https://doi.org/10.1016/j.rse.2020.111973

FSDAF 2.0: Improving the performance of retrieving land cover changes and preserving spatial details

Dizhou Guo[a], Wenzhong Shi[b,*], Ming Hao[a,*], Xiaolin Zhu[b]

[a] NASG Key Laboratory of Land Environment and Disaster Monitoring, China University of Mining and Technology, Xuzhou 221116, China

[b] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China

Abstract

Spatiotemporal fusion is a feasible solution to resolve the tradeoff between the temporal and spatial resolutions of remote sensing images. However, the development of spatiotemporal fusion algorithms has not yet reached maturity, and existing methods still face many challenges, e.g., accurately retrieving land cover changes and improving the robustness of fusion algorithms. The Flexible Spatiotemporal DAta Fusion (FSDAF) method proposed by Zhu et al. in 2016 solved the abovementioned problems to some extent. However, FSDAF has two shortcomings that can be further improved: (1) FSDAF is prone to losing spatial details and predicting a "blurrier" image due to the input of coarse pixels containing type change information and a large amount of boundary information for unmixing calculation, and (2) FSDAF does not optimize the areas of land cover change. In this paper, an improved FSDAF method incorporating change detection technology and an optimized model for changed-type areas (FSDAF 2.0) was proposed to improve the aforementioned problems.

Based on the existing FSDAF algorithm, FSDAF 2.0 excludes changed pixels and boundary pixels for unmixing calculation, and establishes a model to optimize the changed pixels. Its performance was compared with that of the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM), the original FSDAF, and the enhanced FSDAF that incorporates sub-pixel class fraction change information (SFSDAF). Two sites consisting of landscapes with heterogeneous and large-scale abrupt land cover changes were employed for testing. The results of the experiments demonstrate that FSDAF 2.0 effectively improves the shortcomings of FSDAF, blends synthetic fine-resolution images with higher accuracy than that of the other three methods at two different sites, and strengthens the robustness of the fusion algorithm. More importantly, FSDAF 2.0 has a powerful ability to retrieve land cover changes and provides a feasible way to improve the performance of retrieving land cover changes. Consequently, FSDAF 2.0 has great potential for monitoring complex dynamic changes in the Earth's surface.

Keywords

Spatiotemporal fusion; FSDAF; Change detection; Landsat; MODIS

1. Introduction

With the rapid development of remote sensing technology in the past decade, remote sensing has played an increasingly important role in monitoring urbanization (Taubenböck et al., 2012), ecological system dynamic changes (Shen et al., 2011; Zhu et al., 2019), natural disasters (Rudorff et al., 2018; Zhang et al., 2014), crop yield estimation (Battude et al., 2016) and other applications. Acquiring satellite images with high spatial and temporal resolution means that high-precision monitoring of Earth systems by dense time-series can be achieved, which greatly improves the value of remote sensing images in applications. However, limited by relevant budget and satellite sensor technology, the spatial resolution of available satellite images can only be improved at the expense of other performance (Zhang et al., 2015), for instance, sacrifices of temporal and spectral resolution. Accordingly, existing remote sensing satellites have difficulty obtaining images with both high temporal resolution and high spatial resolution, which means that available satellite images cannot satisfy the needs for studying high-frequency changes on the Earth's surface, especially in heterogeneous landscapes and in areas of frequent change (Zhu et al., 2018). The lack of high spatiotemporal resolution images greatly limits the application scenarios of remote sensing.

Moreover, a growing problem currently exists in the use of remote sensing data: the remote sensing community has accumulated a large amount of historical data since the first remote sensing satellite launched. However, due to the influence of thick cloud contamination and other factors (e.g., SLC-off problem in Landsat 7 ETM+), limited remote sensing data can be employed directly, which further increases the difficulty of obtaining dense time series with high spatial resolution data.

To solve the abovementioned problems, launching more satellites or improving the performance of sensors in a short period of time is impractical, while the spatiotemporal fusion of multisource images from multiple satellites to obtain high spatial resolution and dense time-series data is a feasible solution. Compared with traditional pansharpening fusion, spatiotemporal fusion is a relatively new concept (Zhu et al., 2018) and can obtain high spatial and temporal resolution images by blending high-frequency but low-spatial-resolution images with high-spatial-resolution but low-frequency images. For convenience, images with low-spatial-resolution but high-frequency are referred to as “coarse-resolution images”, and the pixels in these images are referred to as “coarse pixels”. Correspondingly, the images with high-spatial-resolution but low-frequency are called “fine-resolution images”, and their pixels are referred to as “fine pixels”.

Due to a large number of satellite images being freely available to the public (e.g., Landsat, MODIS, and Sentinel) and the large demand for Earth monitoring with high spatial resolution and dense time series, in the last two decades, there has been increasing interest in spatiotemporal fusion. Recently, existing spatiotemporal fusion methods have been classified into five groups based on the specific principle: weight function-based, unmixing-based, Bayesian-based, learning-based, and hybrid methods. A comprehensive review of spatiotemporal fusion methods in these five groups can be found in the literature (Zhu et al., 2018).

The Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) is the weight function-based method developed first (Gao et al., 2006). Most other weight function-based methods are developed by the principle of STARFM and improve the corresponding defects; for example, the Spatial Temporal Adaptive Algorithm for mapping Reflectance Change (STAARCH) (Hilker et al., 2009) and the Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model (ESTARFM) (Zhu et al., 2010) were developed to improve the performance in disturbed landscapes and heterogeneous landscapes, respectively.

ATPPK-STARFM increases the performance in abrupt changes and heterogeneous landscapes (Wang et al., 2017), and the spatiotemporal fusion method by using a linear injection model and local neighborhood information retrieves land cover changes effectively (Sun et al., 2018). Among the unmixing-based methods, the Multisensor Multiresolution Technique (MMT) is the original method (Zhukov et al., 1999). Other unmixing-based methods subsequently developed, e.g., the Unmixing-Based Data Fusion (UBDF) (Zurita-milla et al., 2008), the Spatial Temporal Data Fusion Approach (STDFA) (Wu et al., 2012), the Landsat-MERIS fusion method (Amorós-lópez et al., 2013), and the Enhanced Spatial and Temporal Data Fusion Model (ESTDFM) (Zhang et al., 2013) can be considered improved methods of MMT. Bayesian-based methods consider spatiotemporal fusion to be a maximum a posterior (MAP) problem, and several Bayesian-based methods have been proposed and have acquired high accuracy; for example, the Bayesian Maximum Entropy method (BME) blends synthetic sea surface temperature data effectively (Li et al., 2013). The NDVI-BSFM method performs strength and robustness in providing NDVI datasets (Liao et al., 2016). Learning-based methods have grown considerably in recent years (Tan et al., 2018). The Sparse-representation-based Spatiotemporal reflectance Fusion Model (SPSTFM) is perhaps the first dictionary-pair learning method in the spatiotemporal fusion field (Huang and Song, 2012). Moosavi et al. (2015) proposed the Wavelet-Artificial Intelligence Fusion Approach (WAIFA) to blend land surface temperature data. Sun and Zhang (2019) proposed a two-stage spatiotemporal fusion method to blend the Landsat and MODIS reflectance data. The hybrid methods combine several technologies from the above categories of methods. One of the typical hybrid methods is Flexible Spatiotemporal DAta Fusion (FSDAF) proposed by Zhu et al. (2016). FSDAF integrates the unmixing method, weight function, and thin plate spline (TPS) interpolation method into one framework. As a result, FSDAF requires minimum input data and has satisfactory performance in most cases.

Consequently, spatiotemporal fusion technology has developed rapidly, but as a relatively new research topic in the remote sensing field, existing fusion methods based on different principles have their own strengths and weaknesses. The development of spatiotemporal fusion algorithms has not yet reached maturity, and existing solutions still face many challenges, such as the following typical problems.

(1) Difficulty in retrieving land cover changes

Retrieving land cover changes is a difficult problem for spatiotemporal fusion (Zhu et al., 2018). Unlike phenological changes, land cover type changes are usually caused by natural disasters or human activities, such as urbanization, deforestation, wildfires, floods and other land cover transitions. However, most of the existing spatiotemporal fusion algorithms are based on the assumption that the land cover type does not change during the fusion period. Accordingly, most algorithms, for example, the STARFM, ESTARFM, and MMT, fail to handle land cover changes. In particular, the overall accuracy and reliability of the fusion results are greatly affected by large-scale land cover changes. Among the existing spatiotemporal fusion methods, some learning-based methods, e.g., the SPSTFM, error-bound-regularized sparse coding (EBSPTM) (Wu et al., 2015) and WAIFA, can capture change information to some extent because of their specific principles. However, low computing efficiency and time-consuming problem in learning step is a key factor that limits the development and application of learning-based methods (Zhu et al., 2018). Moreover, the accuracy of learning-based methods decreases when spatial heterogeneity is high and spectral scale differences between coarse- and fine-resolution images are large (Zhu et al., 2016).

(2) Low robustness in different types of landscapes

Currently, existing spatiotemporal fusion algorithms have different advantages and limitations in different landscapes due to their various principles. For example, faced with a site with a heterogeneous landscape, a site with a homogeneous landscape, and a site with large-scale abrupt land cover changes, the blending results of using various spatiotemporal fusion algorithms are quite different; for example, STARFM has promising accuracy in homogeneous landscapes, but it is ineffective in the face of heterogeneous landscapes (Zhu et al., 2010). ESTARFM can produce a synthetic image more accurately in heterogeneous landscapes, but is even worse than STARFM for predicting abrupt changes in land cover types (Emelyanova et al., 2013). STAARCH has high accuracy in disturbed forest areas, but STAARCH cannot detect nonforest disturbance events and is sensitive to surface heterogeneity (Hilker et al., 2009). Fit-FC method can more effectively capture considerable phenological changes than STARFM (Wang and Atkinson, 2018), but it performs worse than FSDAF and STARFM in heterogeneous landscapes (Maolin Liu et al., 2019). Consequently, the robustness and reliability of the algorithm still need to be improved. The ability to guarantee accuracy and reliability in the prediction of images in all cases has become a challenge.

(3) High demand of input data

Many existing spatiotemporal fusion methods, including ESTARFM, STAARCH, STDFA, SPSTFM, etc., need more than one prior coarse- and fine-resolution image pair as input data. For real-time processing, some fusion methods like ESTARFM and STAARCH cannot be used because they need the image after the prediction time. For historical case, due to the influence of thick cloud contamination and other factors, it is difficult to acquire one more high-quality coarse- and fine-resolution image pair that has acceptable temporal distance or does not experience large-scale land cover changes between the image pairs, only one pair of prior images may be available in most cases (Zhu et al., 2016). In addition, finding another pair of prior images is time consuming (Song and Huang, 2013). Therefore, for the future proposed spatiotemporal fusion algorithm, guaranteeing the accuracy on the premise that only one pair of prior images is needed is a challenge.

The FSDAF proposed by Zhu et al. (2016) has solved the abovementioned problems to some extent. FSDAF is based on the spectral linear unmixing theory and thin-plate spline (TPS) interpolation method, combining the traditional unmixing-based method and weight function-based method. Compared with other blending methods, FSDAF requires minimum input data: one pair of coarse- and fine-resolution images acquired at T_1 and one coarse-resolution image at T_2. In addition, the FSDAF algorithm can capture more information of coarse-resolution image at T_2 by using TPS interpolation and obtain higher fusion accuracy in various landscapes, especially in heterogeneous landscapes. Moreover, FSDAF has the ability to predict both gradual change and land cover type change.

Accordingly, FSDAF is considered to be a potential fusion method that can efficiently handle land cover change, as long as the change is detectable in coarse-resolution images (Zhu et al., 2016).

