publicly available datasets proposed by Emelyanova et al. [48] are used to validate the effectiveness and stability of the RISDAF. The dataset consists of two sites, namely the heterogeneous Coleambally Irrigation Area (CIA) and the homogeneous Lower Gwydir Catchment (LGC), which have large-scale land cover changes.

The first study area CIA is located in the southern region of New South Wales, Australia, which is located at $34.0034E$, $145.0675S$ and covers an area of 2193 km^2. The CIA dataset mainly covers areas of agricultural rice fields and woodlands with neat boundaries and large extent, and after the summer season, the plants grow luxuriantly, and the landmarks have more obvious physical and spatial changes. In this experiment, Landsat images from Landsat-7 ETM+ are used as fine images while MODIS images from Terra MODTRAN4 are used as coarse images. The CIA dataset uses the image pairs on November 25, 2001, and the MODIS image on January 12, 2002, as shown in Fig. 3 to predict the Landsat image on January 12, 2002. The actual Landsat image on January 12, 2002, as shown in Fig. 3(d) is used for verification.

The second study area, LGC data are located in the northern region of New South Wales, Australia, which is located at $149.2815E$, $299.0855S$, and covers an area of 5440 km^2. In this experiment, Landsat images from Landsat-5 TM are used as fine images, while MODIS images from Terra MOD09GA are used as coarse images. Due to the flood disaster in December 2004 in the study area, the dynamic changes in the time domain are obvious, and the types of ground objects change significantly, which is representative of the research. The LGC dataset uses the image pairs on November 26, 2004, and the MODIS image on February 14, 2005, as shown in Fig. 4 to predict the Landsat image on February 14, 2005. The actual Landsat image on February 14, 2005, as shown in Fig. 4(d) is used for verification.

The MODIS and Landsat images in the dataset are acquired on the same date and undergo preprocessing, including atmospheric and geographic correction. To match the bands in the dataset, six groups of bands similar to Landsat and MODIS are selected, and a scale factor of 20 is applied between the two images. The image size of the CIA dataset is 800 × 800 pixels, and the image size of the LGC dataset is 1200 × 1200 pixels. To meet the experimental requirements and match the Landsat resolution, the MODIS image is upsampled to 25-m resolution through nearest-neighbor interpolation. In this article, all images use the combination of NIR-red-green bands for RGB visualization, thereby facilitating more straightforward identification of land object types such as vegetation and water.

B. Experimental Settings

1) Implementation Details: The experiments are conducted on A620-G30 servers, each of which is configured with two AMD EPYC 7281 16-Core processors and 256 GB memory. In order to ensure fair comparisons, we use the same settings for all methods that we replicate. Experimental parameters are set to default values, as specified by the authors of each corresponding paper.

2) Experimental Design: In this article, we design a three-part experiment to verify the proposed RISDAF method. First, RISDAF is compared with several widely used algorithms, including the traditional weight-based algorithm STARFM, the flexible mixing-solution-based spatiotemporal fusion algorithm FSDAF, and all of the above algorithms using fine and rough pair of images as input. Although there are many scholars making improvements based on the FSDAF algorithm, the overall algorithm structure is similar, and FSDAF has representativeness and stability, and it has been widely used in various fields. Therefore, FSDAF is chosen as a benchmark algorithm for experimental comparisons. To ensure the fairness and authenticity of the comparative experiments, all algorithms use the default parameters provided by their respective authors during the experimentation process. First, quantitative metrics are employed to compare and analyze the prediction results of the three algorithms. Second, the experimental results are compared at the visual level by visualizing the partially enlarged results within their respective subregions. Scatter plots of the predicted and measured data in NIR bands are also plotted to aid in the analysis. Finally, the usability of the proposed algorithm is separately analyzed through ablation experiments.

3) Accuracy Assessment: To evaluate the effectiveness of the proposed method, we conduct a computational comparison of the experimental results with the corresponding real images. Five metrics are used to measure accuracy in the experiments: the root-mean-square error (RMSE), the correlation coefficient (CC), the structure similarity (SSIM), the spectral angle mapper (SAM), and the enhanced reconstruction of grayscale and aerial signal (ERGAS). These metrics are commonly used for evaluating the spatiotemporal fusion of remote sensing images. Specifically, RMSE and CC measure the differences between predicted values while SAM indicates the degree of spectral distortion, and SSIM measures the degree of texture similarity between spectra. A smaller value of RMSE, SAM, and ERGAS typically corresponds to a larger value for CC and SSIM, which indicates better fusion results.

IV. Result and Analysis

A. Results and Analysis of Heterogeneous Regions

The quantitative measures of the CIA dataset are presented in Table I, and the best results are indicated in bold font. Overall, compared with STARFM and FSDAF, the proposed RISDAF algorithm has the lowest RMSE, ERGAS, and SAM, and the highest CC and SSIM. Among the results calculated in six bands, most of them are the best, except for the ERGAS of the blue and green bands, indicating the overall best performance of the predicted results. As the two sets of data in the experiment are collected during the vegetation growth period, as the cell structure in the vegetation leaves will strongly reflect near-infrared light, resulting in very bright reflections in the NIR band. Therefore, the NIR band is extensively used in vegetation growth monitoring. We calculated the percentage improvement of RISDAF over FSDAF in the six bands across four metrics: 1) RMSE, 2) CC, 3) SSIM, and 4) ERGAS. We found that the improvements in the NIR band outperformed the remaining five bands, with percentage increases sequentially reaching 12.2%, 12.1%, 11.5%, and 11.1%. As vegetation rapidly grows, the NIR band exhibits the greatest uncertainty in RISDAF, indicating that RISDAF is more accurate in capturing heterogeneous land and ecological changes.

In this experiment, the heterogeneous dataset CIA does not undergo any significant category changes, but there are obvious physical changes in two time periods. Therefore, the experimental results focus on observing the ecosystem dynamics of the predicted images and processing the image edges. Fig. 5 shows the original Landsat image and the prediction results of the three algorithms. The experimental outcomes from STARFM exhibit substantial boundary blurring, and distortion is evident in some images. Considering the CIA dataset, which lacks prominent changes in land cover types, the predicted images generated by RISDAF demonstrate higher precision across the full spectral range compared to FSDAF. Compared to FSDAF, the predicted images produced by RISDAF demonstrate higher precision across the entire spectral range. RISDAF more accurately models and predicts spectral diversity under conditions of spatiotemporal heterogeneity.

