Blocks-removed spatial unmixing for downscaling MODIS images

Qunming Wang [a, *], Kaidi Peng [a], Yijie Tang [a, *], Xiaohua Tong [a], Peter M. Atkinson [b, c]

[a] College of Surveying and Geo-Informatics, Tongji University, 1239 Siping Road, Shanghai 200092, China

[b] Faculty of Science and Technology, Lancaster University, Lancaster LA1 4YR, UK

[c] Geography and Environment, University of Southampton, Highfield, Southampton SO17 1BJ, UK

*Corresponding authors. E-mail: wqm11111@126.com (Q. Wang); tangyijie@sina.com (Y. Tang)

Abstract: The Terra/Aqua MODerate resolution Imaging Spectroradiometer (MODIS) data have been used widely for global monitoring of the Earth's surface due to their daily fine temporal resolution. The spatial resolution of MODIS time-series (i.e., 500 m), however, is too coarse for local monitoring. A feasible solution to this problem is to downscale the coarse MODIS images, thus creating time-series images with both fine spatial and temporal resolutions. Generally, the downscaling of MODIS images can be achieved by fusing them with fine spatial resolution images (e.g., Landsat images) using spatio-temporal fusion methods. Among the families of spatio-temporal fusion methods, spatial unmixing-based methods have been applied widely owing to their lighter dependence on the available fine spatial resolution images. However, all techniques within this class of method suffer from the same serious problem, that is, the block effect, which reduces the prediction accuracy of spatio-temporal fusion. To our knowledge, almost no solution has been developed to tackle this issue directly. To address this need, this paper proposes a blocks-removed spatial unmixing (SU-BR) method, which removes the blocky artifacts by including a new constraint constructed based on spatial continuity. SU-BR provides a flexible framework suitable for any existing spatial unmixing-based spatio-temporal fusion method. Experimental results on a heterogeneous region, a homogeneous region and a region experiencing land cover changes show that SU-BR removes the blocks effectively and increases the prediction accuracy obviously in all three regions. SU-BR also outperforms two popular spatio-temporal

fusion methods. SU-BR, thus, provides a crucial solution to overcome one of the longest standing challenges in spatio-temporal fusion.

Keywords: MODIS, Landsat, Downscaling, Spatio-temporal fusion, Image fusion, Spatial unmixing, Block effect.

1. Introduction

Remote sensing technology has shown increasing importance for land cover change detection (Zhang et al., 2018) and environmental monitoring; for example, crop growth (Johnson et al., 2016), agricultural (Hansen et al., 2000) and carbon sequestration monitoring (Lees et al., 2018). Effective monitoring of land surface dynamics places great demands on the quality of remote sensing data, especially in terms of the spatial and temporal resolutions. Due to technical and budget limitations, however, remote sensing satellite sensors trade spatial resolution and temporal resolution. As a result, almost no satellite sensor can meet the demand for both fine spatial and temporal resolutions. For example, the MODIS sensor can acquire images for the same scene at least once per day, but the images are at a coarse spatial resolution of 500 m (250 m for the red and NIR bands). In contrast, Landsat sensors (e.g., Thematic Mapper (TM), Enhanced Thematic Mapper (ETM+) and Operational Land Imager (OLI)) can acquire images at a fine spatial resolution of 30 m, but they have a revisit period of up to 16 days. Also, the impact of cloud and shadow contamination can further limit the number of high-quality Landsat images (i.e., it generally requires more than 16 days to acquire an effective Landsat image) (Ju et al., 2008).

In recent years, spatio-temporal fusion approaches have been developed to create images with both fine spatial and temporal resolutions by blending the available temporally sparse, but fine spatial resolution images

with temporally dense, but coarse spatial resolution images (Belgiu et al., 2019; Chen et al., 2015; Zhu et al., 2018). Spatio-temporal fusion has been used widely in various applications, including prediction of fine spatial and temporal resolution land surface temperature (LST) (Huang et al., 2013; Wang et al., 2020a; Weng et al., 2014; Wu et al., 2015), normalized difference vegetation index (NDVI) (Meng et al., 2013; Tewes et al., 2015) and leaf area index (Houborg et al., 2016; Zhang et al., 2014). Generally, five types of spatio-temporal fusion approaches can be identified: spatial weighting-based (Gao et al., 2016; Hilker et al., 2009; Wang and Atkinson, 2018; Zhu et al., 2010), spatial unmixing-based (Busetto et al., 2008; Wu et al., 2012; Xu et al., 2015; Zhukov et al., 1999; Zurita-Milla et al., 2009), Bayesian-based (Li et al., 2013; Shen et al., 2016; Xue et al., 2017), learning-based (Das et al., 2016; Huang et al., 2012; Liu et al., 2016; Song et al., 2013; Wang et al., 2020b) and hybrid methods (Li et al., 2020a; Liu et al., 2019; Zhu et al., 2016). The spatial weighting-based model is a common spatio-temporal fusion method. The spatial and temporal adaptive reflectance fusion model (STARFM) proposed by Gao et al. (2006) is perhaps the earliest and the most widely-used spatial weighting-based method. The basic assumption of STARFM is that the temporal changes in the coarse and fine spatial resolution images are consistent, in which case the prediction can be seen simply as a combination of the known fine spatial resolution image and the fine spatial resolution temporal change image predicted from the coarse version. Based on STARFM, several approaches have been developed to enhance the performance of spatio-temporal fusion for heterogeneous areas and areas which include land cover changes (Hilker et al., 2009; Luo et al., 2018; Tang et al., 2020; Wang and Atkinson, 2018; Zhu et al., 2010).

Another main category of spatio-temporal fusion model is spatial unmixing. The basic principle of spatial unmixing-based methods is to predict the value (reflectance hereafter) of fine spatial resolution pixels (fine pixels hereafter) by applying unmixing algorithms to each coarse pixel (Gevaert and Garc á-Haro, 2015). The multisensor multiresolution technique (MMT) proposed by Zhukov et al. (1999) is one of the first spatial unmixing-based methods, and it underpins most existing spatial unmixing-based methods. The algorithm includes four operations: 1) classification of the available fine spatial resolution images to produce the

thematic land cover map; 2) calculation of the proportions of each land cover class in each coarse pixel by upscaling the thematic map produced in 1); 3) spatial unmixing of each coarse pixel to obtain the reflectance of each land cover type within it; 4) reconstruction of the fine spatial resolution image by assigning the predicted reflectance according to the land cover type of the fine pixel (Zhukov et al., 1999). Based on the MMT algorithm, increasing efforts have been made to develop spatial unmixing-based methods in recent years. Busetto et al. (2008) considered both the spatial distance and the spectral similarity between the neighboring coarse pixel and the target pixel when unmixing coarse pixels, where the spectral similarity is quantified using the spectral information of the known fine spatial resolution image. Zurita-Milla et al. (2008) applied the unmixing-based data fusion (UBDF) model to fuse Landsat TM and MERIS images for vegetation monitoring over heterogeneous landscapes. As an alternative to the use of the known fine spatial resolution image, Zurita-Milla et al. (2009) introduced the LGN5 land use database to derive the fractional composition of land cover classes within each coarse pixel. The Spatial Temporal Data Fusion Approach (STDFA) proposed by Wu et al. (2012) made fuller use of the known fine resolution image, which predicts the fine spatial resolution temporal change image from the coarse temporal change image by spatial unmixing. Amorós-López et al. (2013) utilized a regularization term in the cost function of the unmixing model to restrict the solution of the reflectance of each class using a pre-defined spectrum extracted from pure pixels in the coarse image. Gevaert and García-Haro (2015) introduced a Bayesian approach to constrain the unmixing process using the available prior spectral information. Xu et al. (2015) proposed an approach to reduce unmixing error by incorporating the class spectra predicted by other reliable spatial and temporal data fusion approaches such as STARFM. The linear spectral unmixing-based spatiotemporal data fusion model proposed by Liu et al. (2020) predicts a fine spatial resolution proportion image for each class (rather than the hard class labels in the methods mentioned above) by implementing linear spectral unmixing on the known fine spatial resolution image. Then, the fine spatial resolution proportion image is degraded to produce coarse proportions in the spatial unmixing model.

The spatial unmixing-based methods have several unique advantages. On the one hand, they have a light dependence on the number of available images. More specifically, most of the spatial unmixing-based methods require only one fine spatial resolution image at the known time to produce the land cover classification map, together with a coarse image at the prediction time for unmixing. Therefore, this type of method has limited data-dependence and is, thus, more flexible. This is different from spatial weighting-based methods, where at least one pair of coarse-fine spatial resolution images is required. On the other hand, the spatial unmixing-based methods do not require the coarse and fine spatial resolution images to have corresponding spectral bands (i.e., the same wavelength) (Gevaert and Garc á-Haro, 2015), while the spatial weighting-based methods place a strict requirement for the correspondence of spectral bands. This characteristic brings two benefits. First, spatial unmixing can be performed on coarse bands whose wavelengths are not available in the observed fine spatial resolution images, resulting in an increase in the spectral resolution of the fine spatial resolution images (Gevaert and Garc á-Haro, 2015). Second, auxiliary datasets such as fine (or even finer) spatial resolution land cover maps can be treated as a supplement or even replacement of the classification map produced from the fine spatial resolution multispectral images (e.g., Landsat images in most cases) to further increase the accuracy (Zurita-Milla et al., 2011).

Despite the above advantages, there exists a widely acknowledged problem in spatial unmixing-based methods: the block effect (Ma et al., 2018; Wang et al., 2020c), which means pixels of the same land cover class present different reflectances in spatially adjacent coarse pixels, resulting in visually obvious blocky artifacts within an object. The block effect exists commonly in spatial unmixing predictions. The reason for this phenomenon is that unmixing of different coarse pixels is implemented using different local windows. This means that different coarse pixels containing different spectral properties of land cover (even for the same class) are involved in unmixing spatially adjacent center pixels. As a result, the same land cover class in the spatially adjacent coarse pixels may be assigned different reflectances, which leads to blocky artifacts. Also, intra-class spectral variation, which caused mainly by heterogeneous spatial patterns and temporal changes in

land cover (especially for the same class), is responsible for blocky artifacts, as only one reflectance value is predicted for each land cover class in spatial unmixing. Thus, for the same class, the prediction of reflectance may have multiple equal realizations, and it always differs in the unmixing model for each coarse pixel. Generally, blocky artifacts occur most obviously at the boundary between neighboring coarse pixels in the prediction.

The block effect has been a main obstacle in spatial unmixing, which greatly influences the visual appearance of the predictions and, more importantly, the accuracy of spatio-temporal fusion. Several studies attempted to enhance the performance of spatial unmixing-based methods, such as by making fuller use of the known fine spatial resolution image and performing unmixing on temporal change image (Wu et al., 2012), exerting additional constraints to the prediction of reflectances (Xu et al., 2015) and combining with spatial weighting-based predictions (Zhu et al., 2016). Nevertheless, these approaches are not designed for tackling the blocky artifacts, which remain in the predictions.