Although FSDAF performs excellently in spatiotemporal fusion, it has two problems that can be improved: (1) in the unmixing process of FSDAF, coarse pixels with change values within the range of the 0.1–0.9 quantiles (or a narrower range, e.g., 0.2-0.8) are selected to participate in the unmixing calculation to filter out the changed pixels. However, this strategy is empirical and not strict. In addition, the coarse pixels containing a large amount of boundary information are not excluded. These pixels would introduce the wrong spectral information into the unmixing calculation once selected. As a result, FSDAF is prone to reducing the contrast between different objects, losing spatial details and predicting a "blurrier" image. This problem is particularly acute in the case of the large-scale type change occurring during the fusion period. (2) FSDAF can capture part of the change information from the coarse-resolution image at T_2 by using the TPS interpolation method (Zhu et al., 2018); however, the FSDAF algorithm does not judge whether the land cover type has changed in a fine-resolution image. Therefore, FSDAF does not have the capacity to determine the crisp boundary of land cover type change and accurately estimate the values of the changed pixels. The accuracy and reliability of the blending result of FSDAF would be affected if the fusion processes do not include a land cover change detection module. Consequently, FSDAF needs further improvement and optimization.

To address the above problems, an improved FSDAF method incorporating change detection technology and an optimized model for changed-type areas (FSDAF 2.0) is proposed in this paper. Its goal is to overcome the disadvantages of FSDAF and solve the three typical problems mentioned above. Specifically, FSDAF 2.0 employs change detection technology to find the changed-type pixels by detecting two coarse-resolution images of different phases and excludes the coarse pixels containing changed-type areas and a large amount of boundary information in the unmixing process. Furthermore, FSDAF 2.0 establishes an optimized model that performs targeted optimization on the prediction values of changed-type pixels. To validate the effectiveness of the proposed method, we compared the performance of FSDAF 2.0 with the STARFM, the original FSDAF, and the enhanced FSDAF that incorporates sub-pixel class fraction change information (SFSDAF) (Li et al., 2020) at two different sites, including a site with a heterogeneous landscape, and a site with large-scale abrupt land cover change.

2. Methodology

FSDAF 2.0 only requires one pair of coarse- and fine-resolution images acquired at T_1 and one coarse-resolution image at T_2. The flowchart of FSDAF 2.0 is shown in Fig. 1. The main idea of FSDAF 2.0 is employing a change detection algorithm to find changed-type areas and perform targeted optimization. These additional steps of FSDAF 2.0 are within yellow boxes in Fig. 1. Other steps remain the same as those of the original FSDAF.

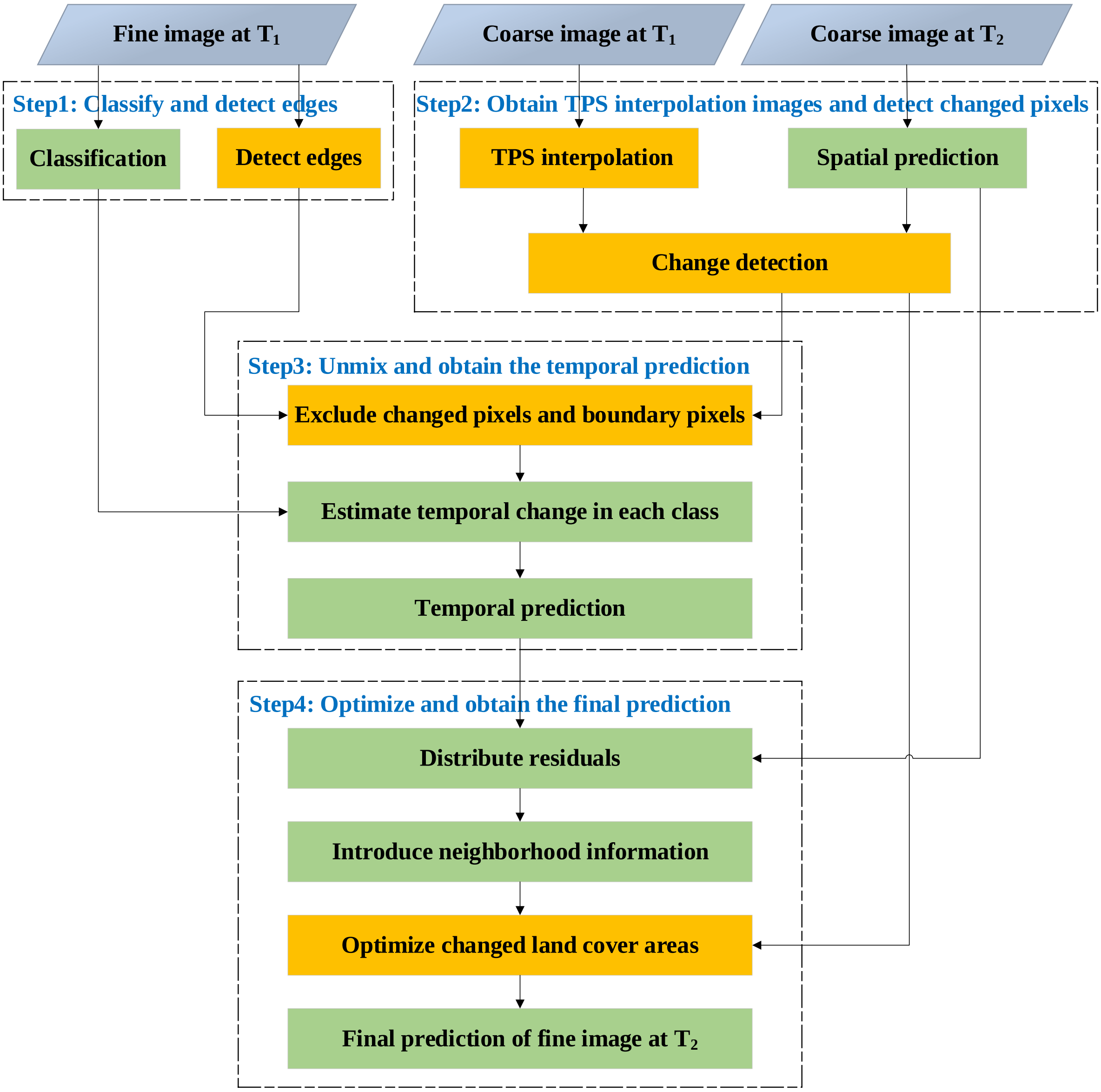


Fig. 1. Flowchart of FSDAF 2.0

FSDAF 2.0 includes four main steps: (1) classify and detect edges; (2) obtain thin plate spline (TPS) interpolation images and detect changed pixels; (3) unmix and obtain the temporal prediction; and (4) optimize and obtain the final prediction. The specific steps and theories of FSDAF 2.0 are as follows:

2.1.Classify and detect edges

This step involves acquiring the classification map and the edge image. The fraction of each class within one coarse pixel, which needs to be used in the subsequent unmixing process, can be obtained from the classification map. In this paper, the unsupervised classifier ISODATA is used to classify the fine-resolution image at T_1. The edge detection algorithm is employed to extract the features of the object boundary, i.e., obtain the edge image of the fine-resolution image at T_1. For convenience, the edge of the surface is referred to as the "boundary area". The fine pixels inside the boundary area are called "boundary pixels". The boundary pixels can be found by using the threshold method in the edge image. The boundary pixel is mixed with two or more types of features, generally due to being located at the edges of different objects. Accordingly, its spectral features are quite different from the type to which it belongs. Therefore, the prediction accuracy can be affected once coarse pixels containing a large number of boundary pixels are employed to estimate the temporal change of each class, specifically, increasing the error of the unmixing calculation and reducing the contrast between different objects in the blending image. To avoid the above problems, determining the boundary pixels is a key process. FSDAF 2.0 employs the Sobel operator to obtain the edge image. The pixels in the edge image with values within the range of the 0.96-1.0 quantiles are defined as boundary pixels.

2.2.Obtain TPS interpolation images and detect changed pixels

Thin plate spline (TPS) interpolation is a kind of spatial interpolation method based on spatial dependence, and is a tool for interpolating surfaces from scattered datasets. TPS interpolation can produce a "smooth" interpolation image and capture the spatial patterns and land cover type change signals. More information about TPS can be found in the literature (Dubrule, 1984). The interpolation image of the coarse-resolution image at T_2 is defined as "spatial prediction" based on the above features. Different from the original FSDAF that downscales the coarse-resolution image of the prediction phase only, FSDAF 2.0 uses the TPS interpolation method to downscale the coarse-resolution images of two phases, and the interpolation images are used in the following process of change detection.

The ability of capturing change information in FSDAF mainly results from the TPS interpolation of coarse-resolution image at T_2 (i.e., spatial prediction). However, FSDAF distributes residuals on the assumption that errors depend mainly on the homogeneity of the surface in step 3. In other words, the original FSDAF does not make targeted optimization to changed-type areas. The lack of this process affects the accuracy and reliability of the blending result when facing a site with land cover type changes. The key to settling this problem is to find the changed-type pixels during the fusion period. Therefore, it is reasonable to employ change detection technology to solve this problem. The selection of the change detection algorithm depends on many factors, e.g., image size, resolution, scale of the type changes, and calculation efficiency. For convenience, the fine pixels that have type changes during the fusion period are referred to as "changed pixels", while other pixels are called "unchanged pixels".

In this paper, two thresholding algorithms were employed: the thresholding method based on the Gaussian distribution model and OTSU (OTSU, 1979). The thresholding method based on the Gaussian model assumes difference values are in accordance with the Gaussian distribution mathematical model and judges the probability of the change in pixel classification by the three-sigma rule. Specifically, when the difference value is larger than the sum of double standard deviations and the average change value or smaller than the difference between the average change value and double standard deviations, it has a 95.45% probability of classification change. It is reasonable to consider that these pixels have experienced type changes. The threshold values Q of the Gaussian model can be calculated as:

$$\begin{aligned} Q_{neg} &= mean(C_d) - 2 \times stddv(C_d) \;\; if \;\; C_d < 0, \\ Q_{pos} &= mean(C_d) + 2 \times stddv(C_d) \;\; if \;\; C_d \geq 0, \end{aligned} \tag{1}$$

where C_1 and C_2 are the coarse-resolution images at T_1 and T_2, respectively; $C_d = (C_2 - C_1)$; $mean(C_d)$ is the average value of C_d; and $stddv(C_d)$ is the standard deviation of C_d. The thresholds are calculated separately for two cases.

In most cases, the difference values of remote sensing images in two phases approximately agree with the Gaussian distribution (Song et al., 2000), but it is false when large-scale change occurs on the land surface because large-scale change (e.g., floods and earthquakes) usually changes the boundaries of objects and has irregular spectral variations in the image. To address this limitation, OTSU was employed as a complementary algorithm. The OTSU algorithm is considered one of the most successful methods for image thresholding because of its simple calculation (Lai and Rosin, 2014). In the field of remote sensing change detection, OTSU is an adaptive thresholding method that is sensitive to spectral change. The threshold values of OTSU can be calculated as:

$$Q_{neg} = max(\omega_0 \times (\mu_0 - \mu)^2 + \omega_1 \times (\mu_1 - \mu)^2) \;\; if \;\; C_d < 0, \tag{2}$$

$$Q_{pos} = max(\omega_0 \times (\mu_0 - \mu)^2 + \omega_1 \times (\mu_1 - \mu)^2) \;\; if \;\; C_d \geq 0,$$

where ω_0 is the ratio of the unchanged pixels to the number of total pixels, ω_1 is the ratio of the changed pixels to the number of total pixels, μ_0 is the average value of the unchanged pixels, μ_1 is the average value of the changed pixels, and μ is the average value of the total pixels. OTSU employs the traversing method to obtain the threshold values. The thresholds are calculated separately for two cases.

Compared with the Gaussian distribution model, OTSU tends to mistakenly judge phenological changes as classification changes, but it is suitable for detecting areas where the land cover type changes on a large scale. Consequently, FSDAF 2.0 chooses the change detection algorithm according to whether the difference values of TPS interpolation result in two phases in accordance with the Gaussian distribution model. There are many methods that can judge whether the difference values agree with the Gaussian distribution model, such as the Shapiro-Wilk test (ROYSTON, 2000), Kolmogorov-Smirnov test (Lilliefors, 2012) and histogram judgment method.