Fig. 6 shows the zoomed-in orange and yellow areas of Fig. 5 to compare the differences between the predicted results and

TABLE I
RESULTS OF CONTRAST EXPERIMENT ON CIA DATASET

Band	STARFM					FSDAF					Proposed RISDAF				
	RMSE	CC	SSIM	ERGAS	SAM	RMSE	CC	SSIM	ERGAS	SAM	RMSE	CC	SSIM	ERGAS	SAM
Blue	0.01805	0.79093	0.87462	0.09847	-	0.01523	0.87284	0.89382	**0.07012**	**-**	**0.01502**	**0.87916**	**0.90300**	0.07756	-
Green	0.02706	0.80088	0.83159	0.16928	-	0.02404	0.84966	0.85723	**0.12599**	**-**	**0.02371**	**0.86160**	**0.87528**	0.12744	-
Red	0.04366	0.84682	0.73647	0.26468	-	0.04005	0.86814	0.75521	0.21034	-	**0.03945**	**0.89633**	**0.79477**	**0.19033**	**-**
NIR	0.06540	0.47712	0.56489	0.30968	-	0.06915	0.47438	0.56494	0.25921	-	**0.06072**	**0.53169**	**0.62975**	**0.23041**	**-**
SWIR1	0.05207	0.86041	0.70884	0.34530	-	0.04597	0.89863	0.77685	0.28461	-	**0.04509**	**0.89974**	**0.78120**	**0.25955**	**-**
SWIR2	0.04179	0.86141	0.75872	0.39961	-	0.03777	0.88106	0.77730	0.33379	-	**0.03582**	**0.88225**	**0.79962**	**0.30173**	**-**
All bands	0.04416	0.77293	0.74586	0.77422	0.13082	0.04214	0.80745	0.77089	0.70760	0.11574	**0.03961**	**0.82513**	**0.79727**	**0.67275**	**0.10885**

The best results are indicated by bold font.

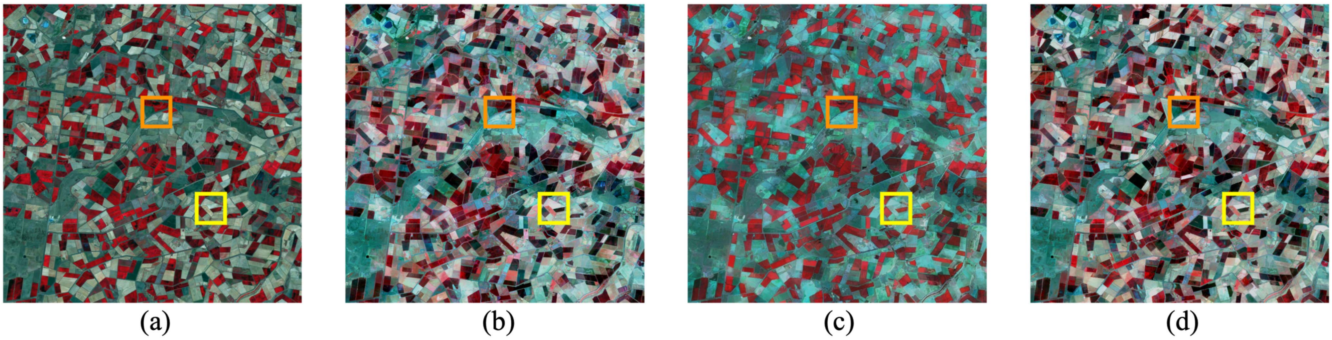

Fig. 5. Original Landsat image of January 12, 2002, and the prediction results of the three algorithms. (a) Original Landsat image. (b) Predicted images by STARFM. (c) Predicted images by FSDAF. (d) Predicted images by proposed RISDAF.

the actual images. Both regions exhibit significant phenological changes during the vegetation growth period, and the predictions generated by the three algorithms are generally consistent with the real Landsat images. Based on visual interpretation, all three methods accurately capture the phenological changes between November 25, 2001, and January 12, 2002. However, the structure of the proposed RISDAF algorithm is seemed to be closer to the original images than those of STARFM and FSDAF. The enlarged orange area is shown in Fig. 6(a)–(d). The RISDAF algorithm can capture the small irregular objects and pixel value changes more accurately and has more significant advantages in monitoring small farmland. Only the proposed method can correctly predict the physical changes and boundary areas in the orange area in Fig. 5(a), STARFM and FSDAF boundaries are blurred. This is because the proposed RISDAF combines temporal and spatial increments in its computation process, leading to advancements in restoring spatial details. Additionally, the magnified yellow regions in Fig. 5(a)–(d) correspond to what is shown in Fig. 6(e)–(h). Although the results after spectral rendering imaging predicted by the three algorithms are not exactly close to the original image, the image predicted by the proposed RISDAF algorithm has the clearest image structure as well as acquires sharper image boundaries, which is superior to STARFM and FSDAF.

In vegetation remote sensing, the reflection in the near-infrared region is highly influenced by the internal structure of the leaves. Therefore, we select the predicted and actual data of the near-infrared band to create a scatter plot. In the CIA dataset, the scatter plots of the NIR bands based on STARFM, FSDAF, and the proposed RISDAF are shown in Fig. 7. As can be seen from Fig. 7, there is no significant bias among the three algorithms. However, by calculating R^2, the results from the proposed RISDAF are superior to STARFM and FSDAF, being closer to the 1:1 line, indicating better fitting performance and smaller errors between the actual values and predicted values.

B. Results and Analysis of Homogeneous Regions

The experimental quantitative evaluation indicators of the homogeneous LGC dataset are shown in Table II. Compared with STARFM and FSDAF, the results of RISDAF prediction have the lowest RMSE, ERGAS, and SAM, the highest CC and SSIM, and the best effect. This shows that the proposed RISDAF algorithm has more powerful spectral retrieval and image reconstruction capabilities when the land scale changes on a large scale. The NIR band shows the most tremendous uncertainty in RISDAF and FSDAF, and the prediction effect is the best. In addition to the NIR band, each band also shows a good performance. The accuracy improvement in SWIR1 band and SWIR2 band is obvious, second only to the NIR band, and the RMSE single band index is increased by 5.5%. Therefore, the proposed RISDAF algorithm can predict large-scale land changes better compared with other algorithms.

Fig. 8 shows the Landsat image on February 14, 2005, and the prediction results of the three algorithms. It can be seen from the map that due to the impact of floods in two time periods, after the flood, the recovery of the surface caused the change of land cover to a certain extent, and some features did not recover as before. Fig. 9 shows the enlarged orange and yellow areas of Fig. 8 to compare the differences between the predicted results