This paper proposes a blocks-removed spatial unmixing (SU-BR) method to remove the blocky artifacts in spatial unmixing-based methods, and further, increase the accuracy of spatio-temporal fusion. SU-BR considers both the residual errors in the unmixing model and the difference in reflectances between the same land cover class in the neighboring pixels. It is an optimization method requiring a number of iterations to approach the optimal solution. There are two main advantages of SU-BR:

1) SU-BR can remove the blocky artifacts and increase the prediction accuracy simultaneously. SU-BR removes the blocks in spatial unmixing by exerting a new constraint according to the spatial continuity of land cover. The information (i.e., reflectance prediction) provided by neighboring pixels further enhances the reflectance predicted by the original spatial unmixing, thus, ensuring the spatial continuity and increasing the prediction accuracy. This method is performed by deeper spatial information mining of the observed data, and it does not require any additional data or prior knowledge.

2) SU-BR provides a general model for removing the blocky artifacts in spatial unmixing-based methods. It is a strategy applicable to any spatial unmixing-based methods, such as UBDF and STDFA. Furthermore, it is also compatible with other existing enhanced versions using different constraints (e.g., the class reflectance extracted from pure coarse pixels (Xu et al., 2015)). That is, the constraint of spatial continuity in SU-BR can potentially be jointly considered with many other constraints.

The remainder of this paper is organized into four sections. Section 2 summarizes the mechanisms of three typical spatial unmixing-based methods, explores the block effect problem and introduces explicitly the proposed SU-BR method. Section 3 implements experiments on three datasets to compare the performance of SU-BR with other blocks-removed methods. SU-BR is also compared with several popular spatio-temporal fusion methods. Section 4 further discusses the findings from the experiments and potential future research, followed by a conclusion in Section 5.

2. Methods

2.1. Existing spatial unmixing-based methods

This section illustrates briefly the common principle of three typical spatial unmixing-based methods, including UBDF, STDFA and the virtual image pair-based spatio-temporal fusion (VIPSTF) with spatial unmixing (VIPSTF-SU) recently proposed by Wang et al. (2020c). Two key assumptions can be summarized for spatial unmixing-based methods. The first is that the observed reflectance of a mixed pixel can be treated as the weighted sum of the sub-pixel level reflectances of different land cover classes within the pixel (i.e., the linear mixture model). The second is that the distribution of land cover remains stable between the known and prediction times. To predict the fine spatial resolution reflectance of land cover classes conveniently, we

usually solve a set of linear equations using the mixed reflectance of coarse pixels in a local window, by assuming that the neighboring coarse pixels share the same reflectance for the same land cover class. The calculation is performed for each band sequentially. For convenience, we illustrate the principles of the spatial unmixing-based methods based on a unified model for a single band. Specifically, the general linear mixture model can be written as

$$\mathbf{Q} = \mathbf{PE} + \boldsymbol{\varepsilon} \tag{1}$$

where $\boldsymbol{\varepsilon}$ is the residual error term. $\mathbf{Q}$ is an $N \times 1$ vector composed of the observed reflectances of the coarse pixels, where N is the number of coarse pixels in the local window. $\mathbf{E}$ is a $C \times 1$ vector composed of reflectances for all land cover classes (class reflectance hereafter) that needs to be solved and C is the number of land cover classes. $\mathbf{P}$ is an $N \times C$ matrix composed of the coarse proportions of the C classes in the N coarse pixels. $\mathbf{E}$ in Eq. (1) can be solved by the least squares method based on the objective function

$$\hat{\mathbf{E}} = \arg\min_{\mathbf{E}} R = \left\| \mathbf{PE} - \mathbf{Q} \right\|_2^2 \tag{2}$$

where R is the object quantifying the residual error of the linear mixture model. Spatial unmixing-based methods utilize a fine spatial resolution thematic map temporally close to the prediction time to synthesize the coarse proportions in $\mathbf{P}$. For $\mathbf{Q}$ and $\mathbf{E}$, however, they have different meanings in the three spatial unmixing-based methods, which are explained in detail in Appendix A. Note that for simplicity, $\mathbf{E}$ is called class reflectance hereafter, but its specific meaning for different spatial unmixing-based methods should be borne in mind.

2.2. The block effect in spatial unmixing-based methods

In spatial unmixing, the observed reflectances in the local window are used to predict the class reflectance $\mathbf{E}$ in Eq. (2). It is performed on each coarse pixel independently and the predicted class reflectance will be assigned only to the center coarse pixel in the moving window. For neighboring pixels containing the same

class, the predicted class reflectance may be different, rendering obvious regular blocky artifacts with a spatial size of the coarse pixel. This phenomenon is called the block effect as introduced above. It is produced mainly due to intra-class spectral variation (caused by heterogeneity of spatial pattern and gradual temporal changes in land cover) and differences in the coarse data involved in the spatial unmixing models (i.e., the observed coarse data vector **Q** and coarse proportion matrix **P** in Eq. (2)) for neighboring coarse pixels. More precisely, for neighboring coarse pixels, there are two adjacent cases (i.e., side- and vertex-adjacent), where the proportions of different observed coarse data in the unmixing models need to be distinguished.

Fig. 1 shows an example to illustrate the two adjacent cases. In fact, if two coarse pixels are contiguous on one side, for the window size of $w \times w$ pixels, the different coarse pixels between the two windows account for a proportion of

$$F_1 = w \times 1/w^2 = 1/w . \tag{3}$$

In the other case where two neighboring coarse pixels are connected by a vertex, the proportion of different pixels is

$$F_2 = 1-(w-1)^2/w^2 = (2w-1)/w^2 . \tag{4}$$

These distinct coarse pixels are the essential reason for the block effect. Specifically, they correspond to different elements in **Q** and **P** in Eq. (2). Due to the intra-class spectral variation, these distinct pixels are essentially mixed with different class reflectances, even for the same class. Thus, the calculation based on Eq. (2) can lead to different solutions of class reflectances in **E** for the two adjacent pixels. The blocky artifacts reflect the intra-class spectral variation in fusion predictions at a coarse resolution, which is neglected within a coarse pixel.

An example is exhibited in Fig. 2 to illustrate the block effect, where w=3 is considered and the trapezoid represents an object shared by six neighboring coarse pixels. Fig. 2(a) is a spatial unmixing prediction with blocky artifacts, while Fig. 2(b) is the reference for the trapezoid object (i.e., the object is characterized by a constant gray value). For spatial unmixing of the center pixel I in Fig. 2(a), the nine pixels in the 3×3 window

marked in red are used. For coarse pixel II, it utilizes the six pixels in the red box and another three pixels in the blue box that represent the local window for pixel II. The three different coarse pixels in the adjacent regions contribute to different predictions of the values for the trapezoid object in pixels I and II.

As seen from Eqs. (3) and (4), the intensity of the block effect is related to the size of the moving window. When the window size w is larger, for both adjacent cases, the proportion of different coarse pixels between the two windows becomes smaller. That is, the larger window size, the less obvious is the phenomenon. A larger the window size, however, means more distant pixels are involved in the unmixing process, where all pixels are assumed to share the same class reflectance. According to Tobler's First Law of Geography (Tobler, 1970), the relation between two observations decreases gradually as the distance increases. The inclusion of more distant pixels will, thus, reduce the inherent intra-class spectral variation in the fusion results, and exacerbate the performances of spatial unmixing.

The block effect reduces the spatial continuity and dramatically affects the visual presentation. It limits the application of spatial unmixing-based methods in the field of spatio-temporal fusion. There is, therefore, a great need for a solution to remove the blocks to enhance spatial unmixing-based methods.

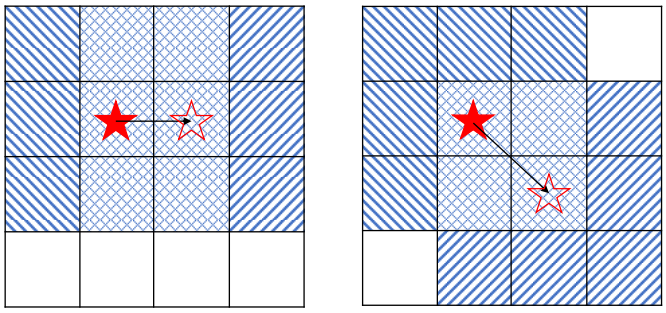
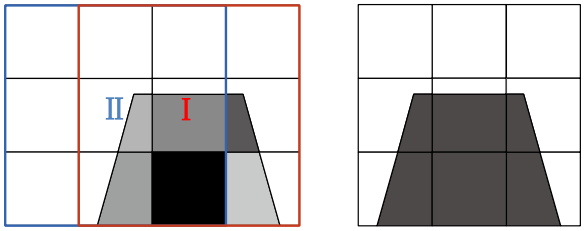


Fig. 1. An example for illustration of two adjacent cases (w=3). (a) and (b) represent the side- and vertex-adjacent cases, respectively. The pixels covered by diagonals at minus 45° represent distinct coarse pixels in a 3×3 window centered at the pixel marked by the red solid star. The pixels covered by diagonals at 45° represent distinct coarse pixels in a 3×3 window centered at the pixel marked by the red hollow star. The pixels covered by checks represent shared coarse pixels of the two local widows.



(a) (b)

Fig. 2. An example for illustration of the block effect. The trapezoid represents an object shared by neighboring coarse pixels. (a) is a prediction in which each part displays different colors. (b) is the reference image with fixed color.

2.3. The proposed constraint for removing blocks

In this section, a new constraint is proposed for removing blocks in spatial unmixing. The block effect essentially represents the difference in class reflectances between adjacent pixels. According to the spatial continuity of land cover, however, it can be assumed that the reflectances for the pixels belonging to the same class should be similar when the pixels are spatially adjacent (see, for example, the object in Fig. 2(b)). Based on this assumption, we can define a constraint by minimizing the difference between the reflectances for the same class in a local window for each coarse pixel, as shown in Eq. (5)

$$D_i = \frac{\sum_{c=1}^{C}\sum_{j=1}^{N_0}\left[I_{i,j,c}(E_{i,c}-E_{j,c})\right]^2}{\sum_{c=1}^{C}\sum_{j=1}^{N_0} I_{i,j,c}} \tag{5}$$

where D_i is the mean of the differences in reflectances for all classes in the local window centered at location $\mathbf{X}_i$, N_0 is the number of the neighbors in the local window (N_0 =8 is considered in this paper). $E_{i,c}$ and $E_{j,c}$ are the reflectances of class c for the center pixel at $\mathbf{X}_i$ and its neighboring pixel at $\mathbf{X}_j$, respectively. $I_{i,j,c}$ is an indicator function describing the relationship between the target coarse pixel at $\mathbf{X}_i$ and its neighboring pixel at $\mathbf{X}_j$

$$I_{i,j,c} = \begin{cases} 1, & \text{if pixels at } \mathbf{X}_i \text{ and } \mathbf{X}_j \text{ both contain class } c \\ 0, & \text{otherwise} \end{cases}. \quad (6)$$

D_i can be considered as the mean deviation of class reflectance between the target coarse pixel at $\mathbf{X}_i$ and its neighboring pixels, which is divided by the total number of deviations. The larger D_i, the more severe the spatial discontinuity in the window. To reduce the blocks, we need to minimize D_i for each window centered at $\mathbf{X}_i$. Thus, an objective function with the new constraint is proposed in the next section.