In the above threshold calculation, coarse-resolution images are used instead of TPS interpolation images. We found that there is negligible difference between the threshold values obtained by using interpolation images and coarse-resolution images as input. This strategy effectively reduces the calculation amount. After thresholds are obtained, the difference image of two interpolation images is employed to make the change detection binary image. Specifically, the values of pixels in the range of the thresholds are selected as changed pixels.

The specific steps of change detection include (1) acquiring a difference image by subtracting two coarse-resolution images; (2) judging whether the difference values are accordance with the Gaussian distribution model; (3) calculating thresholds; and (4) determining the changed pixels in the difference image of interpolation images. In this paper, the thresholds of each band are obtained, which need to be used to limit the results of the unmixing calculation in step 3. In the change detection process, the shortwave infrared band (e.g., SWIR1 or SWIR2 is chosen in Landsat 7 ETM+) is employed to calculate difference values only. The shortwave infrared band is often employed to distinguish rock (Y. Yamaguchi and Naito, 2010), soil water content (Sadeghi et al., 2015) and different types of vegetation (Panigrahy and Panigrahy, 2009), and the image of this band has a high contrast. Moreover, using the shortwave infrared band can effectively detect the changes caused by geological disasters such as floods and landslides.

2.3.Unmix and obtain the temporal prediction

This step employs linear unmixing technology to estimate the change value between two phases and then calculate the temporal prediction. In the classification process, the class fractions within a coarse pixel $f_c(x_i, y_i)$ are acquired. According to the spectral linear unmixing theory, and assuming no type change occurs during the blending period, the temporal change of a coarse pixel is the weighted sum of the temporal change of all classes within it:

$$\Delta C(x_i, y_i, b) = \sum_{c=1}^{n} f_c(x_i, y_i) \times \Delta F(c, b), \tag{3}$$

where n means the number of classes and $\Delta F(c,b)$ indicates the change value of class c in band b. Assuming the temporal change in each class is the same, theoretically, we can choose m $(m > n)$ coarse pixels to construct the following matrix Eq. (3), and solve it by using the least square method.

$$\begin{bmatrix} \Delta C(x_1,y_1,b) \\ \vdots \\ \Delta C(x_i,y_i,b) \\ \vdots \\ \Delta C(x_m,y_m,b) \end{bmatrix} = \begin{bmatrix} f_1(x_1,y_1) & f_2(x_1,y_1) & \cdots & f_n(x_1,y_1) \\ \vdots & \vdots & & \vdots \\ f_1(x_i,y_i) & f_2(x_i,y_i) & \cdots & f_n(x_i,y_i) \\ \vdots & \vdots & & \vdots \\ f_1(x_m,y_m) & f_2(x_m,y_m) & \cdots & f_n(x_m,y_m) \end{bmatrix} \times \begin{bmatrix} \Delta F(1,b) \\ \vdots \\ \Delta F(c,b) \\ \vdots \\ \Delta F(n,b) \end{bmatrix}, \quad (4)$$

$$with\ s.\ t.\ \ Q_{neg} \le \Delta F \le Q_{pos},$$

Considering that the changed pixels involved in the inversion calculation can affect the accuracy, the original FSDAF excludes the pixels with ΔC outside of the range of the 0.1-0.9 quantiles. This strategy is empirical and has no theoretical basis. Instead, FSDAF 2.0 excludes the changed pixels according to the result of change detection in step 2 and limits the change value of class ΔF in the closed interval $[Q_{neg},\ Q_{pos}]$. Furthermore, FSDAF 2.0 takes into account the effect of boundary pixels, and coarse pixels containing more than 10% of boundary pixels are also excluded. The process of filtering pixels is shown in Fig. 2.

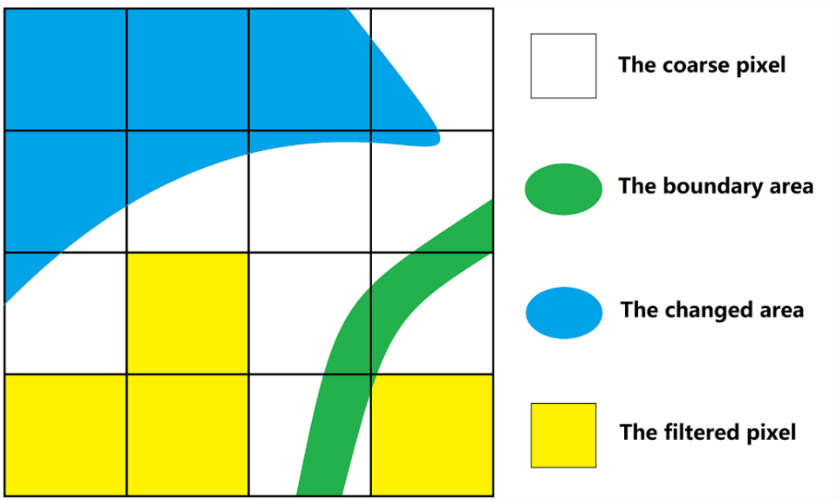


Fig. 2. The process of filtering pixels for unmixing calculation in FSDAF 2.0

After the calculation of temporal change is completed, and the change value of each class is assigned to fine pixels in the T_1 phase of the corresponding class, and the temporal prediction can be obtained.

2.4.Optimize and obtain the final prediction

Similar to FSDAF, FSDAF 2.0 distributes residuals on the assumption that errors depend mainly on the homogeneity of the surface. Consequently, the change value of each class is corrected. However, the above calculation is on a pixel-by-pixel basis; thus, the neighborhood information is employed to reduce the block effect and obtain a more robust prediction F_c. The specific processes of distributing residuals and using neighborhood information to obtain robust prediction can be found in the literature (Zhu et al., 2016).

Theoretically, the TPS interpolation result of the coarse-resolution image at T_2 preserves most of the actual information of the fine-resolution image at T_2 in homogeneous areas. In that case, using the TPS interpolation result to replace the changed pixels is a reasonable optimization method. The TPS interpolation result, however, only uses the spatial dependence of the coarse pixels, which means it produces a “smooth” result. Compared with the real fine-resolution image at T_2, the TPS interpolation result loses many spatial details, and may lead to many spectral and structural errors. In addition, the spectral differences between two sensor platforms should also be considered. Therefore, replacing the changed pixels directly by TPS interpolation results is not rigorous.

To address this problem, FSDAF 2.0 proposes the TPS reliability coefficient (TRC) to describe the reliability degree to which the changed pixels are replaced by the TPS interpolation result. The similarity, homogeneity and consistency index between two images in different phases are used to calculate the TPS reliability coefficient.

The similarity index (SI) describes the spectral similarity between the TPS interpolation image and the real fine-resolution image. Specifically, the similarity index not only describes the difference in pattern between the fine-resolution image and TPS interpolation image but also reflects the spectral difference of different sensor platforms. Theoretically, more similarity between the real image and TPS interpolation image leads to more reliable employment of the TPS interpolation result at T_2 to correct the changed pixels. Logically, the values of SI in the T_2 phase need be obtained. However, the fine-resolution image at T_2 is unknown; instead, the images at T_1 are employed to calculate the SI. Before calculating the similarity index, the difference values of the TPS interpolation image and fine-resolution image at T_1 need to be obtained, as shown in the following equation:

$$F_d(x_{ij}, y_{ij}, b) = F_1^{TPS}(x_{ij}, y_{ij}, b) - F_1(x_{ij}, y_{ij}, b), \tag{5}$$

where (x_{ij}, y_{ij}, b) is the coordinate index of the jth fine pixel within the coarse pixel at location (x_i, y_i) in band b, $F_1^{TPS}(x_{ij}, y_{ij}, b)$ is the value of the TPS interpolation result at T_1, and $F_1(x_{ij}, y_{ij}, b)$ is the value of the fine-resolution image in T_1 phase.

To simplify the calculation, we assume that the difference values $F_d(x_{ij}, y_{ij}, b)$ are in accordance with the Gaussian distribution model and consider that there is no spectral similarity between the TPS interpolation image and the real image when the difference value is beyond the triple standard deviations of the average difference. In that case, the similarity index is 0. For other changed pixels, the calculation process is as follows:

$$SI(x_{ij}, y_{ij}, b) = 1 - \left|F_d(x_{ij}, y_{ij}, b) - mean(F_d)\right| / (3 \times stddev(F_d)), \quad (6)$$

where $mean(F_d)$ is the average difference value and $stddev(F_d)$ is the standard deviation of the difference value. The SI ranges from 0 to 1, and larger values indicate more spectral similarity between the TPS interpolation image and the real image.

The homogeneity index reflects the complexity of the land surface. Logically, the higher the homogeneity of the image is, the less spatial details the surface has; thus, less information can be lost by TPS interpolation. In that case, it is more suitable to use the TPS interpolation result to modify the value of the changed pixel. FSDAF 2.0 uses a modified version of the homogeneity index in the original FSDAF to describe the homogeneity of the fine-resolution image in the T_1 phase:

$$MHI(x_{ij}, y_{ij}) = sin\left[\left(\frac{1}{k}\sum_{p=1}^{k} I_p\right) \times \pi/2\right], \quad (7)$$

where $I_p = 1$ means the pth fine pixel within a moving window with the same land cover type as the central pixel (x_{ij}, y_{ij}); otherwise, $I_p = 0$. k is the number of fine pixels within one coarse pixel; MHI ranges from 0 to 1, and larger values indicate a more homogenous landscape (Zhu et al., 2016); sin is the sine function; and π is PI. Using the above empirical formula to determine the weight parameter may not be the most accurate solution. However, this strategy can simplify the calculation and achieve satisfactory results, reasonably balancing the calculation efficiency and fusion performance.

Because the similarity index and homogeneity index at the prediction phase cannot be calculated without the fine-resolution image of the T_2 phase, instead, the similarity index and homogeneity index mentioned above are the values of the T_1 phase. These indexes of the two phases are different mainly because of the land cover changes. The reliability of the blending result cannot be guaranteed by using the calculated value of the T_1 phase to correct the values of the changed pixels. To solve this problem, the consistency index (CI) of the two coarse-resolution images is proposed to reflect the consistency of the spatial information and structural relations in different phases. A larger consistency index indicates a smaller change in the internal spatial relation between two phases; furthermore, a larger index means that the similarity index and homogeneity index of the two phases are closer. The calculation formula of CI is as follows:

$$CI(b) = 1 - |stddev(C_2) - stddev(C_1)| / \left(stddev(C_2) + stddev(C_1)\right), \quad (8)$$

where $stddev(C_1)$ is the standard deviation of the coarse-resolution image value at T_1 and $stddev(C_2)$ is the standard deviation of the values of the coarse-resolution image at T_2. The standard deviation can describe the internal relationship characteristics of the image data. Theoretically, CI can reflect the changes in the internal characteristics of the two phases.

The TPS reliability coefficient TRC is the product of the similarity index, homogeneity index and consistency index, as follows:

$$TRC(x_{ij}, y_{ij}, b) = SI(x_{ij}, y_{ij}, b) \times MHI(x_{ij}, y_{ij}, b) \times CI(b). \quad (9)$$

The list optimization model for changed pixels is as follows:

$$F_{op}(x_{ij}, y_{ij}, b) = [1 - TRC(x_{ij}, y_{ij}, b)] \times F_c(x_{ij}, y_{ij}, b) + TRC(x_{ij}, y_{ij}, b) \times \quad (10)$$

$$F_2^{TPS}(x_{ij}, y_{ij}, b),\ if\ (x_{ij}, y_{ij})\ belongs\ to\ changed\ pixel,$$

where (x_{ij}, y_{ij}, b) is the index of the changed pixel, $F_c(x_{ij}, y_{ij}, b)$ is the value of the changed pixel in the robust prediction, and $F_2^{TPS}(x_{ij}, y_{ij}, b)$ is the TPS interpolation result in the coarse-resolution image at T_2, i.e., the spatial prediction. After optimizing each changed pixel, the final synthetic image is obtained.