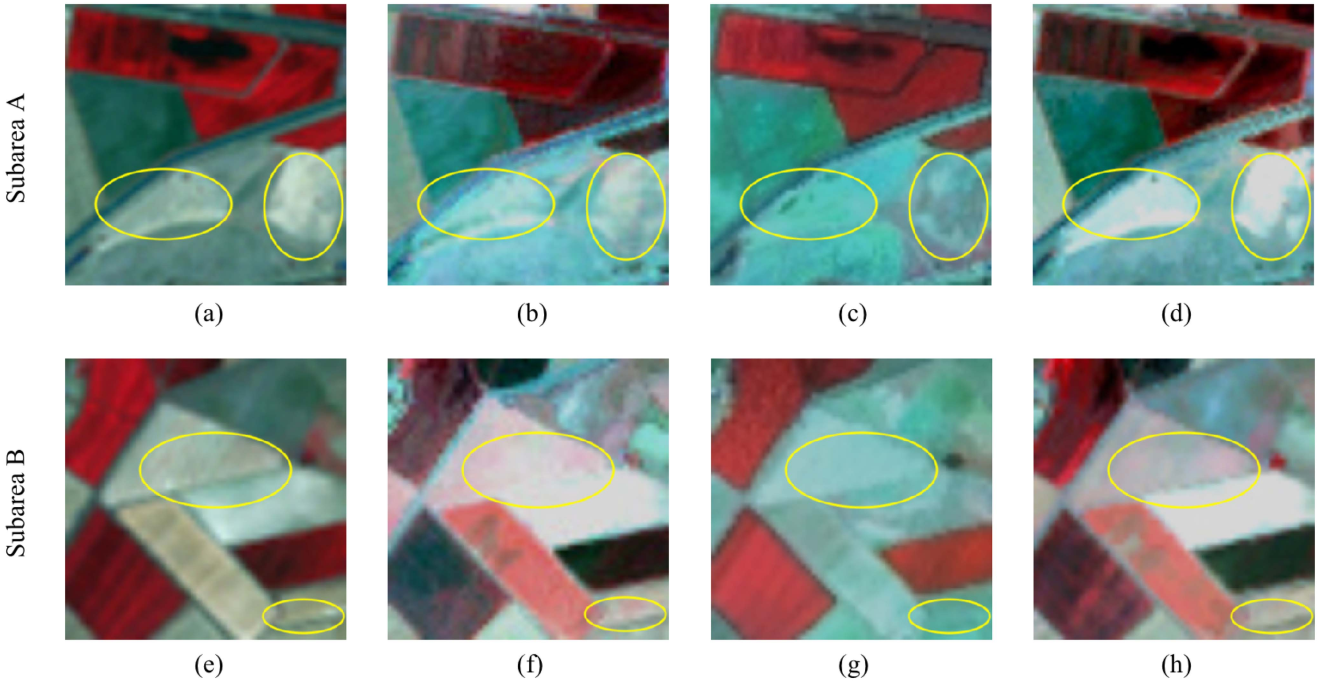

Fig. 6. Zoomed images of subareas in orange and yellow shown in Fig. 5. The corresponding results for the subarea in orange in Fig. 5(a)–(d) are (a)–(d), the corresponding results for the subarea in yellow in Fig. 5(a)–(d) are (e)–(h).

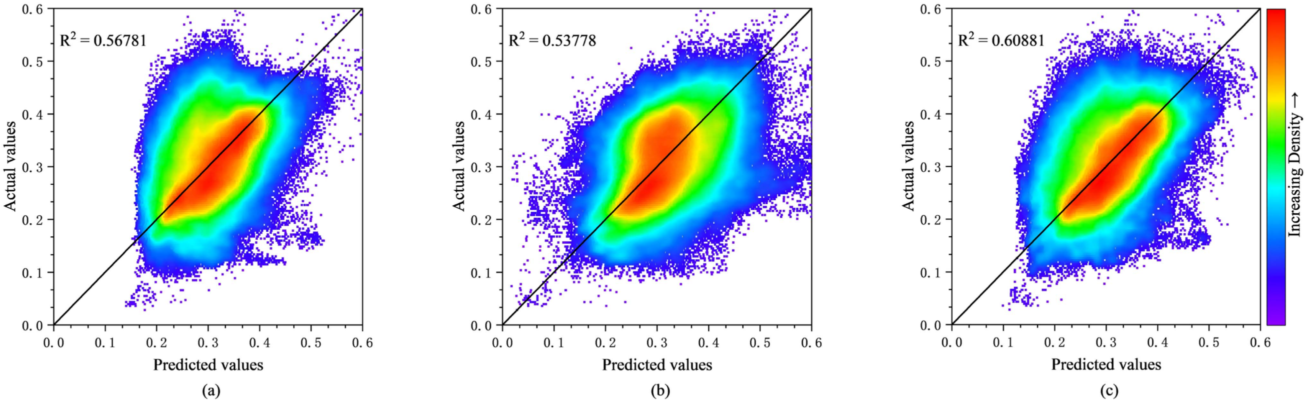

Fig. 7. Scatter plots of the actual and predicted values in the experiment feeding simulated coarse data for the NIR band in the CIA dataset (Closer to red indicates a higher density of points, the line is 1:1 line). (a) STARFM. (b) FSDAF. (c) Proposed RISDAF.

TABLE II
RESULTS OF CONTRAST EXPERIMENT ON LGC DATASET

Band	STARFM					FSDAF					Proposed RISDAF				
	RMSE	CC	SSIM	ERGAS	SAM	RMSE	CC	SSIM	ERGAS	SAM	RMSE	CC	SSIM	ERGAS	SAM
Blue	0.01215	0.80260	0.94226	**0.02843**	-	0.01221	0.80543	0.94750	0.02871	-	**0.01214**	**0.83425**	**0.95864**	0.02931	-
Green	0.01668	0.82420	0.92559	0.05676	-	**0.01393**	0.84943	0.93820	**0.04848**	-	0.01395	**0.86375**	**0.94910**	0.04913	-
Red	0.02178	0.81947	0.88546	0.09324	-	0.01789	0.84854	0.90554	0.07310	-	**0.01698**	**0.86083**	**0.92851**	**0.07130**	-
NIR	0.04142	0.87161	0.78180	0.12764	-	0.03906	0.87340	0.80542	0.10368	-	**0.03085**	**0.92351**	**0.87109**	**0.08952**	-
SWIR1	0.03782	0.87120	0.83924	0.14840	-	**0.03293**	0.82810	0.82010	0.11941	-	0.03792	**0.87468**	**0.84806**	**0.11039**	-
SWIR2	0.03143	0.88359	0.84353	0.17546	-	0.03312	0.84368	0.82477	0.14946	-	**0.03131**	**0.88609**	**0.84512**	**0.13811**	-
All bands	0.02897	0.84545	0.86965	0.51302	0.07677	0.02699	0.84143	0.87359	0.48149	0.08724	**0.02580**	**0.87387**	**0.90009**	**0.45516**	**0.07168**

The best results are indicated by bold font.

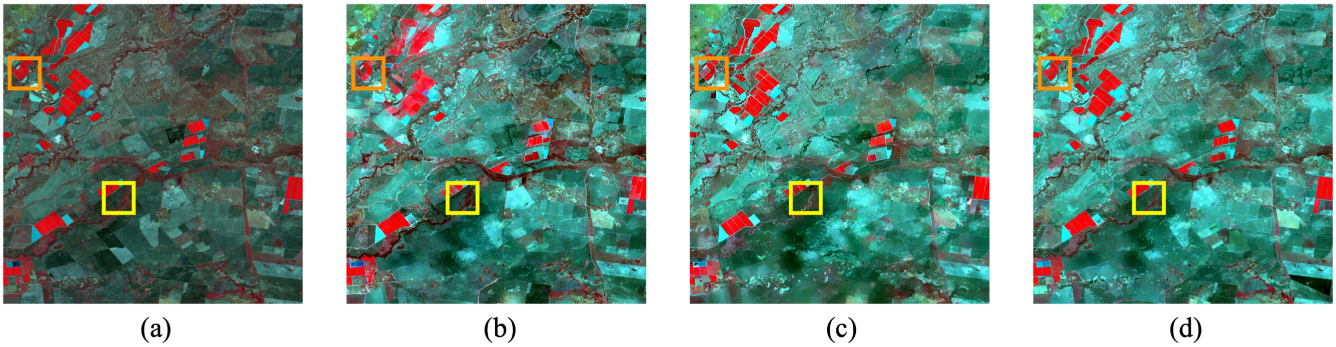

Fig. 8. Original Landsat image of February 14, 2005, and the prediction results of the three algorithms. (a) Original Landsat image. (b) Predicted images by STARFM. (c) Predicted images by FSDAF. (d) Predicted images by RISDAF.