2.4. The proposed blocks-removed spatial unmixing (SU-BR) method

The main objective of the proposed blocks-removed spatial unmixing (SU-BR) method is also to minimize the residual error in the spatial unmixing model, as shown in Eq. (2). However, the constraint introduced in Section 2.3 is exerted on the new objective function to ensure the spatial continuity of class reflectance, thus, removing the blocks. Based on these two aspects, the new objective function for the proposed SU-BR method is provided below

$$\begin{aligned} \hat{\mathbf{E}}_i^{(t)} &= \arg\min_{\mathbf{E}_i^{(t)}} J_i = \alpha R_i^{(t)} + (1-\alpha) A D_i^{(t)} \\ &= \alpha \left\| \mathbf{P}\mathbf{E}_i^{(t)} - \mathbf{Q} \right\|_2^2 + (1-\alpha) A \frac{\sum_{c=1}^{C} \sum_{j=1}^{N_0} \left[I_{i,j,c} (E_{i,c}^{(t)} - E_{j,c}^{(t-1)}) \right]^2}{\sum_{c=1}^{C} \sum_{j=1}^{N_0} I_{i,j,c}} \end{aligned} \quad (7)$$

where α is a balancing parameter taking a value between 0 and 1, A is a magnitude regularization parameter and t is the iteration number. The value of indicator function $I_{i,j,c}$ is calculated based on the degraded thematic map at the known time.

SU-BR is performed for each coarse pixel in turn. Moreover, it is an optimization process based on iteration, as the class reflectance of the neighboring pixel is updated one-by-one in the visit, changing the constraint dynamically. The prediction based on the original spatial unmixing method is used directly for initialization

(i.e., the case of $t=0$). The optimization process terminates when one of the convergence conditions is satisfied: 1) the number of iterations reaches the pre-defined maximum number; 2) the difference between three consecutive realizations is smaller than a pre-defined threshold. With the iterative scheme in Eq. (7), the difference in reflectance for the same class can be reduced gradually to alleviate the block effect. Note that for heterogeneous areas, even though there may be several classes in the whole image, only a very small number of classes will cover a small region in a local window (e.g., with a size of 3×3 pixels in this paper), and the model in Eq. (7) is constructed adaptively for each coarse pixel centered at the local window. Therefore, the convergence for the model constructed in Eq. (7) can be guaranteed in this case.

It is necessary to determine appropriately the magnitude regularization parameter A, due to the difference in magnitudes of the two terms of R and D in the objective function. In this paper, it is proposed to be calculated by comparing the statistical information of R and D for all coarse pixels in the prediction of the original method. Specifically, the parameter A is determined by comparing the modes of the values of D and R.

To further understand Eq. (7), the first term R reflects the ability to preserve the original coarse spatial resolution image at the prediction time, which is called the data fidelity term. The second term D reflects the deviation in reflectance of the same class between the target coarse pixel and adjacent coarse pixels, which is the spatial continuity constraint term. The proposed SU-BR method makes a balance between maintaining the original coarse image and reducing the influence of the block effect. By changing the balancing parameter α, the influence of the two terms on the solution can be adjusted. With a larger balancing parameter α, the solution guarantees greater data fidelity, but may fail to remove blocky artifacts to the largest extent. A smaller balancing parameter may be able to remove the blocky artifacts satisfactorily, but may lead to a larger bias relative to the original data, resulting in lower accuracy of spatio-temporal fusion.

A flowchart describing the whole process of the SU-BR method is given in Fig. 3. The method is applicable to any spatial unmixing-based methods (e.g., UBDF, STDFA, and VIPSTF-SU investigated in this paper), based on the explicit definition of $\mathbf{Q}$ and $\mathbf{E}$, as illustrated in Appendix A. For UBDF, the predicted $\mathbf{E}$ is exactly

the final prediction for the method. For STDFA and VIPSTF-SU, the prediction represents temporal changes of the reflectances for classes and need to be added to the known fine spatial resolution image and virtual fine spatial resolution image, respectively, to achieve the final prediction of spatio-temporal fusion.

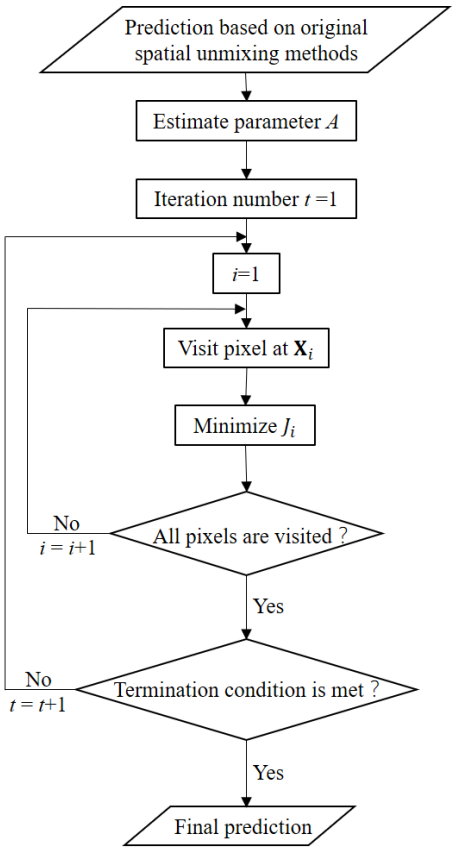


Fig. 3. Flowchart of the proposed SU-BR method. All bands of the image follow this scheme one by one.

2.5. Benchmark methods

This section focuses on two potential blocks-removed algorithms: neighbor mean (SU-NM) and spatial filtering (SU-SF). To the best of our knowledge, they have not been applied to remove blocks in spatial unmixing to-date. They can be implemented straightforwardly on fusion predictions of three typical spatial unmixing methods. The principles are introduced briefly below.

316 *1) SU-NM*

317 In the SU-NM method, the mean of the reflectances of the same class in a moving window will be assigned
318 to this class in the target coarse pixel. This process also needs $I_{i,j,c}$ in Eq. (6) to define the pixels containing
319 the same class. Its mechanism is similar to mean filtering in digital image processing, but it needs to identify
320 effective neighbors that cover the same class.

321 *2) SU-SF*

322 We apply the spatial filtering model in STARFM and enhanced STARFM (ESTARFM) to remove the
323 blocky artifacts by acknowledging the similarity of fine spatial resolution pixels and enabling similar pixels
324 (e.g., pixels belonging to the same class) to have close reflectance. This model was investigated in our
325 previous research (Wang and Atkinson, 2018) to remove the blocky artifacts produced from the local fitting
326 process (substantially different from the spatial unmixing process in this paper) and was shown to be a
327 satisfactory solution. The prediction of SU-SF is a linear combination of the reflectances of spectrally similar
328 neighboring pixels found in a moving window, weighted by the inverse spatial distance. However, it is
329 inappropriate to use the images with blocky artifacts to search for spectrally similar neighboring pixels.
330 Alternatively, the fine spatial resolution image at the known time is used, based on the assumption of stable
331 land cover boundaries during the period.

332 These two methods for removing blocks are applied in our experiments to provide a comparison with the
333 proposed SU-BR method and to validate its effectiveness.

334

335

336 **3. Experiments**

337

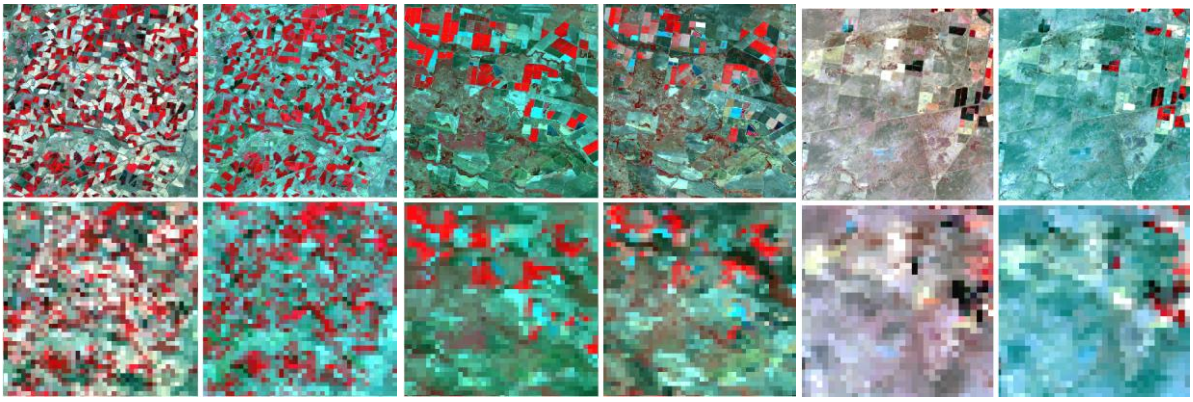
338 *3.1. Data and experimental setup*

339

To examine the performance of the proposed SU-BR method in areas with various spatial patterns, three datasets covering different spatial landscapes were used. The first region is located in southern New South Wales, Australia, and two Landsat 7 ETM+ images and two corresponding MODIS images were used. The spatial extent is 2 km by 2 km. The acquisition times of the two image pairs are 5 January 2002 and 13 February 2002. The second region is located in northern New South Wales, Australia, and has the same spatial size as the first region. The two image pairs were acquired on 14 February 2005 and 3 April 2005. As for the third region located in southern New South Wales, Australia, two MODIS and Landsat ETM+ image pairs covering a spatial extent of 1.8 km by 1.8 km were used. The acquisition times are 4 December 2001 and 5 January 2002. False color composites of the Landsat images and their corresponding MODIS images for the three regions are displayed in Fig. 4. The objective of the experiments is to predict the latter Landsat image, using the former MODIS-Landsat image pair and the latter MODIS image, and the known latter Landsat image is used as reference to evaluate the prediction. The former and latter times in this case are also called the known and prediction times hereafter.

It can be noticed that the first region shows obvious heterogeneity while the third region presents greater homogeneity. For the second region, due to the difference in acquisition seasons, many changes exist between the images acquired at the two times. Table 1 lists the correlation coefficients (CC) between the Landsat images at the two times for the three regions. It is obvious that the homogeneous region provides the greatest CC of 0.8593 between the dates, while the region with a greater number of land cover changes has the smallest CC of 0.6059. The CC of the heterogeneous region lies between the other two regions. Generally, the small CC between the known and prediction times will bring great challenges to the prediction. Three sub-sections (Sections 2.2-2.4) are included in the remainder of Section 3. Section 3.2 provides the results of the different blocks-removed methods based on existing UBDF, STDFA and VIPSTF-SU for the three regions. The blocks-removed methods for testing include the SU-BR, SU-SF and SU-NM methods. Section 3.3 compares

the performances of the proposed SU-BR method with the popular STARFM and FSDAF methods. Section 3.4 analyzes the impact of two parameters in SU-BR on the accuracy of the predictions.