3. Testing experiment

3.1. Study area and data

FSDAF 2.0 was tested by two challenging landscapes that were intercepted from Irina V. Emelyanova's open spatiotemporal fusion experimental data (Emelyanova et al., 2013), including a site with a heterogeneous landscape, and a site with large-scale abrupt land cover change.

The heterogeneous site is located in the southern part of New South Wales, Australia (145.0675° E, 34.0034° S), as shown in Fig. 3. This site has many small patches of farmland with apparent spectral differences. These images (600 × 600 pixels, 15 km × 15 km, resampling resolution is 25 m) were acquired by Landsat 7 ETM+ on 04 December 2001 (T_1) and 12 January 2002 (T_2). Farmland is the main type in this study area; its terrain is flat and the soil is fertile, suggesting that it is conducive to crop growth. After the crops went through the summer growing season in the Southern Hemisphere, the land surface had rapid phenological changes. The MOD09GA images were employed as coarse-resolution images and were oversampled to 25 m resolution to match the Landsat data resolution.

The site with large-scale abrupt land cover changes is located in the northern part of New South Wales, Australia (149.2818° E, 29.0855° S), as shown in Fig. 4. Two Landsat 5 TM images (2400 × 2400 pixels, 60 km × 60 km, resampling resolution is 25 m) on 26 November 2004 (T_1) and 12 December 2004 (T_2) were employed. In mid-December 2004, a flood occurred in the farmland, which resulted in a large-scale type change in the land surface, and more than half of the land surface was affected by floods. Different from those of the first site, the MODIS images employed for spatiotemporal fusion were derived from Landsat resampling instead of MOD09GA data. Some information on land cover change was inconsistent between Landsat image and MODIS image in the T_2 phase (Shi et al., 2019). Using simulated MODIS-like images as input data can eliminate the influence of this problem on visual contrast and make the study focus on spatiotemporal fusion itself (Zhu et al., 2016). However, considering the need to further test the performance of FSDAF 2.0 to fuse real data, MOD09GA data were also used for fusion as an additional experiment, and only quantitative analysis was conducted.

There were four images from each landscape: two pairs of Landsat and MODIS images at T_1 and T_2. The Landsat image at T_2 was employed to compare the experimental results and calculate the blending accuracy. All experimental images were pre-processed for radiation calibration and atmospheric correction before experimentation.

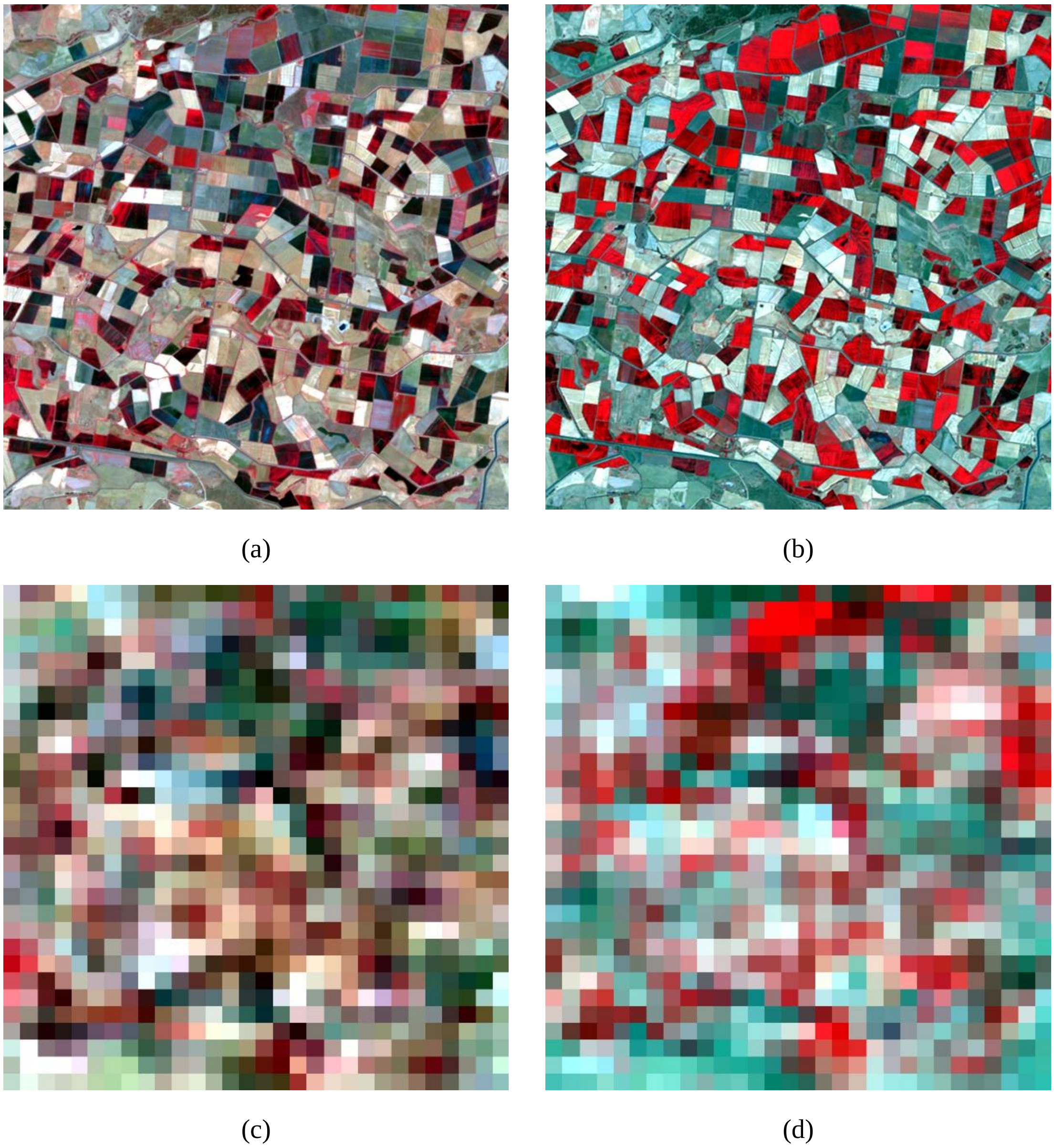


Fig. 3. Experimental data in a heterogeneous landscape: Landsat images (600×600 pixels) were acquired on (a) 04 December 2001 and (b) 12 January 2002; (c) and (d) are MOD09GA images. All images use NIR-red-green as RGB, and MOD09GA images are resampled to have the same size as the Landsat images.

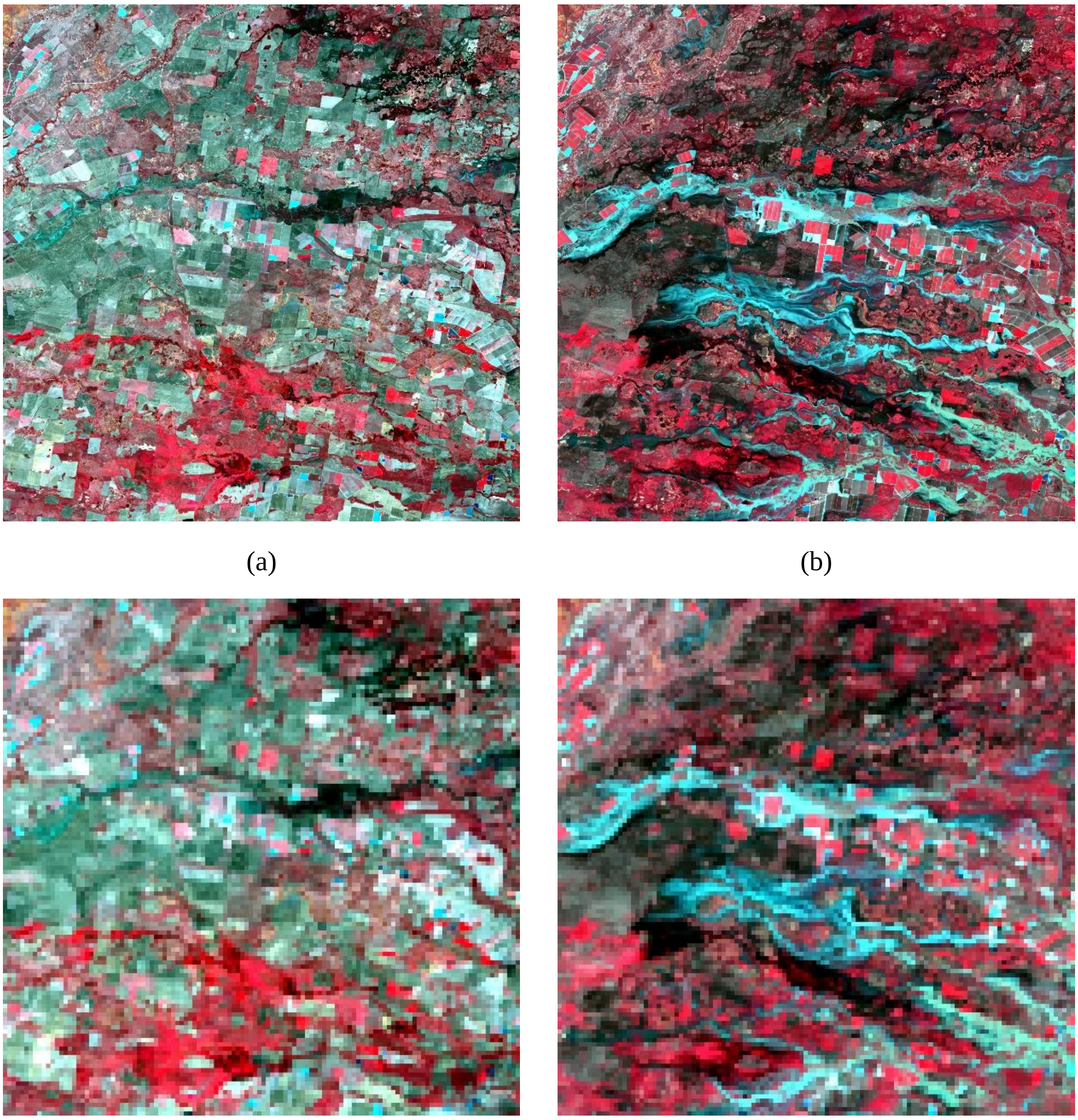


Fig. 4. Experimental data in a large-scale abrupt land cover change landscape: Landsat images (2400 × 2400 pixels) were acquired on (a) 26 November 2004 and (b) 12 December 2004, (c) and (d) are MODIS-like images aggregated from (a) and (b). All images use NIR-red-green as RGB, and MODIS-like images are resampled to have the same size as the Landsat images.

3.2.Comparison and evaluation

The performance of FSDAF 2.0 was compared with that of the STARFM, FSDAF and

SFSDAF algorithms. Each method requires the same input data: one pair of coarse- and fine-resolution images and one coarse-resolution image in the prediction phase. Blended images predicted by the four methods were compared with the real image in the T_2 phase visually and quantitatively in this section. It should be noted that SFSDAF only participated in the comparison of inputting simulated MODIS-like images in experiment 2. Because SFSDAF requires the input MODIS pixel should be the original pixel, while the process of resampling MOD09GA images to 25 m may incorrectly determine the range of the MODIS pixel. This registration error could affect the final result of SFSDAF.

Visual analysis of the fusion results was used to judge the similarity between the synthetic image and the real image in the spectrum and structure of objects by visual comparison, with the purpose of comprehensively evaluating the advantages of the improved algorithm.