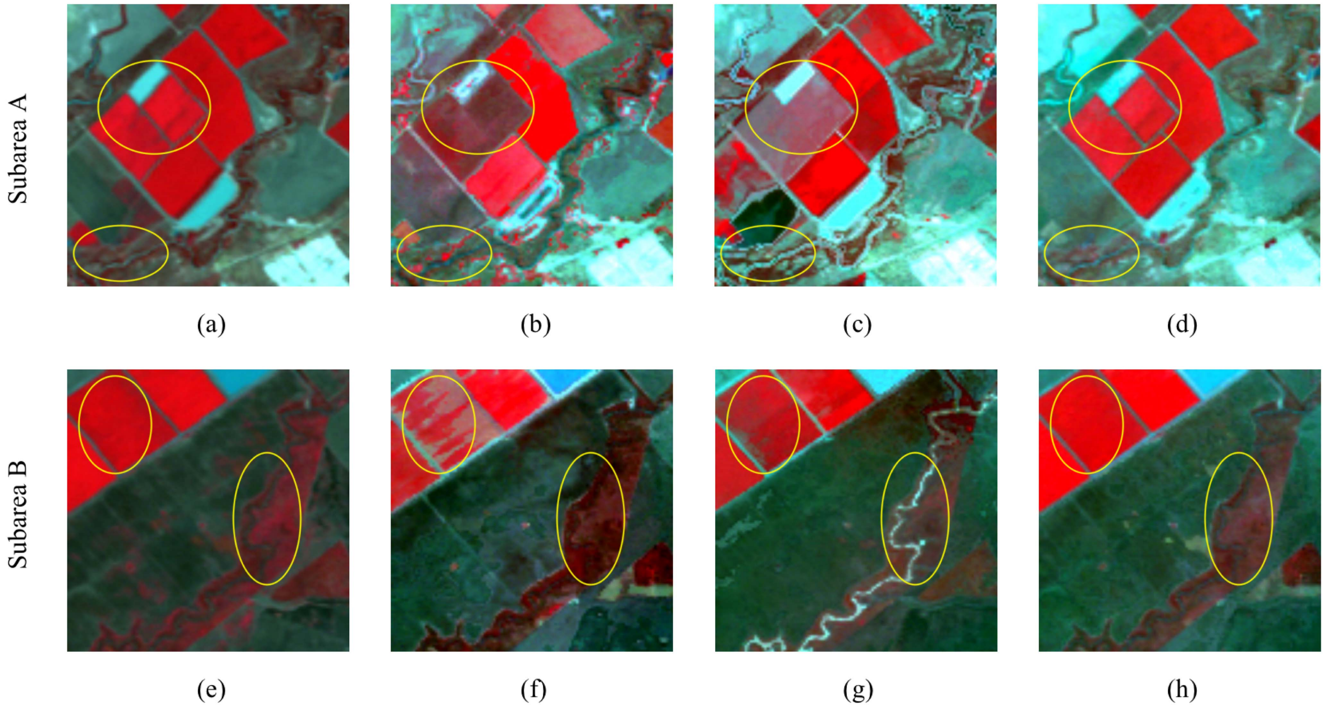

Fig. 9. Zoomed images of subareas in orange and yellow shown in Fig. 8. The corresponding results for the subarea in orange in Fig. 8(a)– (d) are (a)–(d), the corresponding results for the subarea in yellow in Fig. 8(a)–(d) are (e)–(h).

and the actual images. As can be seen in Fig. 9(a)–(d), Although STARFM and FSDAF also accurately capture physical changes due to seasonal changes, the predictions in the border areas are not accurate, small fields in some areas appear to be mixed, and there are traces of flooding that have not recovered in the FSDAF predicted images. By visual comparison, the STARFM predictions are largely accurate in Fig. 9(e)–(h), but the results of STARFM and FSDAF simulations generate images with less spatial detail, compared to RISDAF, which retains more image details with sufficient spectral similarity.

In the LGC dataset, the scatter plots of the NIR bands based on STARFM, FSDAF, and RISDAF are shown in Fig. 10. It can be seen from Fig. 10 that the proposed RISDAF is obviously closer to the 1:1 line. Furthermore, R^2 reached 0.91628, which was notably higher than that of STARFM and FSDAF, indicating that the detailed information can be better retained and the prediction accuracy can be improved in the case of surface-type mutation.

C. *Algorithm Ablation Experiment*

Fig. 11 shows the intermediate results of the proposed RISDAF algorithm on the CIA dataset, including time prediction, joint spatial prediction, residual correction, and spatial filtering prediction based on mixed pixel decomposition. The RMSE values are 0.04314, 0.04179, 0.04092, and 0.03961, respectively, which indicate that the prediction accuracy of the model is gradually increasing. Based on the fact that global phenological

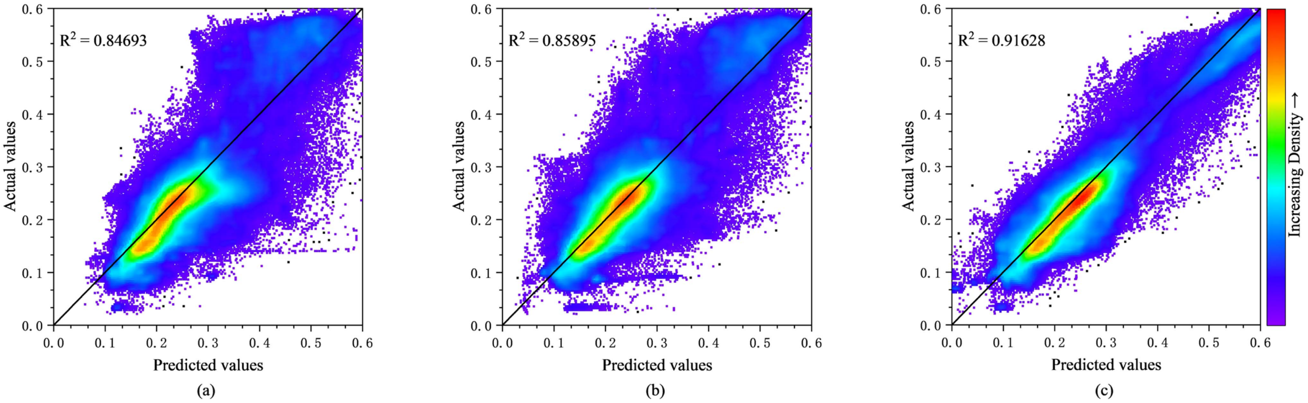

Fig. 10. Scatter plots of the actual and predicted values in the experiment feeding simulated coarse data for the NIR band in the LCG dataset (Closer to red indicates a higher density of points, the line is 1:1 line). (a) STARFM. (b) FSDAF. (c) Proposed RISDAF.