(a) (b) (c) (d) (e) (f)

Fig. 4. Landsat (first line) and MODIS (second line) images for the heterogeneous region acquired on (a) 5 January 2002 and (b) 13 February 2002, for the region with land cover changes acquired on (c) 14 February 2005 and (d) 3 April 2005, and for the homogeneous region acquired on (e) 4 December 2001 and (f) 5 January 2002. All images use NIR-red-green as RGB.

Table 1 CCs between the Landsat images at the known and prediction times

	Heterogeneous region	Region with land cover changes	Homogeneous region
CC	0.7392	0.6059	0.8593

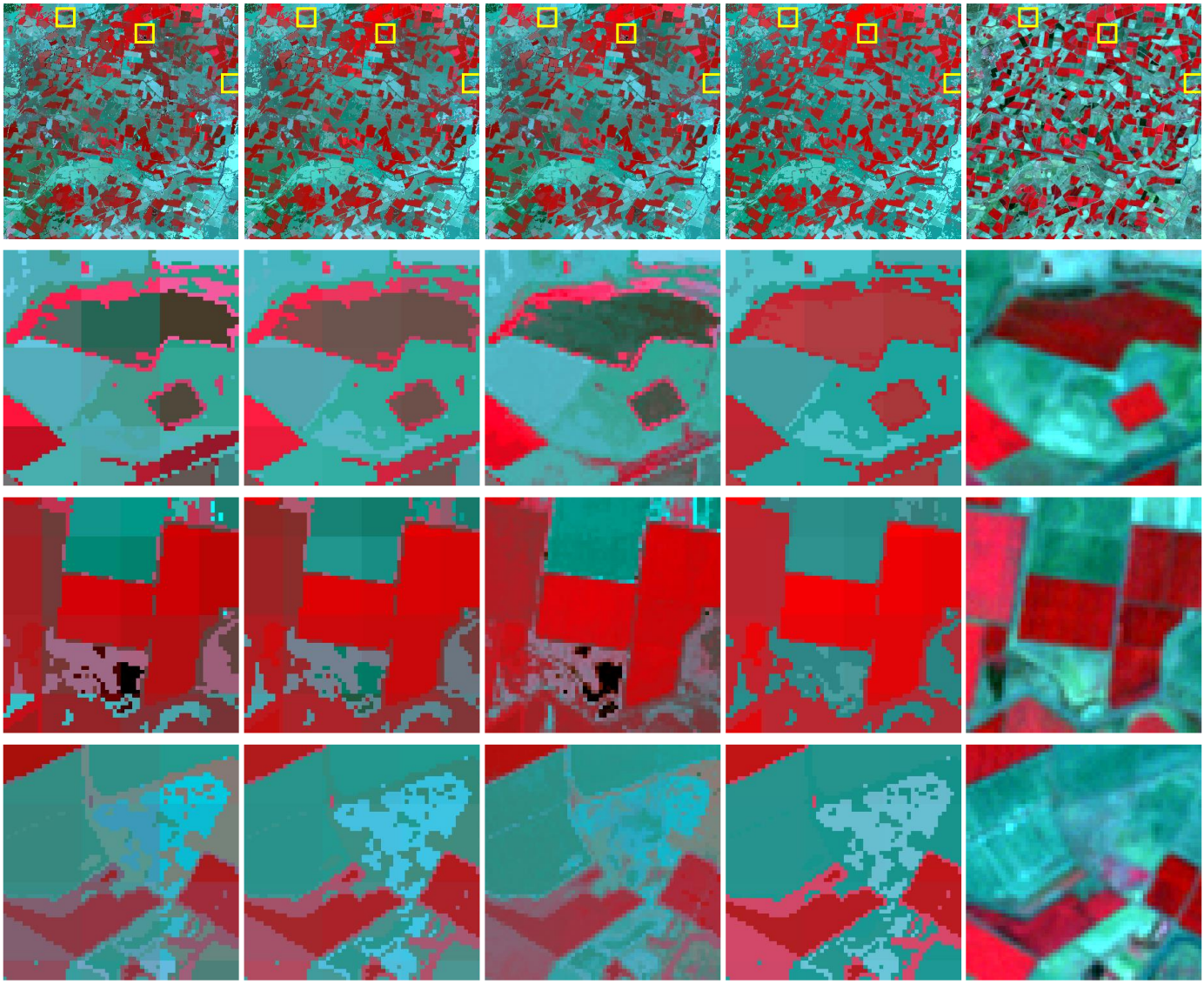
3.2. Comparison between different blocks-removed methods

3.2.1. Results for the heterogenous region

Figs. 5, 6 and 7 display the predictions of the three different blocks-removed methods (i.e., SU-BR, SU-NM and SU-SF) based on the three spatial unmixing methods (i.e., UBDF, STDFA and VIPSTF-SU). For clearer visual comparison between the results, three sub-areas covering 60 by 60 Landsat pixels are shown for each case. In Fig. 5, the UBDF-NM, UBDF-SF and UBDF-BR methods can all remove the blocks to some extent. Moreover, the UBDF-BR prediction is visually much closer to the reference than the other two blocks-removed methods, especially in the restoration of spectral properties; for example, the land cover in the

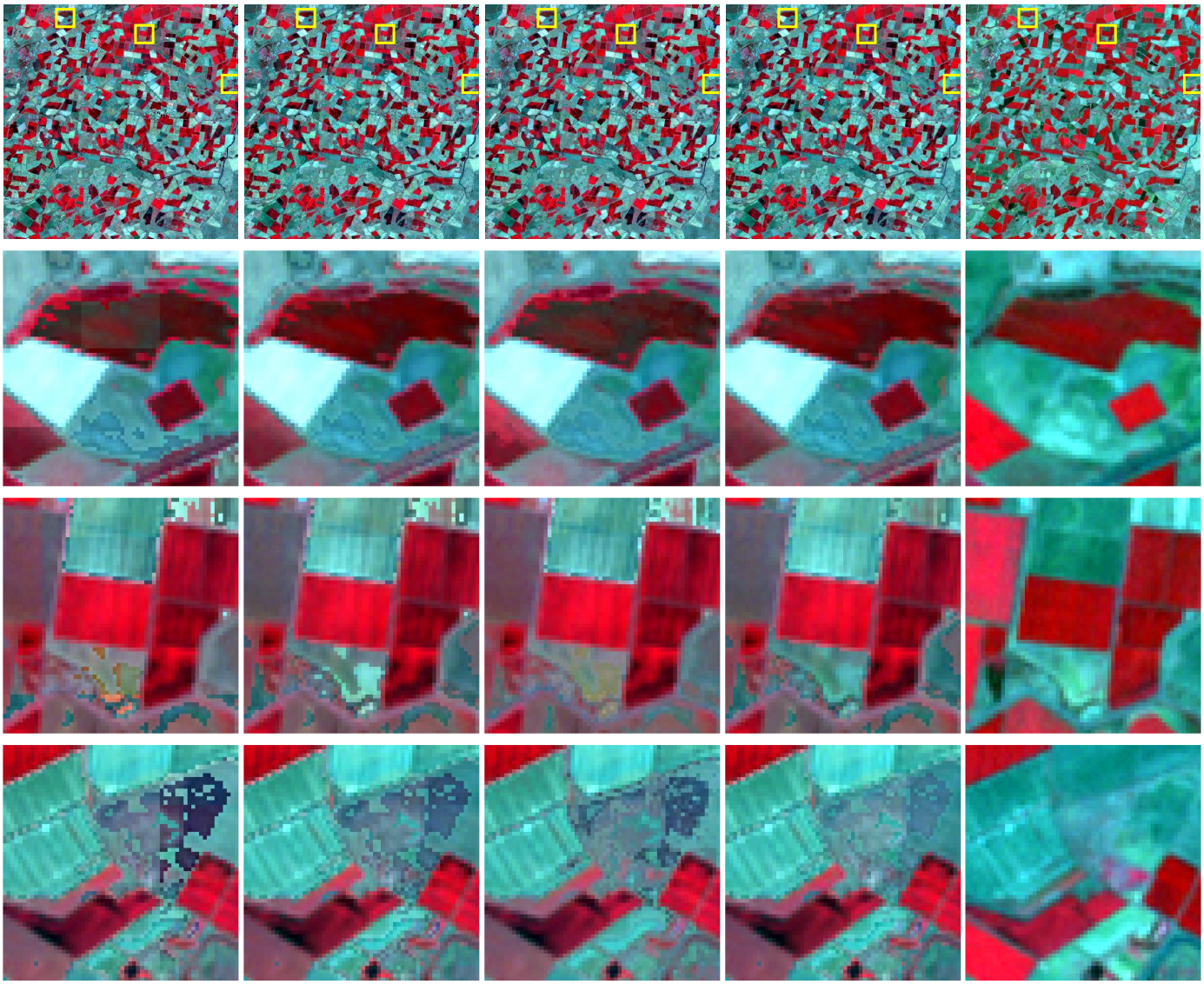
second sub-area which should appear as green, but is inappropriately a light green or black color in the UBDF, UBDF-NM and UBDF-SF predictions. It is also worth noting that the UBDF-based predictions cannot reproduce spatial variance within each object as UBDF assumes that the pixels for the same class in a coarse pixel share the same reflectance and assign the predicted class reflectance directly to the fine pixels. In the predictions of STDFA-based blocks-removed methods in Fig. 6, it is seen that all three methods can remove the blocks satisfactorily and STDFA-BR outperforms STDFA-NM and STDFA-SF. Meanwhile, the spectral distortion of STDFA-NM and STDFA-SF is also more noticeable than STDFA-BR when referring to the reference (e.g., the restoration of the green patch in the bottom of the second sub-area). Compared with the predictions in Figs. 5 and 6, the predictions in Fig. 7 are visually more satisfactory in preserving both the spatial and spectral information, especially for VIPSTF-SU-BR (e.g., the green patch in the third sub-area in Fig. 7(d), which is inappropriately predicted as blue in Fig. 7(a), Fig. 7(b) and Fig. 7(c)). Specifically, the color of the VIPSTF-SU-based predictions in Fig. 7 are closer to the reference than the STDFA-based predictions in Fig. 6, and the VIPSTF-SU-based predictions present more spatial variance and detail than the UBDF-based predictions in Fig. 5. Furthermore, comparison of the predictions in Fig. 7 reveals that VIPSTF-SU-BR is also more accurate than VIPSTF-SU-NM and VIPSTF-SU-SF.

For STDFA and VIPSTF-SU, the spatial unmixing process is essentially performed on the temporal change images. Thus, to further show the effectiveness of the proposed SU-BR method for STDFA and VIPSTF-SU, the temporal change images of STDFA, STDFA-BR, VIPSTF-SU and VIPSTF-SU-BR are shown in Fig. 8, where the results for the red band are provided, with one sub-area zoomed for convenience of visual comparison. It is clear that the blocky artifacts are removed considerably and most of the object boundaries are preserved.