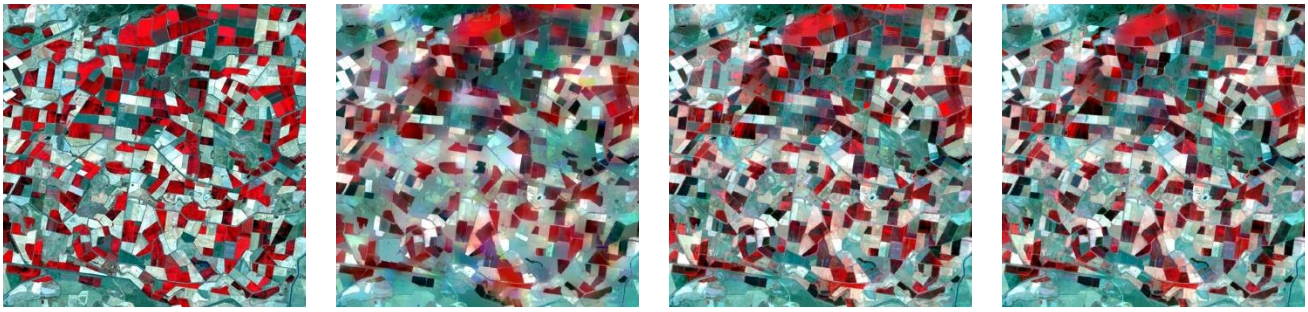
To achieve quantitative analysis, three precision indexes were proposed to reflect different aspects of accuracy. The root mean square error (RMSE) was used to gauge the difference between the predicted reflectance and the actual reflectance and describe the overall errors in the spectrum. The visual assessment index structure similarity (SSIM) was used to evaluate the similarity of the overall structure between the real and blended images. In addition, the correlation coefficient (r) was used to show the linear relationship between the predicted and actual reflectance. Theoretically, a smaller value of RMSE and larger values of SSIM and r indicate a more accurate blending result.

4. Results

4.1. Blending results and evaluation in a heterogeneous landscape

In this experiment, there was no distinct type change in the heterogeneous images but rapid phenological changes between two time periods. Therefore, the experimental comparison of blended heterogeneous images focused on the observation of the ecosystem dynamics and overall improvement of FSDAF 2.0 over the original FSDAF. In addition, this experiment also tested the robustness of FSDAF 2.0 and its performance in blending real MODIS images.

The blending results are shown in Fig. 5. It is apparent that the predicted image of STARFM has problems of boundary ambiguity and spectral anomalies, and the predicted images of FSDAF and FSDAF 2.0 are more similar to the original Landsat image. Due to there was no distinct type change in the heterogeneous images, the improvement of FSDAF 2.0 mainly comes from excluding the coarse pixels that contain a large amount of boundary information in unmixing calculation. As a result, the predicted image of FSDAF 2.0 has higher accuracy in overall spectrum than that of original FSDAF.



(a) (b) (c) (d)

Fig. 5. Original Landsat image of 12 January 2002 (a), and the predicted images by STARFM (b), FSDAF (c), and FSDAF 2.0 (d).

The quantitative evaluation data of the experimental results are shown in Table 1. The blended image predicted by FSDAF 2.0 has the smallest RMSE and highest SSIM of six bands compared with those of STARFM and FSDAF. Among all bands, the NIR band, which changes rapidly with vegetation growth cycles, has the largest difference in accuracy between FSDAF 2.0 and other two methods, suggesting that FSDAF 2.0 has higher accuracy in capturing ecosystem dynamics. Consequently, FSDAF 2.0 has higher accuracy than FSDAF in blending heterogeneous images, and shows satisfactory stability in blending real MODIS images.

Table 1

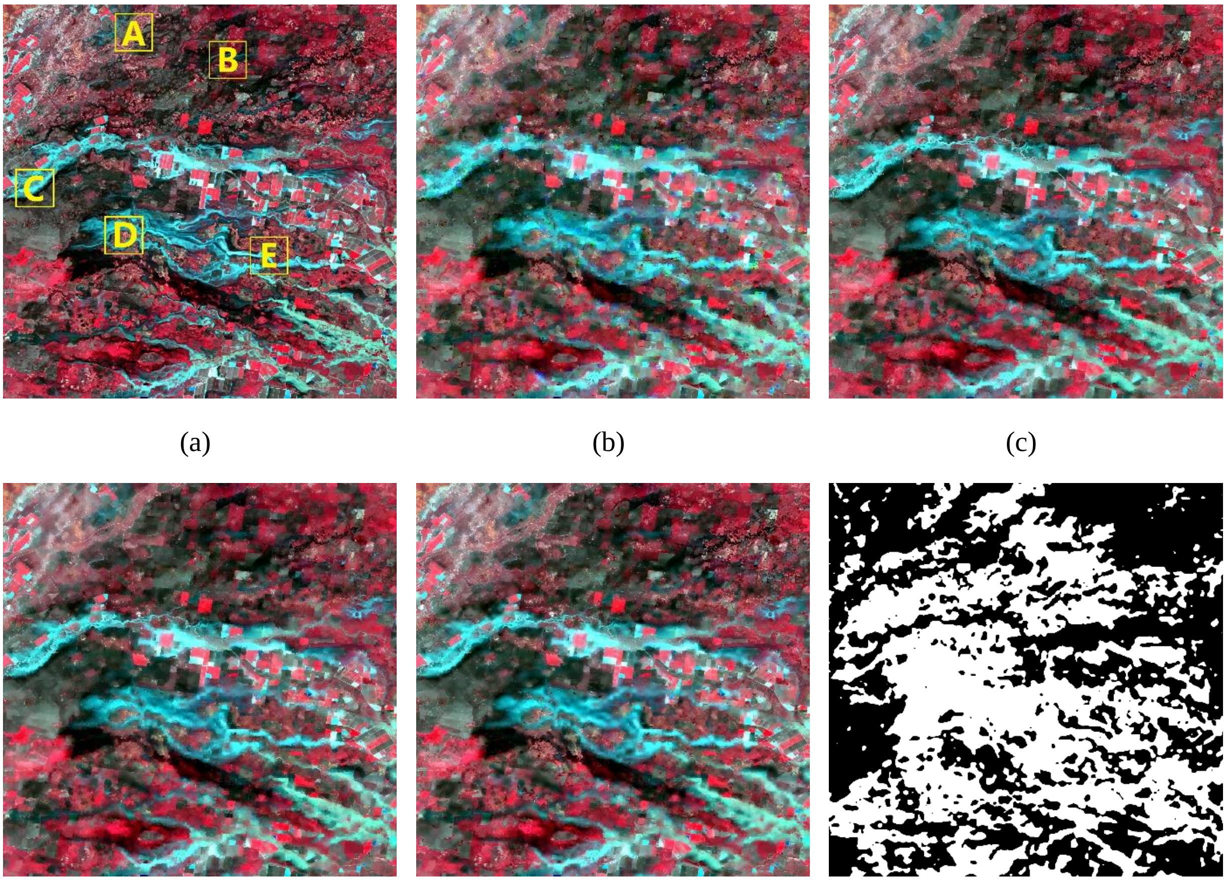
Accuracy assessment of STARFM, FSDAF and FSDAF 2.0 in a heterogeneous landscape. The units are reflectance (RMSE = root mean square error, SSIM = structural similarity, r = correlation coefficient).

	STARFM			FSDAF			FSDAF 2.0		
	RMSE	SSIM	r	RMSE	SSIM	r	RMSE	SSIM	r
Blue	0.0193	0.9077	0.8166	0.0167	0.9309	0.8611	**0.0163**	**0.9348**	**0.8690**
Green	0.0268	0.9294	0.8380	0.0234	0.9425	0.8668	**0.0229**	**0.9457**	**0.8735**
Red	0.0427	0.8678	0.8697	0.0360	0.8953	0.9009	**0.0353**	**0.9019**	**0.9054**
NIR	0.0609	0.8387	0.5218	0.0606	0.8432	0.5785	**0.0592**	**0.8487**	**0.5946**
SWIR1	0.0535	0.8776	0.8603	0.0475	0.8835	0.8894	**0.0472**	**0.8873**	**0.8911**
SWIR2	0.0423	0.8830	0.8542	0.0368	0.8984	**0.8984**	**0.0362**	**0.9055**	0.8949

4.2.Blending results and evaluation of a landscape with large-scale abrupt land cover change

Fig. 6(b), (c), (d) and (e) present blended images of STARFM, FSDAF, SFSDAF and FSDAF 2.0, respectively. Fig. 6(d) is the binary image of the change detection result in FSDAF 2.0. It can be found that the change detection algorithm is extremely sensitive in detecting areas affected by flooding.

Fig. 7 presents the enlarged areas (the areas inside the yellow bounding box in Fig. 6(a)) of the synthetic images. The predicted image of FSDAF 2.0 is the most accurate visually. Specifically, in subarea A, which is unaffected by flooding, FSDAF 2.0 preserves more spatial details than those of the other three methods. For example, only the predicted image of FSDAF 2.0 can distinguish the river indicated by the yellow arrow in Fig. 7(e). In subarea B, which was impacted by small-scale floods, it can be found that there are many spectral errors in the predicted image of FSDAF, e.g., the areas highlighted in yellow circles in Fig. 7(h). This problem is most likely caused by employing coarse pixels that contain a large number of changed pixels for the unmixing calculation. Compared with the other three methods, FSDAF 2.0 retains the most spatial details. Subarea C presents a widespread flood, and the predicted image of STARFM has obvious errors in the spectrum. This problem also occurs in the flooding areas of subarea D and subarea E. The predicted image of FSDAF has distinct boundaries between the flood area at T_2 and the flood area at T_1, see the example highlighted in the yellow circle in Fig. 7(m), while SFSDAF and FSDAF 2.0 effectively solve this problem. In subarea D, FSDAF misjudges the boundary of the flood, e.g., the area highlighted in a yellow circle in Fig. 7(r), while both SFSDAF and FSDAF 2.0 correctly judged the flood boundary. In subarea E, the predicted image of FSDAF is "blurrier", and it is difficult to distinguish the boundary of flood effects, as shown in Fig. 7(w). Both SFSDAF and FSDAF 2.0 have corrected this error. In addition, the predicted images of FSDAF and SFSDAF have many "spots" on the surface of flowing water, which means that the spectral properties are not uniform. It is apparent that the flowing water in predicted image of FSDAF 2.0 seems to be smoother and closer to the actual flowing water in the real image. Consequently, all four algorithms have the ability to retrieve land cover changes, among which STARFM is the weakest and prone to error in the areas with drastic changes. FSDAF is likely to misjudge the retrieval of type change boundaries, and it is prone to produce errors such as speckle noise in the changed-type areas. Although its ability to retain spatial details is stronger than that of STARFM, it is not satisfactory. SFSDAF has a more powerful performance in retrieving land cover changes than that of STARFM and FSDAF, but its ability to retain spatial details has no advantage in this experiment. As a result, FSDAF 2.0 is better than the other three algorithms in terms of restoring changed features and preserving spatial details.



(a) (b) (c)

(d) (e) (f)

Fig. 6. Original Landsat image of 12 December,2004 (a), its predicted images by STARFM (b), FSDAF (c), SFSDAF (d), FSDAF 2.0 (e), and the binary image of change detection result in FSDAF 2.0 (f).