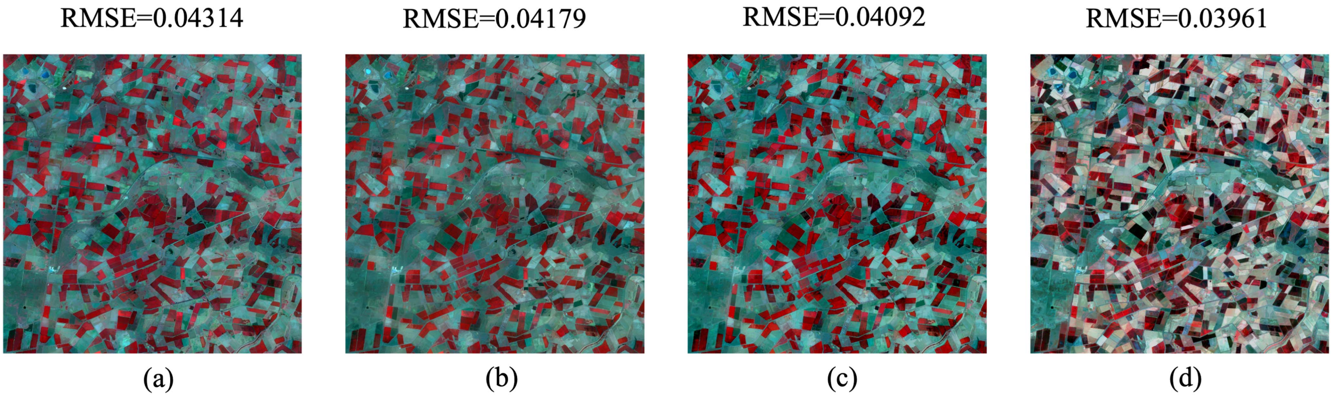

Fig. 11. RMSE of the prediction results of the four different stages of the proposed RISDAF algorithm for heterogeneous sites. (a) Time prediction. (b) Spatial prediction. (c) Residual compensation. (d) Spatial filtering.

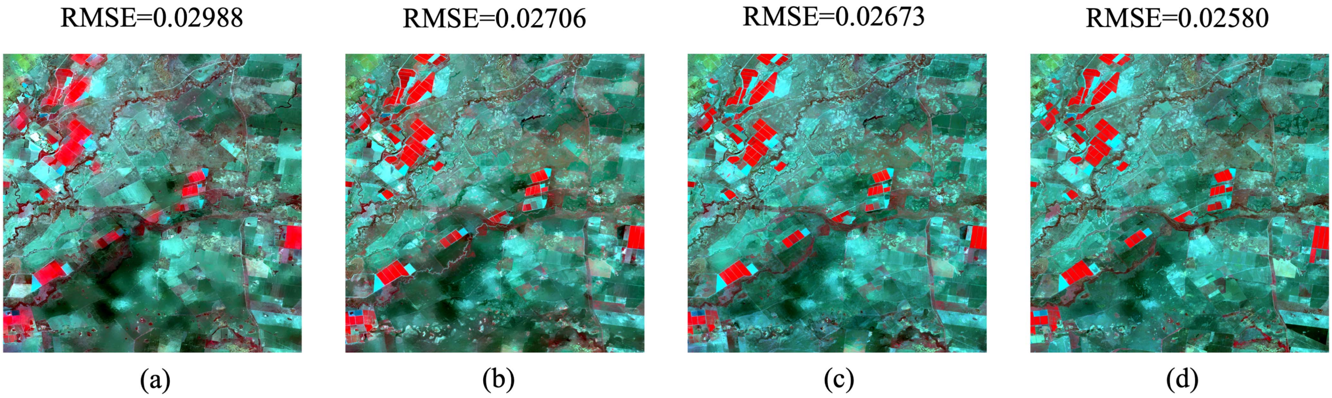

Fig. 12. RMSE of the prediction results of the four different stages of the proposed RISDAF algorithm for homogeneous sites. (a) Time prediction. (b) Spatial prediction. (c) Residual compensation. (d) Spatial filtering.

changes can be captured after decomposing mixed pixels, spatial prediction better preserves the spatial structure and details of the original image, making them clearer. Residual distribution can weaken the impact of time and space prediction as well as sensor differences. Final spatial filtering can eliminate the impact of the block effect. The classification of ground objects is clearer with more distinct boundaries and stronger spatial consistency, which results in more clear images.

Fig. 12 shows the time prediction, joint spatial prediction, residual correction, and spatial filtering prediction results of the proposed RISDAF algorithm on the LGC dataset based on mixed pixel decomposition. The RMSE values are 0.02988, 0.02706, 0.02673, and 0.02580, respectively. The predicted fine image results on February 14, 2005, gradually converge to accuracy. During the period of image prediction, changes occurred due to flooding, leading to alterations in the types of ground cover. As a result, the focus of the ablation study results was on how to restore the flood-stricken areas, track the phenological changes of vegetation, and predict changes in types of ground cover. While the temporal prediction based on spectral unmixing can roughly predict the cover status and physical changes after the flood. However, since the assumption in the temporal prediction is that the type of coverage has not changed, the areas affected by the flood are somewhat blurred, resulting in low precision of the experimental results and poor model robustness and stability. Following spatial prediction, the accuracy of the experimental results significantly improved, the RMSE value increases by 9.4% compared to the first step, which is higher than the CIA dataset, due to the occurrence of flooding. However, RISDAF is able to retain better mutation details and predict more accurately for feature types covered by floods. To account for sensor error and spatiotemporal prediction error, residual correction is employed to recover the previously unrecovered portion. Spatial filtering is also applied to retain more spatial details, which result in improving prediction accuracy and phenological change trend.

V. Discussion

As remote sensing technology continues to develop, numerous spatiotemporal fusion model algorithms have been proposed to integrate remote sensing images with different temporal and spatial resolutions. This integration enables the monitoring of long-time series through remote sensing images. However, in practical applications, fusion accuracy is often compromised by strong surface cover type shape changes, such as floods and disturbances. Through the extensive experiments conducted in this article, the proposed RISDAF algorithm achieves better prediction in both late recovery and strong physical change prediction of heterogeneous data. Our ablation experiments have proven the algorithm's indispensability in every part. In this article, we analyze and discuss the parts that are not analyzed in detail in the previous studies and present some necessary intermediate experimental results during the experiments.

A. Difference Between Proposed RISDAF and FSDAF in Time Increment Prediction

In the process of mixed pixel decomposition, clustering algorithms play a crucial role in endmember extraction and assigning each pixel to appropriate categories based on the similarity of pixels within the same category. However, calculating temporal increments is based on the similarity of pixels within the same category, making clustering algorithms very important. Our article proposes a new approach, the Gaussian mixture model clustering algorithm based on particle swarm optimization (PSO-GMM), to replace the ISODATA clustering algorithm in FSDAF. The ISODATA algorithm requires manual parameter tuning and is vulnerable to noise interference in large-scale calculations. In contrast, the PSO-GMM algorithm is capable of better exploring space [49], allowing the GMM clustering algorithm to quickly converge to optimal solutions and adaptively adjust parameters for different types and scales of datasets. Moreover, running the PSO-GMM algorithm multiple times reduces the impact of random initialization on clustering results and enhances the robustness of the clustering process. By accurately extracting different endmembers, this method significantly improves the accuracy of spectral unmixing.