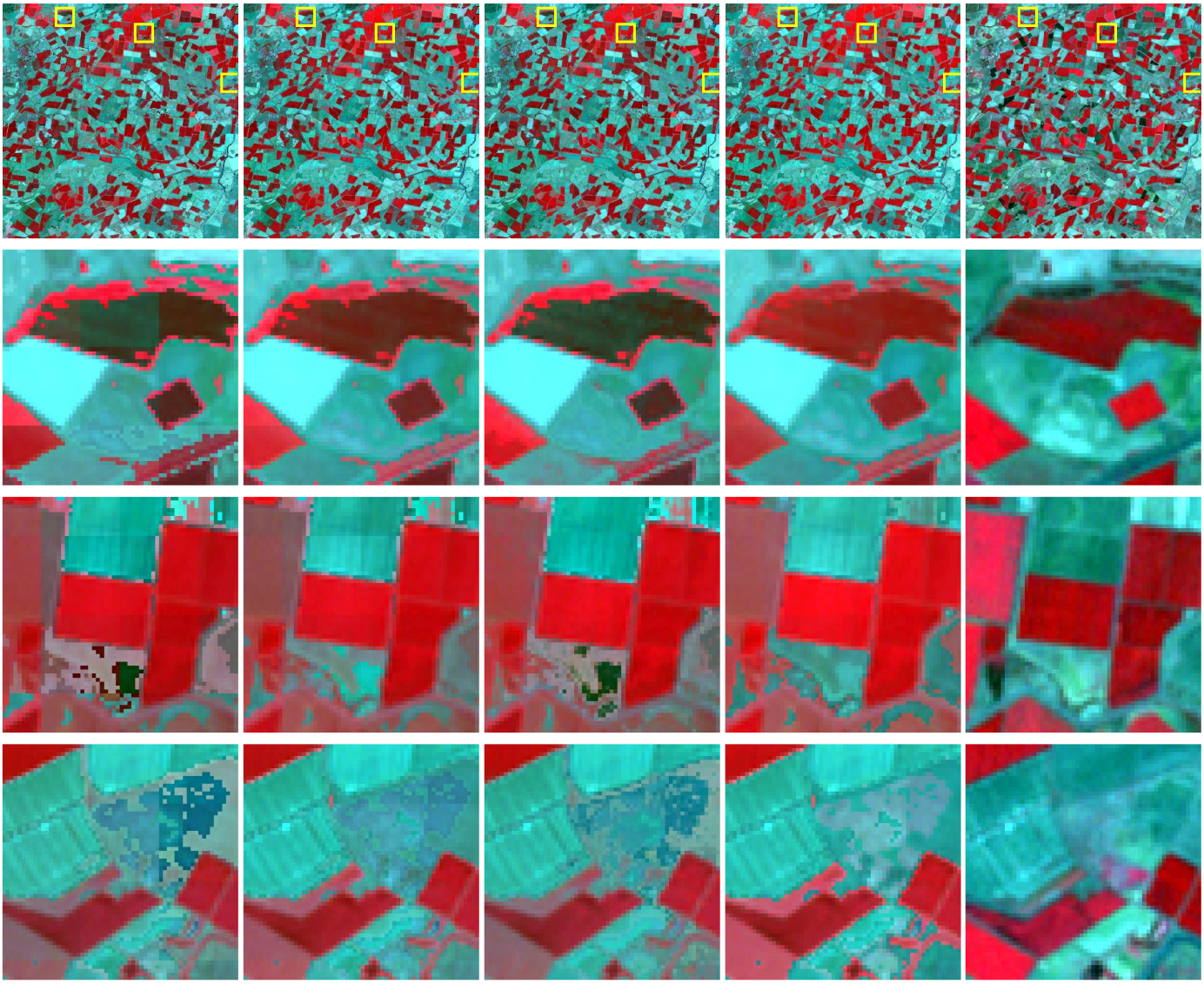
(a) (b) (c) (d) (e)

Fig. 5. Predictions for the heterogeneous region based on UBDF coupled with different blocks-removed methods. (a) UBDF. (b) UBDF-NM. (c) UBDF-SF. (d) UBDF-BR. (e) Reference. The images in the second-to-fourth lines are the corresponding predictions for the three sub-areas marked in yellow in the first line.



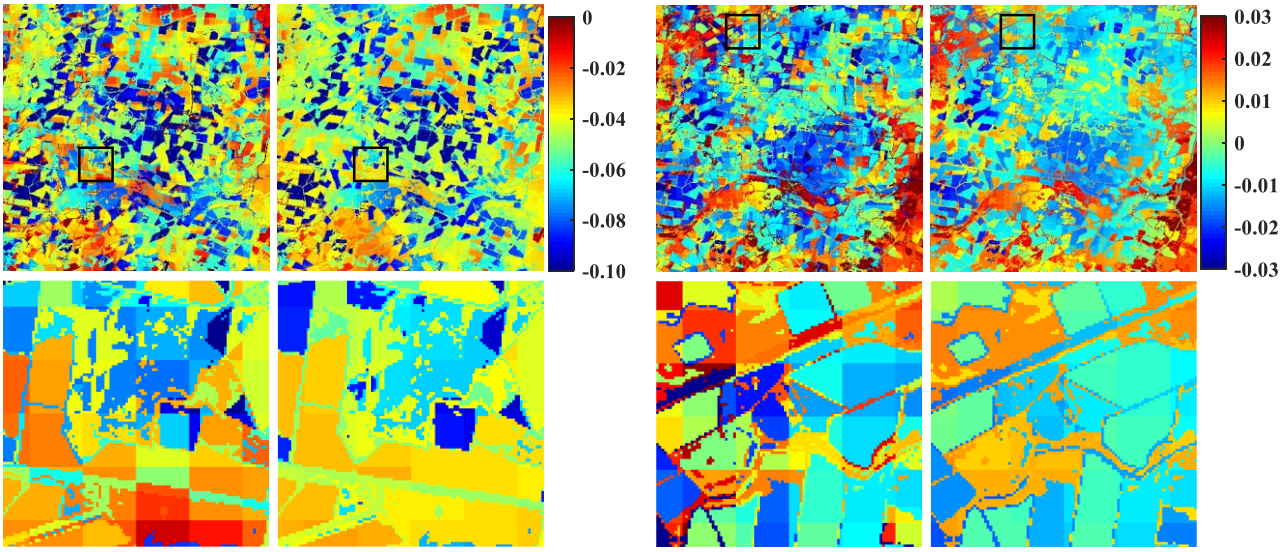
(a) (b) (c) (d) (e)

Fig. 6. Predictions for the heterogeneous region based on STDFA coupled with different blocks-removed methods. (a) STDFA. (b) STDFA-NM. (c) STDFA-SF. (d) STDFA-BR. (e) Reference. The images in the second-to-fourth lines are the corresponding predictions for the three sub-areas marked in yellow in the first line.



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

Fig. 7. Predictions for the heterogeneous region based on VIPSTF-SU coupled with different blocks-removed methods. (a) VIPSTF-SU. (b) VIPSTF-SU-NM. (c) VIPSTF-SU-SF. (d) VIPSTF-SU-BR. (e) Reference. The images in the second-to-fourth lines are the corresponding predictions for the three sub-areas marked in yellow in the first line.



(a) (b)

Fig. 8. Blocks-removed temporal change images for the original spatial unmixing and SU-BR methods. (a) STDFA (left) and STDFA-BR (right) predictions for the red band. (b) VIPSTF-SU (left) and VIPSTF-SU-BR (right) predictions for the red band. The images in the second line are the corresponding predictions for the sub-area marked in black in the first line.

Table 2 Accuracy for the heterogeneous region

		Ideal	Original	SU-NM	SU-SF	SU-BR
CC	UBDF	1	0.7220	0.7675	0.7656	**0.7874**
	STDFA	1	0.8007	0.8151	0.8154	**0.8186**
	VIPSTF-SU	1	0.8181	0.8400	0.8398	**0.8446**
RMSE	UBDF	0	0.0418	0.0399	0.0401	**0.0394**
	STDFA	0	0.0403	0.0380	0.0385	**0.0372**
	VIPSTF-SU	0	0.0343	0.0324	0.0325	**0.0321**
ERGAS	UBDF	0	1.6030	1.5356	1.5384	**1.5133**
	STDFA	0	1.5985	1.5163	1.5416	**1.4868**
	VIPSTF-SU	0	1.2963	1.2291	1.2359	**1.2175**
UIQI	UBDF	1	0.6474	0.6614	**0.6642**	0.6610
	STDFA	1	0.7833	0.7988	0.7974	**0.8026**
	VIPSTF-SU	1	0.8005	0.8125	**0.8144**	0.8120
SAM	UBDF	0	0.2244	0.2157	0.2139	**0.2103**
	STDFA	0	0.1722	0.1590	0.1661	**0.1552**
	VIPSTF-SU	0	0.1615	0.1518	0.1504	**0.1494**

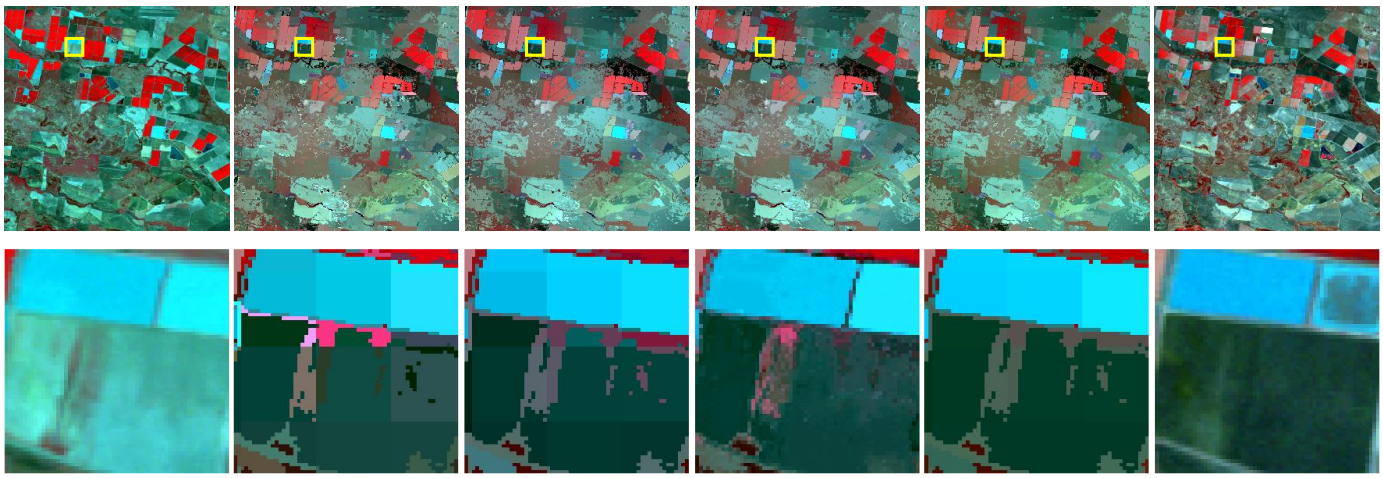
The results of quantitative assessment for the methods are listed in Table 2, where five indices were used, including CC, root mean square error (RMSE), relative global-dimensional synthesis error (ERGAS) (Ranchin and Wald, 2000), universal image quality index (UIQI) (Wang and Bovik, 2002) and spectral angle mapper (SAM). These five indices have been applied widely for quantitative evaluation of image fusion methods

(Amorós-López et al., 2013; Chiman et al., 2018; Wang et al., 2020d). The results in Table 2 support the findings of visual inspection. More precisely, for all the three blocks-removed methods, greater accuracies are produced compared with the original spatial unmixing methods. For example, in comparison with STDFA, the CC values of STDFA-NM and STDFA-SF increase by 0.0144 and 0.0147, respectively. Using the SU-BR method, the gains in CCs are 0.0654, 0.0179 and 0.0265 for UBDF, STDFA, and VIPSTF-SU, respectively. For the other four indices, the gains for SU-BR are also noticeable. Moreover, the SU-BR method can produce fusion results with larger CC and UIQI values, and smaller RMSE, ERGAS and SAM values than the SU-NM and SU-SF methods, indicating SU-BR is more accurate than SU-NM and SU-SF.