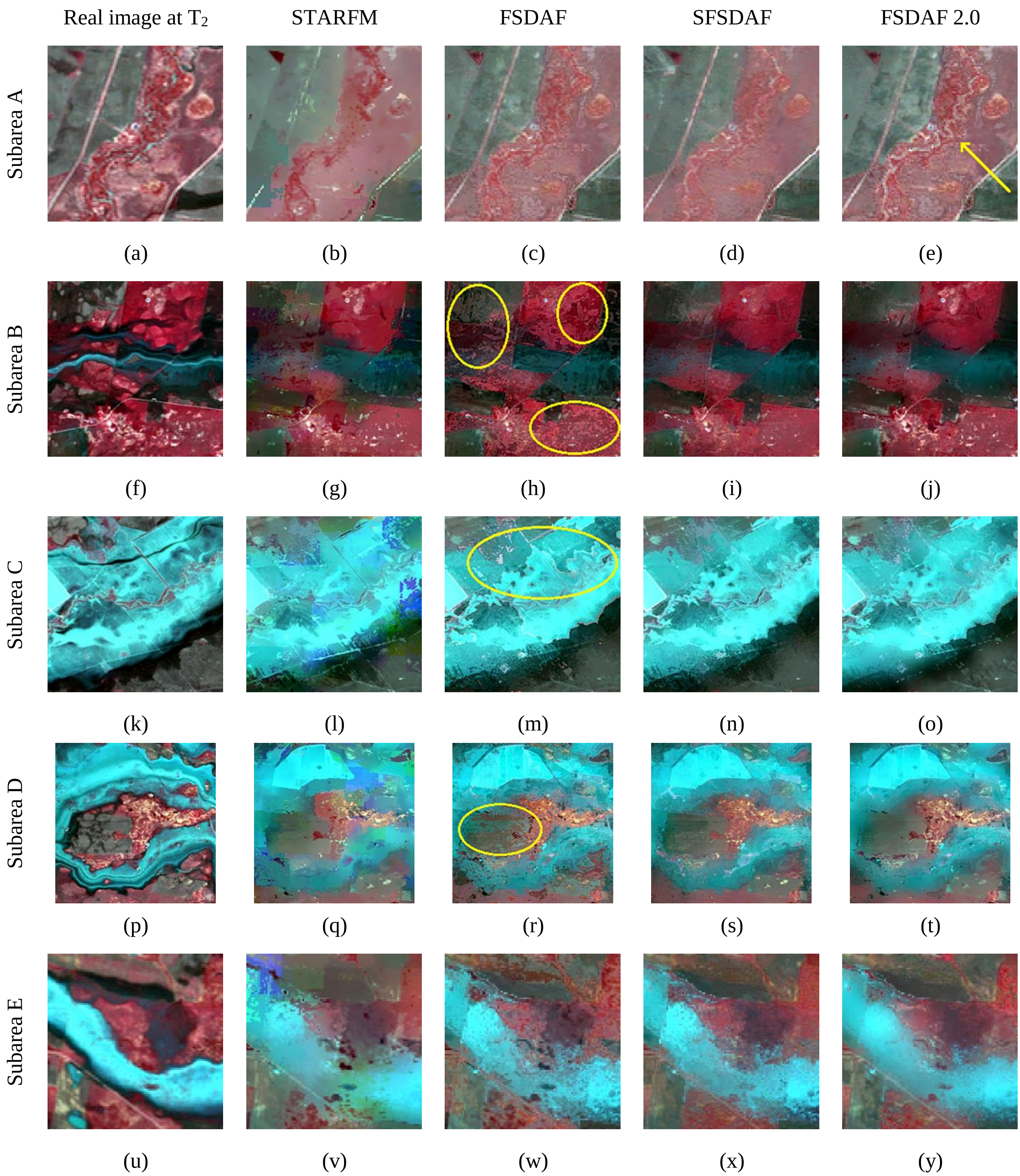


Fig. 7. The subarea images marked in Fig. 6(a).

The quantitative evaluation data of the experimental results are shown in Table 2. For all six bands, the predicted image of FSDAF 2.0 has the smaller RMSE, higher SSIM and r compared with those of STARFM and FSDAF. This suggests that FSDAF 2.0 is more powerful for retrieving spectral and structural information of the surface. Furthermore, FSDAF 2.0 has better performance than SFSDAF except in the NIR band. Taking the blue band as an example, FSDAF 2.0 has a 6.0% improvement over FSDAF and a 4.1% improvement over SFSDAF according to RMSE. Scatterplots of the blue band shown in Fig. 8 also suggest that the values predicted by FSDAF 2.0 are closer to the actual values than those predicted by the other three methods (e.g., the area indicated by the blue arrow in the scatter plot of FSDAF 2.0). Moreover, to further test the robustness of FSDAF 2.0 in the face of real data, MOD09GA images were also used for fusion as an additional experiment. Similar to the experiment in heterogeneous landscape, the SFSDAF was not used to participate in this comparison. The experimental results are shown in Table 3. Apparently, FSDAF 2.0 provided the most accurate prediction, and the progress is obvious in the SWIR1 band and SWIR2 band, which had the most change when flooded. Consequently, FSDAF 2.0 can better retrieve pixels that have undergone large-scale land cover type-change events.

Table 2

Accuracy assessment of STARFM, FSDAF, SFSDAF and FSDAF 2.0 in a landscape with large-scale change in land cover type by inputting simulated MODIS data as coarse-resolution images. The units are reflectance (RMSE = root mean square error, SSIM = structural similarity, r = correlation coefficient).

Band	STARFM			FSDAF			SFSDAF			FSDAF 2.0		
	RMSE	SSIM	r	RMSE	SSIM	r	RMSE	SSIM	r	RMSE	SSIM	r
Blue	0.0106	0.9781	0.8454	0.0100	0.9783	0.8617	0.0098	0.9794	0.8696	**0.0094**	**0.9812**	**0.8792**
Green	0.0153	0.9694	0.8486	0.0146	0.9689	0.8642	0.0142	0.9710	0.8721	**0.0138**	**0.9728**	**0.8791**
Red	0.0190	0.9589	0.8515	0.0180	0.9589	0.8680	0.0174	0.9623	0.8774	**0.0171**	**0.9639**	**0.8815**
NIR	0.0340	0.9111	0.8476	0.0296	0.9190	0.8868	**0.0284**	**0.9288**	**0.8964**	0.0290	0.9250	0.8925
SWIR1	0.0472	0.8055	0.8138	0.0452	0.7994	0.8311	0.0435	**0.8178**	0.8459	**0.0429**	0.8171	**0.8494**
SWIR2	0.0338	0.8467	0.8113	0.0319	0.8508	0.8352	0.0311	0.8601	0.8444	**0.0305**	**0.8606**	**0.8501**

Table 3

Accuracy assessment of STARFM, FSDAF and FSDAF 2.0 in a landscape with large-scale change in land cover type by inputting MOD09GA data as coarse-resolution images. The units are reflectance (RMSE = root mean square error, SSIM = structural similarity, r = correlation coefficient).

Band	STARFM			FSDAF			FSDAF 2.0		
	RMSE	SSIM	r	RMSE	SSIM	r	RMSE	SSIM	r
Blue	0.0160	0.9577	0.6791	0.0157	0.9577	0.6923	**0.0151**	**0.9608**	**0.6955**
Green	0.0224	0.9507	0.6801	0.0218	0.9512	0.6934	**0.0217**	**0.9526**	**0.6968**
Red	0.0280	0.9360	0.6819	**0.0269**	0.9391	0.6963	**0.0269**	**0.9403**	**0.6988**
NIR	0.0416	0.8994	0.7810	0.0401	0.9027	**0.8022**	**0.0399**	**0.9061**	0.8010
SWIR1	0.0612	0.7400	0.6952	0.0628	0.7396	0.6988	**0.0598**	**0.7530**	**0.7128**
SWIR2	0.0470	0.7568	0.6801	0.0493	0.7565	0.6746	**0.0461**	**0.7790**	**0.6957**

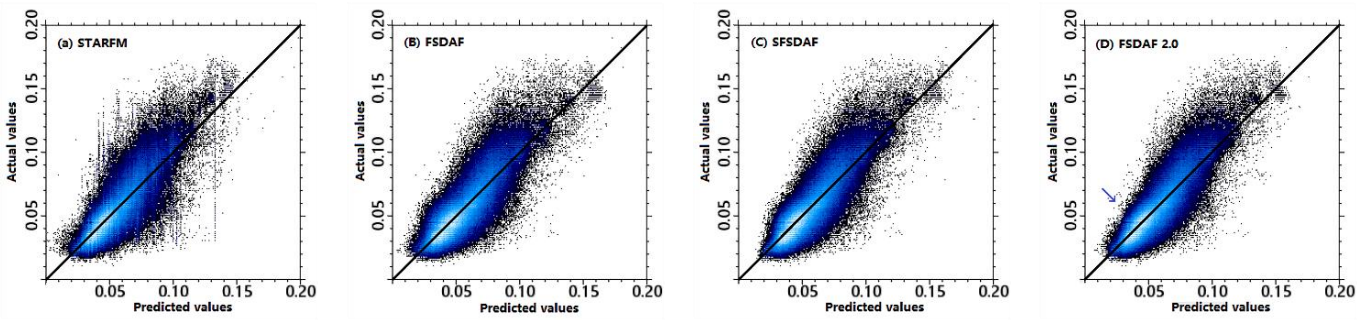


Fig. 8. Scatterplots of the actual and predicted values for the blue band (brighter color indicates a higher density of points, the line is 1:1 line).

5. Discussion

The proposed spatiotemporal data fusion model FSDAF 2.0 shows satisfactory performance in two experiments. In particular, compared with the original FSDAF, FSDAF 2.0 can more accurately capture the ecosystem dynamics and changing type boundaries of objects and retain more details. In this section, a theoretical comparison of the rationale behind the key steps in retrieving land cover changes between FSDAF and FSDAF 2.0 and how FSDAF 2.0 outperforms FSDAF are discussed. Moreover, a comparative experiment of the various steps in FSDAF and FSDAF 2.0 was added. In addition, the efficiency of the algorithm should be considered in the application; thus, a computation time comparison among the four methods was discussed. Finally, we discussed the further improvement of FSDAF 2.0.

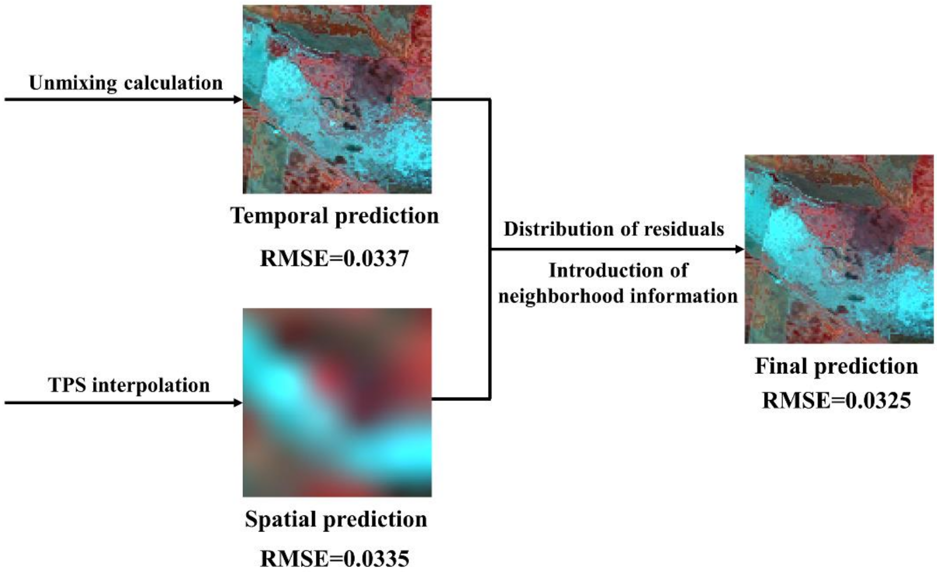
5.1.Comparison of the processes of FSDAF and FSDAF 2.0 in retrieving land cover changes

The results of experiment 2 in section 4 demonstrate that both FSDAF and FSDAF 2.0 have the ability to retrieve land cover changes. The ability of FSDAF to retrieve changed pixels is mainly come from the spatial prediction. Theoretically, spatial prediction describes the information of the real surface in the T_2 phase, which can maintain the signals of land cover type change and local variability in the fusion result (Zhu et al., 2016). In the following process, however, FSDAF distributes residuals on the assumption that errors depend mainly on the homogeneity of the surface. This strategy guarantees that FSDAF can preserve more detailed information but limits its ability to retrieve land cover changes.