Traditional temporal increment calculations in spatiotemporal fusion typically involve directly converting coarse pixel changes into fine pixel changes within a certain timeframe. This approach, however, overlooks the characteristic differences between various sensors and their response disparities to specific cover types, which could lead to errors in the fusion results. Furthermore, the alignment error between remote sensing images can impact the accuracy of spatiotemporal fusion outcomes, as it reflects the discrepancy between observed and underlying variables. To address these issues, our research introduces a linear model that registers the coarse and fine pixels and includes them in the computation of the temporal increment. This allows the model to consider the characteristic differences among sensors during the calculation process, thereby enhancing the precision of the fusion results. Experimental results show that, compared to FSDAF, the proposed RISDAF more effectively retains intraclass spatial details, particularly in predicting images after sudden changes in cover type. RISDAF demonstrates superior adaptability and accuracy, implying that the proposed RISDAF has greater adaptability and predictive capabilities in handling complex spatiotemporal fusion challenges. This offers a novel, more precise solution for spatiotemporal fusion in remote sensing imagery.

B. Advantages of Combining Time and Space Increments

Most hybrid spatiotemporal data fusion algorithms such as FSDAF can capture land cover changes through spatial increments [50]. The IFSDAF builds on this by introducing constrained least squares (CLS) to combine the temporal predictions after spectral unmixing with the number of spatial changes after TPS interpolation to obtain the best predicted amount. In contrast, the RISDAF proposed in this article applies bicubic interpolation to calculate spatial increments, which has the following advantages over TPS interpolation in predicting spatial increments. 1) Due to the variety of surface types of datasets, the bicubic interpolation considers more surrounding data points, leading to the better fitting of local variations and generating smoother results. 2) On large-scale datasets measured in pixels, TPS interpolation requires solving a large-scale linear equation system, which can be relatively slow, especially on large datasets. In contrast, bicubic interpolation can precompute the coefficient matrix and perform interpolation using simple matrix multiplication, resulting in faster computation. The CLSs method is replaced by an SVR algorithm to solve the weights and combine the temporal and spatial increments. The SVR is based on the idea of a support vector machine (SVM), which has better generalization ability on multidimensional data. It maps low-dimensional data to high-dimensional data through kernel functions and flexibly controls the fitting accuracy and robustness of outliers by setting the parameters of the loss function. The proposed RISDAF improves the performance and robustness of the model by calculating the weighted sum of temporal and spatial increments through weight allocation. This approach enables the provision of more accurate and robust prediction results.

C. Improvement of Residual Allocation

In the spatiotemporal fusion model, the residual is defined as the difference between the predicted image and the true image, which is used to guide the generation of the predicted image [51]. Residual correction, as an important step to improve model accuracy, has been widely applied in spatiotemporal fusion algorithms. FSDAF introduces the homogeneity index (HI) for residual allocation to capture land cover changes. However, FSDAF suffers from serious collinearity problems where changes in independent variables can cause variance changes in residuals, thereby affecting the accuracy and reliability of spatiotemporal fusion models.

The proposed RISDAF algorithm introduces sensor errors into the residual calculation to address this issue. As different types of errors exist between different sensors such as systematic errors and random errors, it is necessary to consider their effects when allocating residuals. Linear correction is used to adjust differences between different sensors, and residual coefficients (r_e) are proposed to convert them into the same reflectance values. Specifically, a fitting method is used for linear correction to establish reflectance conversion between different sensors and ensure that residuals are evenly distributed, thereby improving the accuracy and reliability of model prediction results.

D. Algorithm Performance Analysis Influenced by Moving Window Size

Consideration of the remote sensing imaging edge effect is crucial in the spatiotemporal fusion of remote sensing images. Due to the intricate classification of image features, the features of the image typically change during prediction. In computing the pixel value of the target pixel, neighboring pixels with similar spectra within the sliding window can be selected for computation. Generally, selecting a large sliding window increases the computational workload within the window and decreases the correlation between the center target pixel and the edge pixel. On the other hand, selecting a small sliding window may not yield distinct feature calculation results for the central target pixel. Therefore, choosing an appropriate sliding window size to select similar pixels can significantly improve image prediction accuracy.

In this experiment, the heterogeneous dataset CIA and homogeneous dataset LGC are used for the calculation under different window sizes, which are 11 × 11, 21 × 21, 31 × 31, 41 × 41, 51 × 51, and 71 × 71. The experimental results of the CIA dataset with different sliding window sizes on January 12, 2002, are shown in Table III. The experimental accuracy does not improve because the sliding window size increased, and the five evaluation metrics of the proposed RISDAF algorithm are optimal when the sliding window size is 41 × 41 OLI pixels. The experimental results of the LGC dataset with different sliding window sizes are shown in Table IV. The experimental accuracy is optimal when the sliding window size is 41 × 41 OLI pixels, and compared with the window size of 11 × 11 OLI pixels, RMSE improves by 3.78%, CC improves by 3.81%, SSIM improves by 3.39%, ERGAS improves by 5.19%, and SAM improves by 8.02%. Therefore, the sliding window of 41× 41 OLI pixels is the best parameter for the experiment, which achieves the smoothing effect while retaining the spatial details.

TABLE III
RESULTS OF MOVING WINDOW SIZE EXPERIMENT ON CIA DATASET

Window size	RMSE	CC	SSIM	ERGAS	SAM
11	0.04237	0.79323	0.77208	0.71069	0.11329
21	0.04169	0.80410	0.78703	0.69954	0.11373
31	0.04022	0.81411	0.79191	0.68536	0.10935
41	**0.03961**	**0.82513**	**0.79727**	**0.67275**	**0.10885**
51	0.03985	0.82039	0.79477	0.67635	0.10895
71	0.04127	0.81563	0.79120	0.68506	0.11216

The best results are indicated by bold font.

TABLE IV
RESULTS OF MOVING WINDOW SIZE EXPERIMENT ON LGC DATASET

Window size	RMSE	CC	SSIM	ERGAS	SAM
11	0.02682	0.84179	0.87058	0.48006	0.07793
21	0.02666	0.85687	0.87634	0.47902	0.07685
31	0.02658	0.85961	0.88269	0.46338	0.07540
41	**0.02580**	**0.87387**	**0.90009**	**0.45516**	**0.07168**
51	0.02597	0.86104	0.88804	0.45821	0.07293
71	0.02633	0.86375	0.89573	0.47412	0.07429

The best results are indicated by bold font.

E. Further Improvement of RISDAF

The aforementioned experimental results and analysis demonstrate that the proposed spatiotemporal fusion algorithm RISDAF provides an improved solution for heterogeneous data with strong phenological changes and regions with surface-type variations. This enhancement improves the accuracy of fusion, yet it is undeniable that the algorithm has certain limitations. Most of the current spatiotemporal fusion algorithms, including the one presented in this article, are predominantly based on public datasets for experiments and analysis, thereby heavily relying on the quality of the input data. In the process of spectral unmixing, RISDAF depends on the accuracy of land classification. If applied in real-world scenarios, the algorithm's effectiveness might be reduced due to the possibility of multiple types coexisting within a single pixel in a heterogeneous landscape. Furthermore, when handling large real-world datasets or generating and analyzing long-term sequence data, the algorithm demands substantial computational resources. This constrains the feasibility of the model in scenarios where resources or processing time are limited. Therefore, improving model precision and fusion efficiency for real data types will be the focal point of future research.