3.2.2. Results for the region with land cover changes

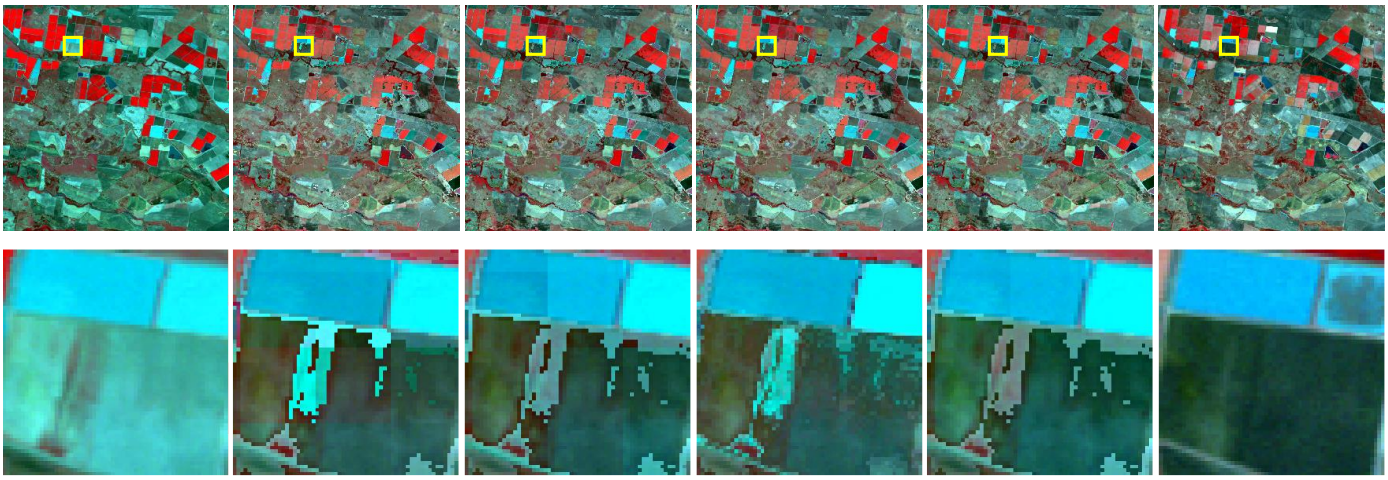
The predictions of the proposed SU-BR method as well as of SU-NM and SU-SF for the region with land cover changes are displayed in Figs. 9, 10 and 11, where a sub-area experiencing noticeable land cover changes is marked in yellow and zoomed for analysis. There exist new artifacts around the boundaries of objects in the SU-SF predictions, presenting noise, especially in the sub-area in Fig. 10(d). With respect to predictions based on SU-NM, the blocky artifacts can still be observed to some extent. Using SU-BR, the blocky artifacts are more satisfactorily removed than with SU-NM, and compared to SU-SF, the SU-BR results contain less noise and are closer to the reference. It should be noted, however, the SU-BR results still present some blocky artifacts, although not very noticeable. This is because our proposed method is implemented based on the assumption of no land cover changes (as for all existing spatial unmixing-based methods), which means if the neighboring pixels do not share the same land cover class with the center pixel at the known time, they also do not participate in constraining the solution of the center pixel at the prediction time, even if some neighbors actually change to share the same class and need to be considered in the constraint. The neglect of these changed pixels can lead to remaining blocky artifacts. Thus, it is challenging to

remove blocky artifacts completely in regions with land cover changes. From the results of quantitative assessment in Table 3, it is clear that all three blocks-removed methods produce greater accuracies than the original spatial unmixing methods, and further, the SU-BR method is more accurate than the other two blocks-removed methods in terms of all five indices, which supports the conclusions from the qualitative assessment. Using SU-BR, the increases in CCs are 0.0704, 0.0481, 0.0589 for the original UBDF, STDFA and VIPSTF-SU methods, respectively. The increases in UIQI and decreases in RMSE, ERGAS and SAM are also substantial.



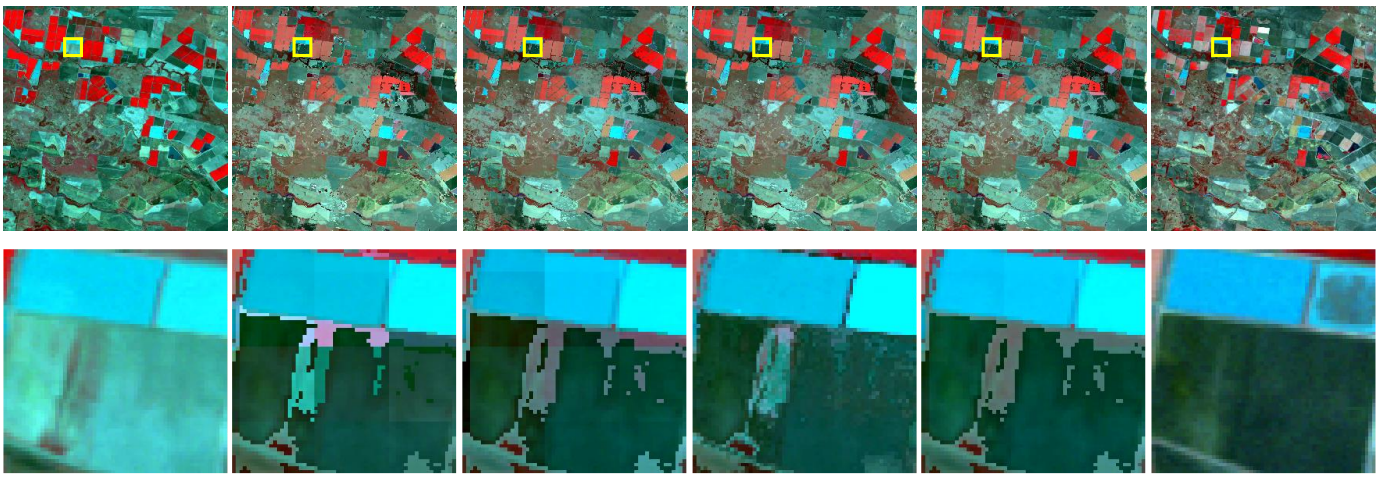
(a) (b) (c) (d) (e) (f)

Fig. 9. Predictions for the region with land cover changes based on UBDF coupled with different blocks-removed methods. (a) Landsat at the known time. (b) UBDF. (c) UBDF-NM. (d) UBDF-SF. (e) UBDF-BR. (f) Reference. The images in the second line are the corresponding predictions for the sub-area marked in yellow in the first line.



(a) (b) (c) (d) (e) (f)

Fig. 10. Predictions for the region with land cover changes based on STDFA coupled with different blocks-removed methods. (a) Landsat at the known time. (b) STDFA. (c) STDFA-NM. (d) STDFA-SF. (e) STDFA-BR. (f) Reference. The images in the second line are the corresponding predictions for the sub-area marked in yellow in the first line.



(a) (b) (c) (d) (e) (f)

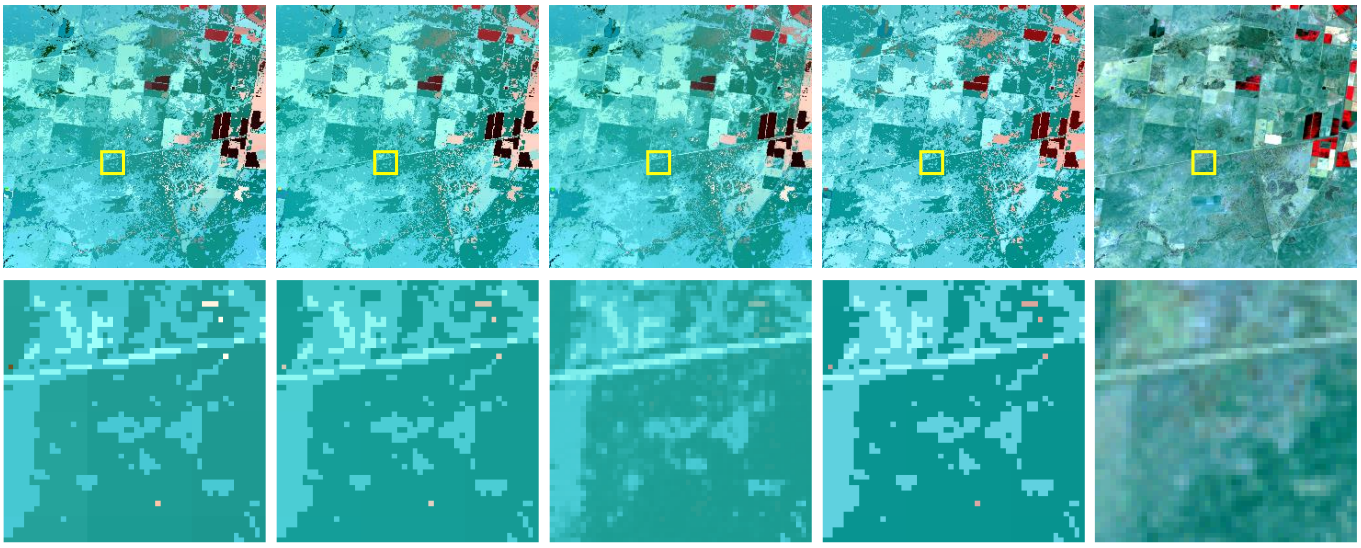
Fig. 11. Predictions for the region with land cover changes based on VIPSTF-SU coupled with different blocks-removed methods. (a) Landsat at the known time. (b) VIPSTF-SU. (c) VIPSTF-SU-NM. (d) VIPSTF-SU-SF. (e) VIPSTF-SU-BR. (f) Reference. The images in the second line are the corresponding predictions for the sub-area marked in yellow in the first line.