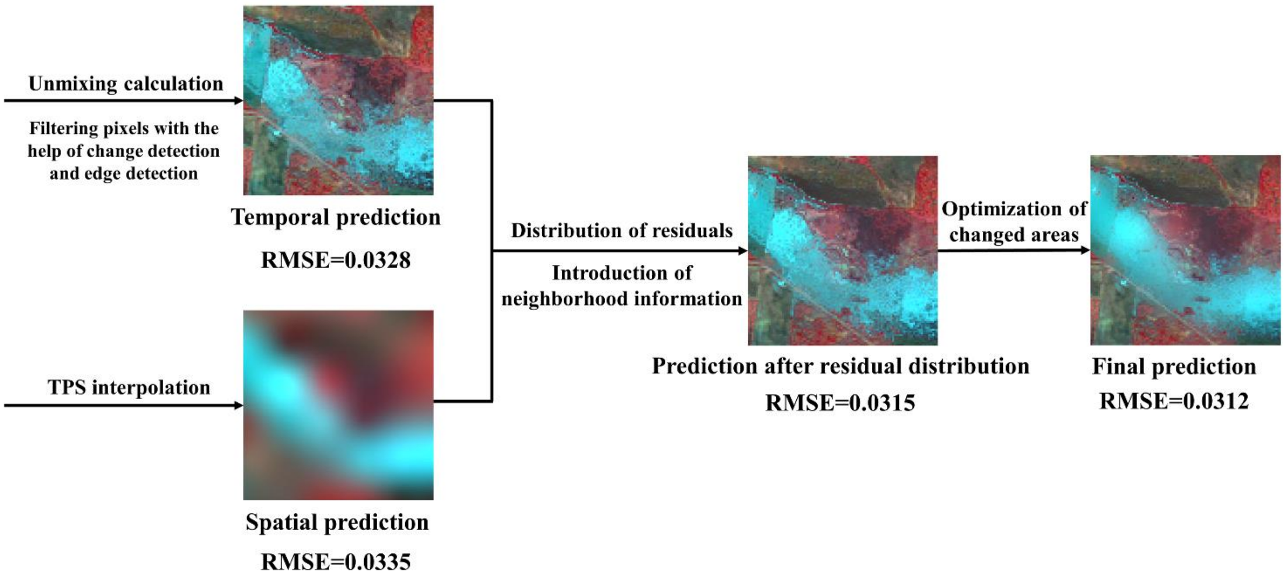
The reconversion capability of FSDAF 2.0 comes from the spatial prediction and the optimization model for the changed pixels. FSDAF 2.0 employs edge detection technology and change detection technology to exclude the coarse pixels that contain changed pixels or more than 10% of the boundary pixels. As a result, FSDAF 2.0 obtains more accurate predicted values in the area where the types of objects are unchanged. These differences are the main reason why FSDAF 2.0 obtains more details and more accurate spectral information than the original FSDAF. In addition, FSDAF 2.0 establishes an optimization model for changed areas in the final step to offset the error caused by unreasonable assumptions in the residual distribution process. Theoretically, FSDAF 2.0 can achieve higher accuracy.

To demonstrate the above points explicitly, a comparison of various steps in FSDAF and FSDAF 2.0 to retrieve land cover changes was added. As shown in Fig. 9, the experimental data (2.5 km × 2.5 km, 100 × 100 pixels) is the subarea E in experiment 2. Fig. 9(a) and (b) show the coarse-resolution images at T_1 and T_2. Fig. 9(c) and (d) show the fine-resolution images at T_1 and T_2. Apparently, this area experienced flooding and phenological changes during fusion period. Fig. 9(e) and Fig. 9(f) show the processes of FSDAF and FSDAF 2.0, respectively, in retrieving land cover changes. Compared with the temporal predictions of the two methods, it is obvious that the temporal prediction of FSDAF is “blurrier”, it is difficult to distinguish the boundary of the flood effects and has significant spectral errors around the flood. In addition, these problems also exist in the final prediction. While the temporal prediction of FSDAF 2.0 is closer to that of the real image than that of FSDAF, not only in spatial details but also in spectrum, quantitative analysis also confirmed this conclusion (average RMSE of 0.0337 vs. 0.0328). Furthermore, the accuracies of the predictions are improved after the following residual distribution process and optimization process, as shown in Fig. 9(f). The flowing water in the final prediction of FSDAF 2.0 seems to be smoother, closer to the actual flowing water in Fig. 9(d), and the newly added flood areas are more visible and easier to distinguish. Quantitative analysis also confirmed that the overall accuracy is gradually improved after residual distribution and optimization, and the average RMSE values are 0.0328, 0.0315, and 0.0312. The average RMSE values of the temporal prediction and final prediction of FSDAF are 0.0337 and 0.0325, respectively, which is worse than that of FSDAF 2.0. Consequently, the processes of improving the temporal prediction and increasing targeted optimization for changed areas in FSDAF 2.0 make the final prediction more accurate.

(a) (b) (c) (d)



(e)



(f)

Fig. 9. The comparison of the key steps in retrieving land cover changes between FSDAF and FSDAF 2.0: the coarse-resolution images of subarea E in experiment 2 at T_1 (a) and T_2 (b), the fine-resolution images of subarea E in experiment 2 at T_1 (c) and T_2 (d), the processes of FSDAF (e) and FSDAF 2.0 (f) in retrieving land cover changes.

5.2.Comparison of computation time

The computation times of the four methods in section 4 are shown in Table 4. The calculation platform used in two experiments is i7-6700HQ (2.60 GHz) and 16 G RAW. The results show that FSDAF 2.0 consumes more time due to more steps than in the original FSDAF, but not too much time; it has comparable efficiency with STARFM and SFSDAF because effective but less computation algorithms were employed in the additional steps. Considering the advantages of FSDAF 2.0 over the other three algorithms, its efficiency is acceptable.

Table 4

The computation time of STARFM, FSDAF, SFSDAF and FSDAF 2.0 in two experiments.

	Experiment 1	Experiment 2
STARFM	228 s	4574 s
FSDAF	189 s	3757 s
SFSDAF	/	4553 s
FSDAF 2.0	245 s	4629 s

5.3.Further improvement of FSDAF 2.0

The results of the experiments in section 4 demonstrate that FSDAF 2.0 can obtain satisfactory overall accuracy in two challenging landscapes: heterogeneous and large-scale abrupt land cover changes. The blending results of the improved method have higher overall spectral accuracy, more similar structure and closer correlation to real images, especially in areas where the types of land cover changed. These improvements are due to overcoming the shortcomings of the original FSDAF. Although FSDAF 2.0 has satisfactory performance, it still has the potential to improve.

First, the improvement of FSDAF 2.0 can mainly be achieved through improved temporal prediction and increased targeted optimization in the final step, but the step of obtaining spatial prediction is consistent with that of the original FSDAF; for example, the result of TPS interpolation is used as the spatial prediction. However, the TPS interpolation image is "smooth" and loses many spatial details; if the spatial prediction could be replaced by a better scale-down algorithm without consuming too much time, FSDAF could theoretically retain more image details.

Second, similar to FSDAF, FSDAF 2.0 still distributes residuals on the assumption that errors depend mainly on the homogeneity of the surface. This strategy is very empirical and has no theoretical basis. It may not be an optimal way to distribute residuals for different scenarios (Meng Liu et al., 2019). Theoretically, a more rigorous method of weight assignment can improve this problem.

Third, on account of the lack of fine-resolution image in the T_2 phase, FSDAF 2.0 employs the TPS interpolation images of coarse-resolution images in two phases to detect changed areas. Therefore, it is difficult to capture tiny land cover changes. Theoretically, fine-resolution images acquired from other satellites can be employed to solve this problem in the future research process. In addition, long time-series observations or the use of more flexible change detection algorithms can also improve the performance and robustness of FSDAF 2.0.

6. Conclusions

This study described the theoretical basis, implementation process and performance of an improved flexible spatiotemporal data fusion method incorporating change detection technology and an optimized model for changed-type areas. Landsat and MODIS images of two different sites were employed to test the performance of the improved method. All results demonstrate that FSDAF 2.0 improves the shortcomings of FSDAF, blends synthetic fine-resolution images with higher accuracy in different landscapes, and strengthens the robustness of the algorithm and the ability of retrieving land cover changes compared with those of the original FSDAF algorithm. In addition, FSDAF 2.0 has acceptable efficiency even though it has more steps than the original FSDAF, because effective but fewer computation algorithms were employed in the additional steps.

The key idea of FSDAF 2.0 is using the change detection technology to label the changed pixels. This is a precondition for the subsequent improvement of the unmixing step and targeted optimization, which effectively helps improve the fusion accuracy in changed-type areas. In the spatiotemporal fusion field, retrieving land cover changes is a challenge, and FSDAF 2.0 provides a feasible way to overcome this problem. Moreover, this field has great potential for improvement, such as improving the accuracy of change detection through long time-series observations or using other satellite data to assist in change detection.

Similar to FSDAF and other spatiotemporal methods, FSDAF 2.0 can also be used to blend other products that are derived from reflectance data, e.g., normalized difference vegetation index (NDVI), surface temperature and leaf area index. FSDAF has been shown to have high accuracy in fusing other products (Meng Liu et al., 2019; Alves et al., 2018). FSDAF 2.0 retains the advantages of FSDAF and improves fusion performance. Theoretically, FSDAF 2.0 can achieve higher accuracy in fusing other products compared with that of FSDAF.

In conclusion, the FSDAF 2.0 algorithm improves the capability for blending fine-resolution remote sensing images, especially for areas of land cover changes. This improvement is beneficial for monitoring the land surface and dynamics of our Earth systems.

Acknowledgements

The work presented in this paper is supported by the Ministry of Science and Technology of the People's Republic of China (2017YFB0503604), and National Natural Science Foundation of China (41701504). The authors would like to thank the Editor and all reviewers whose insightful suggestions have significantly improved this paper.

REFERENCES

Alves, D.B., Llovería, R.M., Pérez-cabello, F., Borini, D., Llovería, R.M., Pérez-cabello, F., 2018. Fusing Landsat and MODIS data to retrieve multispectral information from fire-affected areas over tropical savannah environments in the Brazilian Amazon Fusing Landsat and MODIS data to retrieve multispectral. Int. J. Remote Sens. 39, 7919–7941. https://doi.org/10.1080/01431161.2018.1479790.

Amorós-lópez, J., Gómez-chova, L., Alonso, L., Guanter, L., Zurita-milla, R., Moreno, J., Camps-valls, G., 2013. Multitemporal fusion of Landsat/TM and ENVISAT/MERIS for crop monitoring. Int. J. Appl. Earth Obs. Geoinf. 23, 132–141. https://doi.org/10.1016/j.jag.2012.12.004.

Battude, M., Bitar, A. Al, Morin, D., Cros, J., Huc, M., Sicre, C.M., Dantec, V. Le, Demarez, V., 2016. Estimating maize biomass and yield over large areas using high spatial and temporal resolution Sentinel-2 like remote sensing data. Remote Sens. Environ. 184, 668–681. https://doi.org/10.1016/j.rse.2016.07.030.

Dubrule, O., 1984. Comparing splines and kriging. Computers & Geosciences, 10,327–338. http://dx.doi.org/10.1016/0098-3004(84)90030-X.

Emelyanova, I. V, Mcvicar, T.R., Niel, T.G. Van, Tao, L., Dijk, A.I.J.M. Van, 2013. 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. 133, 193–209. https://doi.org/10.1016/j.rse.2013.02.007.

Gao, F., Masek, J., Schwaller, M., Hall, F., 2006. On the blending of the landsat and MODIS surface reflectance: Predicting daily landsat surface reflectance. IEEE Trans. Geosci. Remote Sens. 44, 2207–2218. https://doi.org/10.1109/TGRS.2006.872081.

Hilker, T., Wulder, M.A., Coops, N.C., Linke, J., Mcdermid, G., Masek, J.G., Gao, F., White, J.C., 2009. A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on Landsat and MODIS. Remote Sens. Environ. 113, 1613–1627. https://doi.org/10.1016/j.rse.2009.03.007.