VI. Conclusion

This article proposes a RISDAF method to address the difficulties and problems of fusing remote sensing images with different resolutions. The RISDAF uses Landsat and MODIS data as fine and coarse images, which are compared with two excellent spatiotemporal fusion algorithms in terms of quantitative metrics and visual interpretation experiments. The RISDAF method provides a better solution for heterogeneous data with strong phenological changes and areas with changes in surface types, improving the accuracy and adaptability of mixed image element decomposition, scalability, and smoothness. Furthermore, through ablation experiments, it has been verified that each part of the proposed model in this article has an irreplaceable role.

In summary, RISDAF provides a more reliable solution to improve the accuracy of mixed pixel decomposition, optimize spatial details by calculating the weighted sum of temporal and spatial increments, and reduce the impact of sensor differences on spatiotemporal fusion, which improves the stability and robustness of the algorithm. This improvement is beneficial for effective dynamic land surface monitoring through satellite imagery.

References

[1] D.-X. Song et al., "Very rapid forest cover change in Sichuan province, China: 40 years of change using images from declassified spy satellites and Landsat," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 10964–10976, Oct. 2021.

[2] S. Francini, G. D'Amico, E. Vangi, C. Borghi, and G. Chirici, "Integrating GEDI and Landsat: Spaceborne Lidar and four decades of optical imagery for the analysis of forest disturbances and biomass changes in Italy," *Sensors*, vol. 22, no. 5, 2022, Art. no. 2015.

[3] B. Li and K.-N. Liu, "Forest biomass estimation based on UAV optical remote sensing," *Forest Eng.*, vol. 38, no. 5, pp. 83–92, 2022.

[4] W. Zou, W. Jing, G. Chen, Y. Lu, and H. Song, "A survey of Big Data analytics for smart forestry," *IEEE Access*, vol. 7, pp. 46621–46636, 2019.

[5] M. S. Dhillon, T. Dahms, C. Kuebert-Flock, E. Borg, C. Conrad, and T. Ullmann, "Modelling crop biomass from synthetic remote sensing time series: Example for the DEMMIN test site Germany," *Remote Sens.*, vol. 12, no. 11, 2020, Art. no. 1819.

[6] N. Liu, D. Wang, and Q. Guo, "Exploring the influence of large trees on temperate forest spatial structure from the angle of mingling," *Forest Ecol. Manage.*, vol. 492, 2021, Art. no. 119220.

[7] M. S. Dhillon, T. C. Dahms, C. Kübert-Flock, I. Steffan-Dewenter, J. Zhang, and T. Ullmann, "Spatiotemporal fusion modelling using STARFM: Examples of Landsat 8 and Sentinel-2 NDVI in Bavaria," *Remote Sens.*, vol. 14, no. 3, 2022, Art. no. 677.

[8] D. Hong, N. Yokoya, N. Ge, J. Chanussot, and X. X. Zhu, "Learnable manifold alignment (LeMA): A semi-supervised cross-modality learning framework for land cover and land use classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 147, pp. 193–205, 2019.

[9] G. Yang et al., "MSFusion: Multistage for remote sensing image spatiotemporal fusion based on texture transformer and convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4653–4666, Jun. 2022.

[10] Y. Li, J. Li, and Z. Shaoquan, "A extremely fast spatio-temporal fusion method for remotely sensed images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 4452–4455.

[11] L. Dong et al., "Very high resolution remote sensing imagery classification using a fusion of random forest and deep learning technique–subtropical area for example," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 113–128, Dec. 2020.

[12] Y. Qu, H. Qi, and C. Kwan, "Unsupervised sparse Dirichlet-Net for hyperspectral image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2511–2520.

[13] D. Jia, P. Gao, C. Cheng, and S. Ye, "Multiple-feature-driven co-training method for crop mapping based on remote sensing time series imagery," *Int. J. Remote Sens.*, vol. 41, no. 20, pp. 8096–8120, 2020.

[14] Y. Cui et al., "A new fusion algorithm for simultaneously improving spatio-temporal continuity and quality of remotely sensed soil moisture over the Tibetan plateau," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 83–91, Dec. 2021.

[15] Y. Tang, Q. Wang, and P. M. Atkinson, "Filling then spatio-temporal fusion for all-sky MODIS land surface temperature generation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1350–1364, 2023.

[16] 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.

[17] 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.

[18] 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.

[19] Y. Tang, Q. Wang, K. Zhang, and P. M. Atkinson, "Quantifying the effect of registration error on spatio-temporal fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 487–503, Jan. 2020.

[20] H. Shen, P. Wu, Y. Liu, T. Ai, Y. Wang, and X. Liu, "A spatial and temporal reflectance fusion model considering sensor observation differences," *Int. J. Remote Sens.*, vol. 34, no. 12, pp. 4367–4383, 2013.

[21] Y. Chen, Y. Yang, X. Pan, X. Meng, and J. Hu, "Spatiotemporal fusion network for land surface temperature based on a conditional variational autoencoder," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, Jun. 2022, Art. no. 5002813.

[22] Z. Zhu, Y. Tao, and X. Luo, "HCNNet: A hybrid convolutional neural network for spatiotemporal image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, May 2022, Art. no. 2005716.

[23] J. Amorós-López et al., "Multitemporal fusion of Landsat/TM and ENVISAT/MERIS for crop monitoring," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 23, pp. 132–141, 2013.

[24] C. M. Gevaert and F. J. García-Haro, "A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion," *Remote Sens. Environ.*, vol. 156, pp. 34–44, 2015.

[25] X. Liu, C. Deng, J. Chanussot, D. Hong, and B. Zhao, "StfNet: A two-stream convolutional neural network for spatiotemporal image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6552–6564, Sep. 2019.

[26] J. Wei, L. Wang, P. Liu, X. Chen, W. Li, and A. Y. Zomaya, "Spatiotemporal fusion of MODIS and Landsat-7 reflectance images via compressed sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7126–7139, 2017.

[27] C. Zhao, X. Gao, W. J. Emery, Y. Wang, and J. Li, "An integrated spatio-spectral–temporal sparse representation method for fusing remote-sensing images with different resolutions," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3358–3370, Jun. 2018.

[28] Z. Tan, P. Yue, L. Di, and J. Tang, "Deriving high spatiotemporal remote sensing images using deep convolutional network," *Remote Sens.*, vol. 10, no. 7, 2018, Art. no. 1066.

[29] X. Zhang and H. Gan, "STF-Net: An improved depth network based on spatio-temporal data fusion for PM2. 5 concentration prediction," *Future Gener. Comput. Syst.*, vol. 144, pp. 37–49, 2023.

[30] Z. Tan, M. Gao, X. Li, and L. Jiang, "A flexible reference-insensitive spatiotemporal fusion model for remote sensing images using conditional generative adversarial network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, Jan. 2022, Art. no. 5601413.

[31] B. Song et al., "MLFF-GAN: A multilevel feature fusion with GAN for spatiotemporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, May 2022, Art. no. 4410816.

[32] H. Gao et al., "cuFSDAF: An enhanced flexible spatiotemporal data fusion algorithm parallelized using graphics processing units," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, May 2022, Art. no. 4403016.

[33] 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.

[34] R. Zurita-Milla, J. G. P. W. Clevers, and M. E. Schaepman, "Unmixing-based Landsat TM and MERIS FR data fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 453–457, Jul. 2008.