Table 3 Accuracy for the region with land cover changes

		Ideal	Original	SU-NM	SU-SF	SU-BR
CC	UBDF	1	0.6306	0.6829	0.6947	**0.7010**
	STDFA	1	0.6937	0.7229	0.7321	**0.7418**
	VIPSTF-SU	1	0.7124	0.7508	0.7534	**0.7713**
RMSE	UBDF	0	0.0372	0.0352	0.0349	**0.0346**
	STDFA	0	0.0358	0.0335	0.0332	**0.0320**
	VIPSTF-SU	0	0.0331	0.0313	0.0312	**0.0305**
ERGAS	UBDF	0	1.2493	1.1895	1.1791	**1.1709**
	STDFA	0	1.1373	1.0606	1.0466	**1.0092**
	VIPSTF-SU	0	1.0906	1.0341	1.0323	**1.0090**
UIQI	UBDF	1	0.6040	0.6332	0.6441	**0.6448**
	STDFA	1	0.6840	0.7158	0.7253	**0.7359**
	VIPSTF-SU	1	0.6991	0.7269	0.7317	**0.7406**
SAM	UBDF	0	0.1365	0.1246	0.1247	**0.1208**
	STDFA	0	0.1551	0.1441	0.1493	**0.1351**
	VIPSTF-SU	0	0.1189	0.1068	0.1104	**0.1023**

3.2.3. Results for the homogeneous region

Fig. 12 shows the predictions based on UBDF for the homogeneous region. The results of one sub-area are zoomed to facilitate visual comparison. It should be noted that the block effect for this region is not as obvious as that for the previous two regions. This is because the intra-class spectral variation for the homogeneous region is not great, leading to smaller differences in class reflectance between adjacent pixels. Checking the results, the ability to remove blocky artifacts of our proposed method is also demonstrated. The result predicted by UBDF-SF presents ambiguous artifacts, which may be helpful for reproducing more spatial variation. The results of quantitative assessment for all three blocks-removed methods are listed in Table 4. It is seen that all three SU-BR methods outperform the original spatial unmixing methods. Furthermore, SU-BR has a comparable performance with SU-SF and both produce greater accuracy than SU-NM.



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

Fig. 12. Predictions for the homogeneous region based on UBDF coupled with different blocks-removed methods. (a) UBDF. (b) UBDF-NM. (c) UBDF-SF. (d) UBDF-BR. (e) Reference. The images in the second line are the corresponding predictions for the sub-area marked in yellow in the first line.

Table 4 Accuracy for the homogeneous region

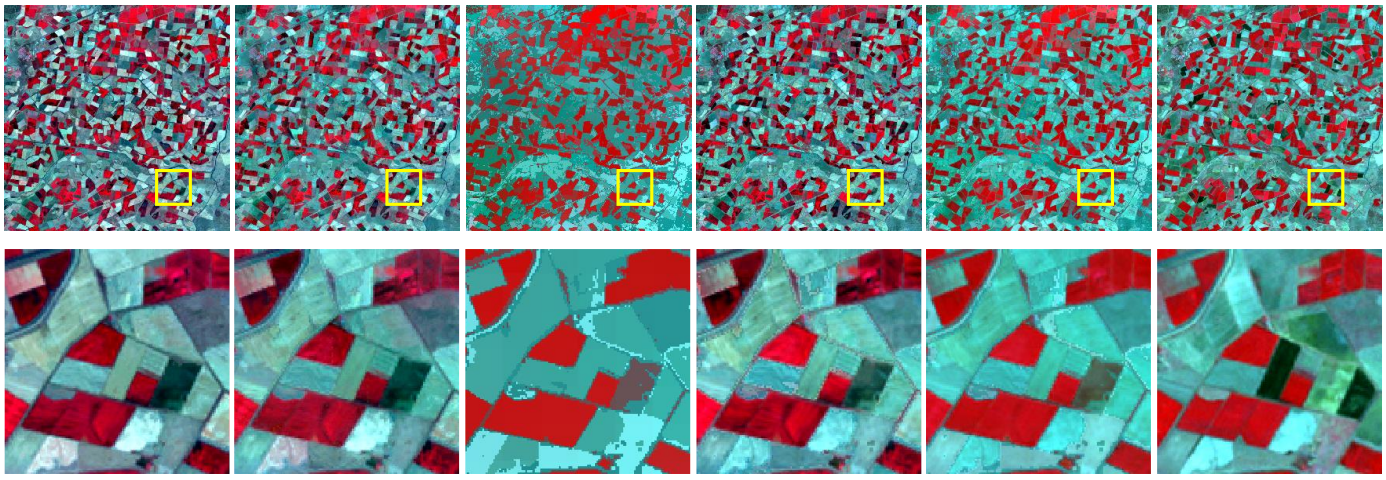
		Ideal	Original	SU-NM	SU-SF	SU-BR
CC	UBDF	1	0.7383	0.7565	**0.7838**	0.7684
	STDFA	1	0.8888	0.8947	0.8965	**0.8971**
	VIPSTF-SU	1	0.8850	0.8911	**0.8954**	0.8930
RMSE	UBDF	0	0.0287	0.0274	0.0268	**0.0266**
	STDFA	0	0.0176	0.0171	0.0170	**0.0168**
	VIPSTF-SU	0	0.0177	0.0171	**0.0169**	**0.0169**
ERGAS	UBDF	0	0.6662	0.6417	**0.6240**	0.6293
	STDFA	0	0.4227	0.4101	0.4071	**0.4053**
	VIPSTF-SU	0	0.4250	0.4116	**0.4049**	0.4083
UIQI	UBDF	1	0.6822	0.6969	**0.7100**	0.6943
	STDFA	1	0.8859	0.8907	**0.8930**	0.8913
	VIPSTF-SU	1	0.8767	0.8811	**0.8855**	0.8804
SAM	UBDF	0	0.0987	0.0967	0.0950	**0.0944**
	STDFA	0	0.0518	0.0507	0.0516	**0.0499**
	VIPSTF-SU	0	0.0541	0.0531	**0.0524**	0.0528

3.3. Comparison with other methods

As demonstrated in Section 3.2, the SU-BR method increases the accuracy of the original spatial unmixing-based methods by reducing the blocky artifacts effectively. Also, it presents greater prediction

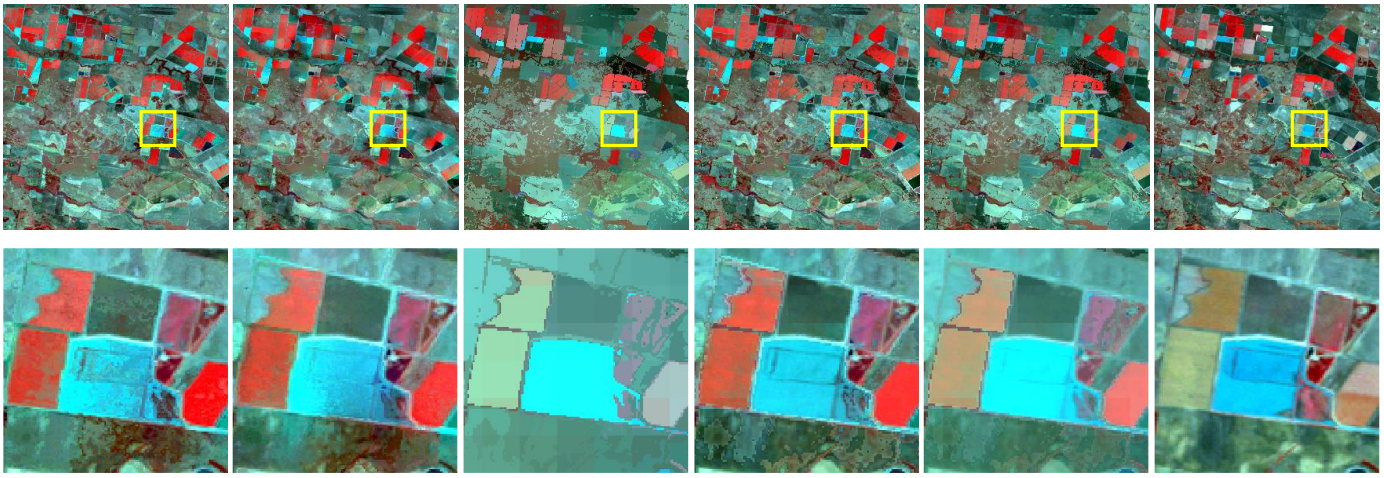
accuracy than other simple blocks-removed algorithms (i.e., SU-SF and SU-NM). Admittedly, the original spatial unmixing-based methods are sometimes inferior to spatial weighting-based methods (e.g., STARFM) and hybrids methods (e.g., FSDAF) due to the effect of blocky artifacts. Since the SU-BR method can remove the blocks effectively, a comparison is warranted between SU-BR and other types of methods. In this paper, two popular methods, including STARFM of the spatial weighting-based method and FSDAF of the hybrid methods were used for comparison. Note that we did not consider ESTARFM as it requires two MODIS-Landsat image pairs for implementation, but the spatial unmixing methods investigated in this paper are all based on a single image pair. For fair comparison, we therefore considered the methods that can also be performed using a single image pair (i.e., STARFM and FSDAF). The SU-BR predictions based on all three choices (UBDF, STDFA and VIPSTF-SU) are included in the comparison. The predictions for the heterogeneous region, the region with land cover changes and the homogeneous region are shown in Figs. 13, 14 and 15, respectively.

As shown in Fig. 13, the predictions of UBDF-BR, STDFA-BR and VIPSTF-SU-BR are visually more similar to the reference (see, for example, the bright red vegetation in these methods). Furthermore, VIPSTF-SU-BR predicts the reflectance of the patches most accurately. On the contrary, the hue as a whole is darker in STARFM and FSDAF compared to the reference. With respect to the region with land cover changes shown in Fig. 14, the prediction is visually less accurate than that for the heterogeneous region due to the great temporal changes between the images at the known and prediction times. Although STARFM and FSDAF seem to predict well the dark blue patch, there exist unexpected red patches when focusing on the left part of the sub-area. The prediction of VIPSTF-SU-BR is more similar to the reference image as a whole, which can be validated by the restoration of the brown patches in the sub-area. Focusing on the predictions for the homogeneous region in Fig. 15, it is obvious that the predictions of STDFA-BR, VIPSTF-SU-BR and FSDAF are closer to the reference image. Moreover, the curved line object in the middle of the sub-area predicted by VIPSTF-SU-BR is the closest to the reference.