Huang, B., Song, H., 2012. Spatiotemporal reflectance fusion via sparse representation. IEEE Trans. Geosci. Remote Sens. 50, 3707–3716. https://doi.org/10.1109/TGRS.2012.2186638.

Lai, Y., Rosin, P.L., 2014. Efficient Circular Thresholding. IEEE Trans. Image Process. 23, 992–1001. https://doi.org/10.1109/TIP.2013.2297014.

Li, A., Bo, Y., Zhu, Y., Guo, P., Bi, J., He, Y., 2013. Blending multi-resolution satellite sea surface temperature (SST) products using Bayesian maximum entropy method. Remote Sens. Environ. 135, 52–63. https://doi.org/10.1016/j.rse.2013.03.021.

Liao, L., Song, J., Wang, Jindi, Xiao, Z., Wang, Jian, 2016. Bayesian Method for Building Frequent Landsat-Like NDVI Datasets by Integrating MODIS and Landsat NDVI. Remote Sens. https://doi.org/10.3390/rs8060452.

Lilliefors, H.W., 2012. On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. J. Am. Stat. Assoc. 1459. https://doi.org/10.2307/2283970.

Liu, Maolin, Ke, Y., Yin, Q., Chen, X., Im, J., 2019. Comparison of Five Spatio-Temporal Satellite Image Fusion Models over Landscapes with Various Spatial Heterogeneity and Temporal Variation. Remote Sens. 11, 2612. https://doi.org/10.3390/rs11222612.

Liu, Meng, Yang, W., Zhu, X., Chen, J., Chen, X., Yang, L., Helmer, E.H., 2019. An Improved Flexible Spatiotemporal DAta Fusion (IFSDAF) method for producing high spatiotemporal resolution normalized difference vegetation index time series. Remote Sens. Environ. 227, 74–89. https://doi.org/10.1016/j.rse.2019.03.012.

Li, X., Foody, G.M., Boyd, D.S., Ge, Y., Zhang, Y., Du, Y., Ling, F., 2020. SFSDAF: An enhanced FSDAF that incorporates sub-pixel class fraction change information for spatio-temporal image fusion. Remote Sens. Environ. 237, 111537. https://doi.org/10.1016/j.rse.2019.111537.

Moosavi, V., Talebi, A., Hossein, M., Rashid, S., Shamsi, F., Niazi, Y., 2015. A wavelet-artificial intelligence fusion approach (WAIFA) for blending Landsat and MODIS surface temperature. Remote Sens. Environ. 169, 243–254.

https://doi.org/10.1016/j.rse.2015.08.015.

OTSU, N., 1979. A Threshold Selection Method from Gray-Level Histograms. IEEE Trans. SYSTREMS, MAN, Cybern. 20, 62–66. https://doi.org/10.1109/tsmc.1979.4310076.

Panigrahy, R.K., Panigrahy, S.S.R.S., 2009. Study on the Utility of IRS-P6 AWiFS SWIR Band for Crop Discrimination and Classification. J. Indian Soc. Remote Sens. 325–333. https://doi.org/10.1007/s12524-009-0026-6.

ROYSTON, P., 2000. Approximating the Shapiro-Wilk W-test for non-normality. Stat. Comput. 3–5. https://doi.org/10.1007/BF01891203.

Rudorff, N., Rudorff, C.M., Kampel, M., Ortiz, G., 2018. Remote sensing monitoring of the impact of a major mining wastewater disaster on the turbidity of the Doce River plume off the eastern Brazilian coast. ISPRS J. Photogramm. Remote Sens. 145, 349–361. https://doi.org/10.1016/j.isprsjprs.2018.02.013.

Sadeghi, M., Jones, S.B., Philpot, W.D., 2015. A linear physically-based model for remote sensing of soil moisture using short wave infrared bands. Remote Sens. Environ. 164, 66–76. https://doi.org/10.1016/j.rse.2015.04.007.

Shen, M., Tang, Y., Chen, J., Zhu, X., Zheng, Y., 2011. Agricultural and Forest Meteorology Influences of temperature and precipitation before the growing season on spring phenology in grasslands of the central and eastern Qinghai-Tibetan Plateau. Agric. For. Meteorol. 151, 1711–1722. https://doi.org/10.1016/j.agrformet.2011.07.003.

Shi, C., Wang, X., Zhang, M., Liang, X., Niu, L., Han, H., Zhu, X., 2019. A Comprehensive and Automated Fusion Method: The Enhanced Flexible Spatiotemporal DAta Fusion

Model for Monitoring Dynamic Changes of Land Surface. Appl. Sci. 1–19. https://doi.org/10.3390/app9183693.

Song, C., Woodcock, C.E., Seto, K.C., Lenney, M.P., Macomber, S.A., 2000. Classification and Change Detection Using Landsat TM Data : When and How to Correct Atmospheric Effects? Remote Sens. Environ. 4257. https://doi.org/10.1016/S0034-4257(00)00169-3.

Song, H., Huang, B., 2013. Spatiotemporal satellite image fusion through one-pair image learning. IEEE Trans. Geosci. Remote Sens. 51, 1883–1896. https://doi.org/10.1109/TGRS.2012.2213095.

Sun, Y., Zhang, H., 2019. A Two-Stage Spatiotemporal Fusion Method for Remote Sensing Images. Photogramm. Eng. Remote Sens. 907–914. https://doi.org/10.14358/PERS.85.12.907.

Sun, Y., Zhang, H., Shi, W., 2018. A spatio-temporal fusion method for remote sensing data Using a linear injection model and local neighbourhood information. Int. J. Remote Sens. 00, 1–21. https://doi.org/10.1080/01431161.2018.1538585.

Tan, Z., Peng, Y., Di, L., Tang, J., 2018. Deriving High Spatiotemporal Remote Sensing Images Using Deep Convolutional Network. Remote Sens. 1–16. https://doi.org/10.3390/rs10071066.

Taubenböck, H., Esch, T., Felbier, A., Wiesner, M., Roth, A., Dech, S., 2012. Monitoring urbanization in mega cities from space. Remote Sens. Environ. 117, 162–176. https://doi.org/10.1016/j.rse.2011.09.015.

Wang, Q., Atkinson, P.M., 2018. Spatio-temporal fusion for daily Sentinel-2 images. Remote

Sens. Environ. 204, 31–42. https://doi.org/10.1016/j.rse.2017.10.046.

Wang, Q., Zhang, Y., Onojeghuo, A.O., Zhu, X., Atkinson, P.M., 2017. Enhancing Spatio-Temporal Fusion of MODIS and Landsat Data by Incorporating 250 m MODIS Data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10, 4116–4123. https://doi.org/10.1109/JSTARS.2017.2701643.

Wu, B., Huang, B., Zhang, L., 2015. An Error-Bound-Regularized Sparse Coding for Spatiotemporal Reflectance Fusion. IEEE Trans. Geosci. Remote Sens. 53, 6791–6803. https://doi.org/10.1109/TGRS.2015.2448100.

Wu, M., Niu, Z., Wang, C., Wu, C., Wang, L., 2012. Use of MODIS and Landsat time series data to generate high-resolution temporal synthetic Landsat data using a spatial and temporal reflectance fusion model. J. Appl. Remote Sens. https://doi.org/10.1117/1.JRS.6.063507.

Y. Yamaguchi, Naito, C., 2010. Spectral indices for lithologic discrimination and mapping by using the ASTER SWIR bands. Int. J. Remote Sens. 1161. https://doi.org/10.1080/01431160110070320.

Zhang, F., Zhu, X., Liu, D., 2014. Blending MODIS and Landsat images for urban flood mapping. Int. J. Remote Sens. 35, 3237–3253. https://doi.org/10.1080/01431161.2014.903351.

Zhang, H.K., Huang, B., Zhang, M., Cao, K., Yu, L., 2015. A generalization of spatial and temporal fusion methods for remotely sensed surface parameters. Int. J. Remote Sens. 36, 4411–4445. https://doi.org/10.1080/01431161.2015.1083633.

Zhang, W., Li, A., Jin, H., Bian, J., Zhang, Z., Lei, G., Qin, Z., Huang, C., 2013. An Enhanced Spatial and Temporal Data Fusion Model for Fusing Landsat and MODIS Surface Reflectance to Generate High Temporal Landsat-Like Data. Remote Sens. 5346–5368. https://doi.org/10.3390/rs5105346.

Zhu, X., Cai, F., Tian, J., Williams, T.K.A., 2018. Spatiotemporal Fusion of Multisource Remote Sensing Data: Literature Survey, Taxonomy, Principles, Applications, and Future Directions. Remote Sens. 10, 527. https://doi.org/10.3390/rs10040527.

Zhu, X., Chen, J., Gao, F., Chen, X., Masek, J.G., 2010. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. Remote Sens. Environ. 114, 2610–2623. https://doi.org/10.1016/j.rse.2010.05.032.

Zhu, X., Helmer, E.H., Gao, F., Liu, D., Chen, J., Lefsky, M.A., 2016. A flexible spatiotemporal method for fusing satellite images with different resolutions. Remote Sens. Environ. 172, 165–177. https://doi.org/10.1016/j.rse.2015.11.016.

Zhu, X., Leung, K.H., Li, W.S., Cheung, L.K., 2019. Monitoring interannual dynamics of desertification in Minqin County , China , using dense Landsat time series. Int. J. Digit. Earth 0, 1–13. https://doi.org/10.1080/17538947.2019.1585979.

Zhukov, B., Oertel, D., Lanzl, F., Reinhäckel, G., 1999. Unmixing-based multisensor multiresolution image fusion. IEEE Trans. Geosci. Remote Sens. 37, 1212–1226. https://doi.org/10.1109/36.763276.

Zurita-milla, R., Clevers, J.G.P.W., Schaepman, M.E., Member, S., 2008. Unmixing-Based Landsat TM and MERIS FR Data Fusion. IEEE Geosci. Remote Sens. Lett. 5, 453–457.

https://doi.org/10.1109/LGRS.2008.919685.

List of Figure Captions

Figure number	Figure caption
Fig. 1.	Flowchart of FSDAF 2.0
Fig. 2.	The process of filtering pixels for unmixing calculation in FSDAF 2.0
Fig. 3.	Experimental data in a heterogeneous landscape: Landsat images (600×600 pixels) were acquired on (a) 04 December 2001 and (b) 12 January 2002; (c) and (d) are MOD09GA images. All images use NIR-red-green as RGB, and MOD09GA images are resampled to have the same size as the Landsat images.
Fig. 4.	Experimental data in a large-scale abrupt land cover change landscape: Landsat images (2400 × 2400 pixels) were acquired on (a) 26 November 2004 and (b) 12 December 2004, (c) and (d) are MODIS-like images aggregated from (a) and (b). All images use NIR-red-green as RGB, and MODIS-like images are resampled to have the same size as the Landsat images.
Fig. 5.	Original Landsat image of 12 January 2002 (a), and the predicted images by STARFM (b), FSDAF (c), and FSDAF 2.0 (d).
Fig. 6.	Original Landsat image of 12 December,2004 (a), its predicted images by STARFM (b), FSDAF (c), SFSDAF (d), FSDAF 2.0 (e), and the binary image of change detection result in FSDAF 2.0 (f).
Fig. 7.	The subarea images marked in Fig. 6(a).
Fig. 8.	Scatterplots of the actual and predicted values for the blue band (brighter color

	indicates a higher density of points, the line is 1:1 line).
Fig. 9.	The comparison of the key steps in retrieving land cover changes between FSDAF and FSDAF 2.0: the coarse-resolution images of subarea E in experiment 2 at T_1 (a) and T_2 (b), the fine-resolution images of subarea E in experiment 2 at T_1 (c) and T_2 (d), the processes of FSDAF (e) and FSDAF 2.0 (f) in retrieving land cover changes.