[35] J. Wu, Q. Cheng, H. Li, S. Li, X. Guan, and H. Shen, “Spatiotemporal fusion with only two remote sensing images as input,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 6206–6219, Oct. 2020.

[36] 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.

[37] 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.

[38] 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.

[39] H. Zhang, Y. Sun, W. Shi, D. Guo, and N. Zheng, “An object-based spatiotemporal fusion model for remote sensing images,” *Eur. J. Remote Sens.*, vol. 54, no. 1, pp. 86–101, 2021.

[40] H. K. Zhang and B. Huang, “A new look at image fusion methods from a Bayesian perspective,” *Remote Sens.*, vol. 7, no. 6, pp. 6828–6861, 2015.

[41] J. Xue, Y. Leung, and T. Fung, “A Bayesian data fusion approach to spatio-temporal fusion of remotely sensed images,” *Remote Sens.*, vol. 9, no. 12, 2017, Art. no. 1310.

[42] H. Yang et al., “Measuring the urban land surface temperature variations under Zhengzhou city expansion using Landsat-like data,” *Remote Sens.*, vol. 12, no. 5, 2020, Art. no. 801.

[43] 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. Geosciences*, vol. 124, pp. 93–102, 2019.

[44] J. Zhou et al., “Sensitivity of six typical spatiotemporal fusion methods to different influential factors: A comparative study for a normalized difference vegetation index time series reconstruction,” *Remote Sens. Environ.*, vol. 252, 2021, Art. no. 112130.

[45] J. Li, Y. Li, L. He, J. Chen, and A. Plaza, “Spatio-temporal fusion for remote sensing data: An overview and new benchmark,” *Sci. China Inf. Sci.*, vol. 63, pp. 1–17, 2020.

[46] L. Shi, “Effects of hydrological connectivity on the growth dynamics of wetland vegetation in Poyang Lake,” Ph.D. dissertation, Beijing Forestry Univ., Beijing, China, 2018.

[47] Z. Cao, S. Chen, F. Gao, and X. Li, “Improving phenological monitoring of winter wheat by considering sensor spectral response in spatiotemporal image fusion,” *Phys. Chem. Earth*, Parts A/B/C, vol. 116, 2020, Art. no. 102859.

[48] I. V. Emelyanova, T. R. McVicar, T. G. Van Niel, L. T. Li, and A. I. J. M. van Dijk, “Assessing the accuracy of blending Landsat–MODIS surface reflectances in two landscapes with contrasting spatial and temporal dynamics: A framework for algorithm selection,” *Remote Sens. Environ.*, vol. 133, pp. 193–209, 2013.

[49] N. Kumar and H. Kumar, “A fuzzy clustering technique for enhancing the convergence performance by using improved fuzzy c-means and particle swarm optimization algorithms,” *Data Knowl. Eng.*, vol. 140, 2022, Art. no. 102050.

[50] D. Jia, C. Cheng, C. Song, S. Shen, L. Ning, and T. Zhang, “A hybrid deep learning-based spatiotemporal fusion method for combining satellite images with different resolutions,” *Remote Sens.*, vol. 13, no. 4, 2021, Art. no. 645.

[51] Y. Lin, J. Qiao, J. Bi, H. Yuan, H. Gao, and M. Zhou, “Hybrid water quality prediction with graph attention and spatio-temporal fusion,” in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, 2022, pp. 1419–1424.

Weipeng Jing (Member, IEEE) received the Ph.D. degree in computer architecture from the Harbin Institute of Technology, Harbin, China, in 2016.

He is currently a Professor with Northeast Forestry University, Harbin, China. He has authored or coauthored more than 100 research articles in refereed journals and conference proceedings. His research interests include modeling and scheduling for distributed computing systems, artificial intelligence, and spatial data mining.

Dr. Jing was the Publication Chair of ICPCSEE (2016, 2017, 2018, 2019, and 2020), BDTA (2015 and 2017), Collaborate 2017, Wicon 2016, Ficloud 2016, etc. He is now a member of the ACM and a Senior Member of the China Computer Federation.

Tongtong Lou received the B.S. degree in computer science and technology, in 2016, and the M.S. degree in forestry information engineering from Northeast Forestry University, Harbin, China, in 2020, where she is currently working toward the Ph.D. degree in forestry information engineering.

Her research interests and expertise include remote sensing and Big Data parallel computing.

Zeyu Wang received the B.S. degree in information management and information system, in 2015, and the M.S. degree in agricultural informatization from Northeast Forestry University, Harbin, China, in 2019, where he is currently working toward the Ph.D. degree in forestry information engineering.

He is currently with the University of Sanya, Sanya, China. He has authored or coauthored several papers in domestic and international journals. His main research interests include machine learning, cloud computing, and artificial intelligence.

Weitao Zou received B.S. degree in information management and information system from Northeast Forestry University, Harbin, China, in 2018, where he is currently working toward the Ph.D. degree in forestry information engineering.

He is currently a visiting Ph.D. student with The University of Sydney, Camperdown, NSW, Australia. He has authored or coauthored many papers in famous journals and conferences, such as IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, ICA3PP, etc. His research interests include parallel and distributed computing, big spatial data, and remote sensing.

Zekun Xu received the M.S. degree in electronic communication engineering from Northeastern Forestry University, Harbin, China, in 2020, where he is currently working toward the Ph.D. degree in forestry information engineering.

He has authored or coauthored several papers in domestic and international journals. His research interests include computer vision and data mining.

Linda Mohaisen received the B.S. degree in computer science from King Abdul Aziz University, Jeddah, Saudi Arabia, in 2004, the M.S. degree in computer engineering from the University of Central Florida, Orlando, FL, USA, in 2011, and the Ph.D. degree in artificial intelligence from the University of Alabama in Huntsville, Huntsville, AL, USA, in 2018.

She is currently with the Faculty of Computing and Information Technology, King Abdul Aziz University. She is an author and coauthor of some publications and a referee of many referred international journals and conferences. Her current research interests include artificial intelligence, wireless communication, mobile network, vehicular ad hoc network, wireless sensor network, hybrid network, cross-layer design, and modeling and performance evaluation of computer networks.

Chao Li received the Ph.D. degree in information and communication systems from the Harbin Institute of Technology, Harbin, China, in 2012.

She is currently the Deputy Director of Artificial Intelligence and a master's tutor with Harbin Northeast Forestry University, Harbin. She was a visiting scholar at the Center for Forestry and Natural Resources, University of West Virginia, Morgantown, West Virginia, USA. She has chaired and completed projects such as the National Natural Science Foundation of China Youth Project. She has long conducted research on signal processing, feature extraction, and multimodal fusion. She authored or coauthored more than 20 SCI/EI retrieval papers being the first author and corresponding author, and won the second prize in Heilongjiang Provincial Science and Technology Progress Award.

Dr. Li is currently a member of the China Computer Federation and the Chinese Society of Forestry.

Jian Wang received the Ph.D. degree from Institute of Software, Chinese Academy of Sciences. He is currently a Senior Engineer with Aerospace Information Research Institute, CAS, Beijing, China. He is also the Head of the National Key Research and Development Program and has been so far responsible for more than 30 projects. His research interests include high-performance geographic computing, parallel computing, remote Big Data service and application, etc.