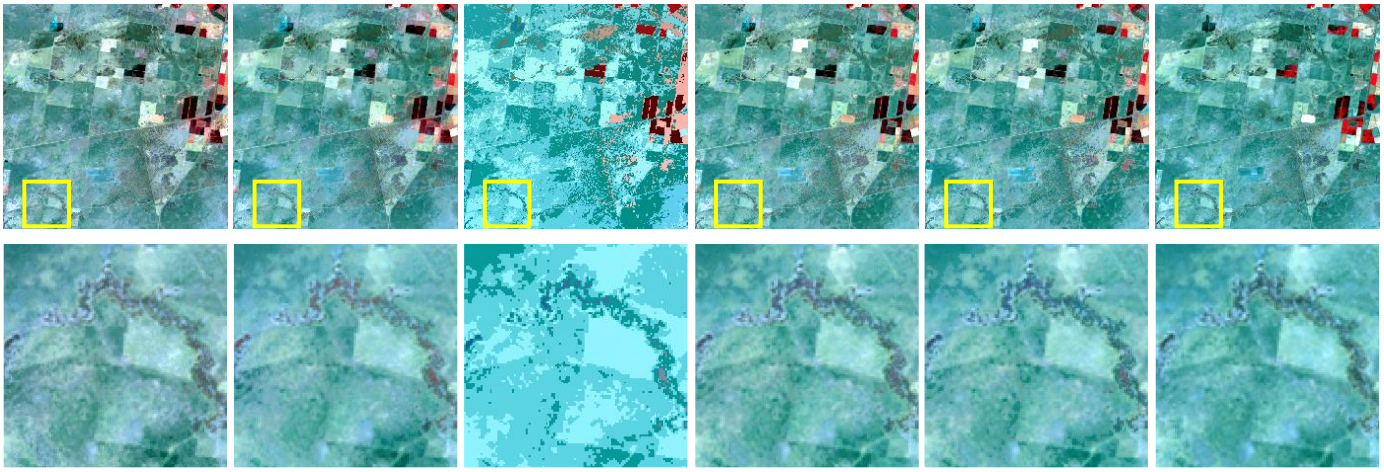
(a) (b) (c) (d) (e) (f)

Fig. 13. Predictions for the heterogeneous region using different methods. (a) STARFM. (b) FSDAF. (c) UBDF-BR. (d) STDFA-BR. (e) VIPSTF-SU-BR. (f) Reference. The images in the second line are the corresponding predictions for the sub-area marked in yellow in the first line.



(a) (b) (c) (d) (e) (f)

Fig. 14. Predictions for the region with land cover changes using different methods. (a) STARFM. (b) FSDAF. (c) UBDF-BR. (d) STDFA-BR. (e) VIPSTF-SU-BR. (f) Reference. The images in the second line are the corresponding predictions for the sub-area marked in yellow in the first line.



(a) (b) (c) (d) (e) (f)

Fig. 15. Predictions for the homogeneous region using different methods. (a) STARFM. (b) FSDAF. (c) UBDF-BR. (d) STDFA-BR. (e) VIPSTF-SU-BR. (f) Reference. The images in the second line are the corresponding predictions for the sub-area marked in yellow in the first line.

Table 5 Accuracy of different methods for the three regions

		Ideal	Heterogeneous region	Region with land cover changes	Homogeneous region
CC	STARFM	1	0.8043	0.7643	0.8897
	FSDAF	1	0.8314	0.7705	0.8946
	UBDF-BR	1	0.7874	0.7010	0.7684
	STDFA-BR	1	0.8186	0.7418	**0.8971**
	VIPSTF-SU-BR	1	**0.8446**	**0.7713**	0.8930
RMSE	STARFM	0	0.0411	0.0323	0.0180
	FSDAF	0	0.0357	**0.0297**	0.0171
	UBDF-BR	0	0.0394	0.0346	0.0266
	STDFA-BR	0	0.0372	0.0320	**0.0168**
	VIPSTF-SU-BR	0	**0.0321**	0.0305	0.0169
ERGAS	STARFM	0	1.6696	1.0154	0.4228
	FSDAF	0	1.4137	**0.9366**	0.4112
	UBDF-BR	0	1.5133	1.1709	0.6293
	STDFA-BR	0	1.4868	1.0092	**0.4053**
	VIPSTF-SU-BR	0	**1.2175**	1.0090	0.4083
UIQI	STARFM	1	0.7753	0.7544	0.8876
	FSDAF	1	**0.8169**	**0.7653**	0.8881
	UBDF-BR	1	0.6610	0.6448	0.6943
	STDFA-BR	1	0.8026	0.7359	**0.8913**
	VIPSTF-SU-BR	1	0.8120	0.7406	0.8804
SAM	STARFM	0	0.1758	0.1494	0.0676
	FSDAF	0	0.1552	0.1244	0.0573
	UBDF-BR	0	0.2103	0.1208	0.0944
	STDFA-BR	0	0.1552	0.1351	**0.0499**
	VIPSTF-SU-BR	0	**0.1494**	**0.1023**	0.0528

Table 5 shows the results of quantitative assessment of the different methods for the three regions together. Overall, the predictions for the homogeneous region have the greatest accuracy while the predictions for the region with land cover changes are the least accurate. Checking the results for the heterogeneous region, VIPSTF-SU-BR produces the greatest CC and the smallest RMSE, ERGAS and SAM. More precisely, the CC for VIPSTF-SU-BR is 0.8446, with an increase of 0.0403 and 0.0132 compared to STARFM and FSDAF. Also, the CC of VIPSTF-SU-BR is 0.0572 and 0.0260 larger than for UBDF-BR and STDFA-BR. VIPSTF-SU-BR produces the smallest ERGAS of 1.2175, which is 0.4521 and 0.1962 smaller than for STARFM and FSDAF. For the region with land cover changes, VIPSTF-SU-BR also produces the greatest CC of 0.7713 and the smallest SAM of 0.1023, which is 0.0471 and 0.0221 smaller than for STARFM and FSDAF. For the homogeneous region, STDFA-BR has the greatest prediction accuracy and VIPSTF-SU-BR has very close accuracy to STDFA-BR. The RMSE of STDFA-BR is 0.0168, which is 0.0012 and 0.0003 smaller than for STARFM and FSDAF.

3.4. Analysis of parameters

3.4.1. The magnitude regularization parameter A

The aim of using A as a coefficient of the spatial continuity constraint term D is to match its magnitude with that of the data fidelity term R. As mentioned earlier in Section 2.4, we utilized statistical information of D and R in predictions of the original spatial unmixing methods to estimate the magnitude regularization parameter A. The histograms of D and R in the original STDFA predictions for the three regions are shown as examples for illustration in Fig. 16. From the histograms, the values of the magnitude regularization parameter A were determined as 100, 10 and 1000 for the heterogeneous region, the region with land cover changes and the homogeneous region, respectively. The value of A for the homogeneous region is the largest, as the original

predictions (e.g., STDFA predictions illustrated here) of reflectances of the same class in neighboring pixels are most similar (i.e., the term of *D* is very small). Thus, *A* tends to be larger to match the magnitude of *D* with *R*. Note that the smallest value of *D* also suggests that the block effect is the weakest for the homogeneous region.

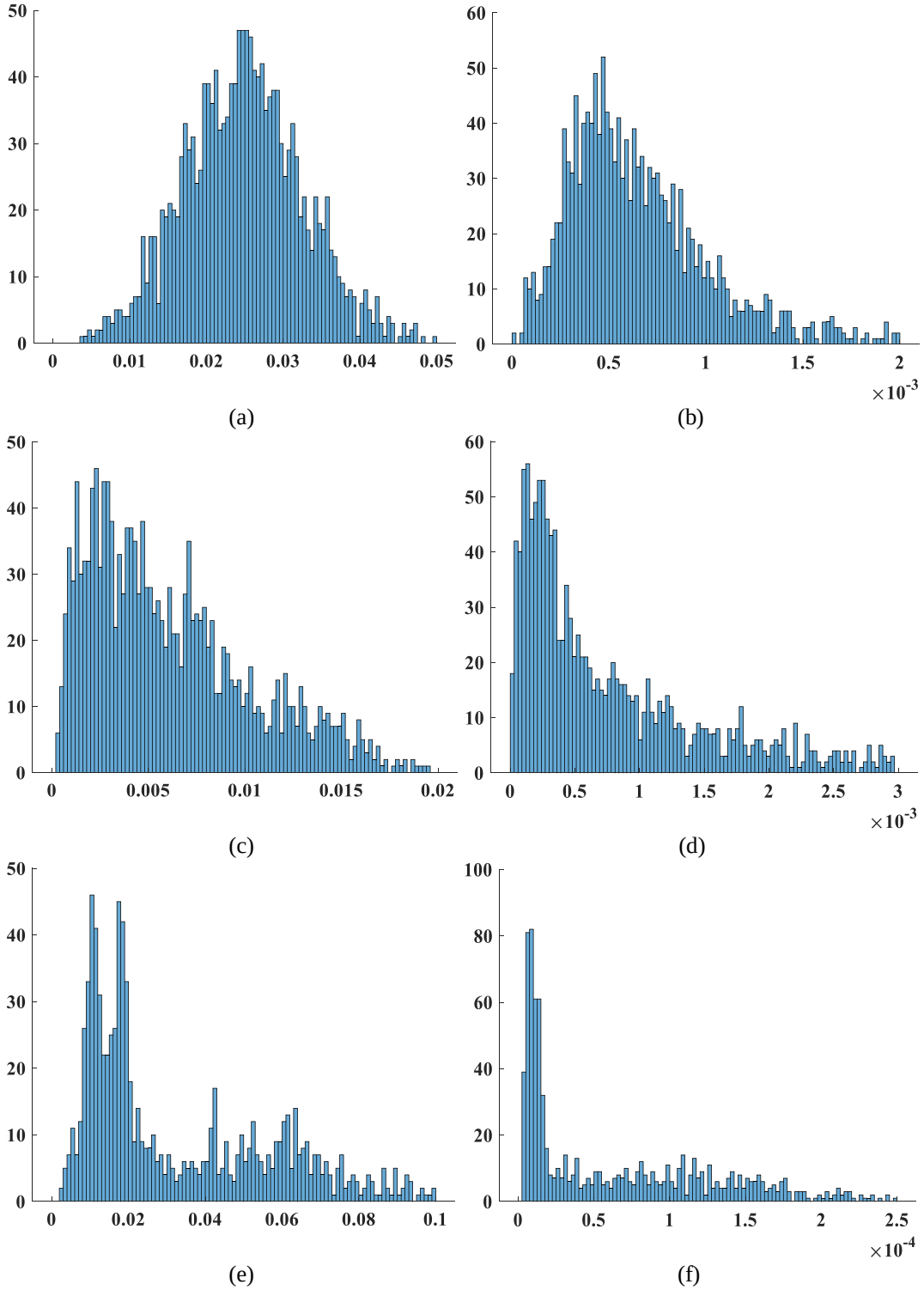


Fig. 16. Histograms of the data fidelity term *R* (left) and the spatial continuity constraint term *D* (right) produced based on the predictions of the original STDFA method. (a) and (b) heterogeneous region. (c) and (d) region with land cover changes. (e) and (f) homogeneous region.

3.4.2. The balancing parameter α

The balancing parameter α is used to control the contributions of the spatial continuity constraint term and the data fidelity term. The CCs of STDFA-BR and VIPSTF-SU-BR in relation to different balancing parameters are shown in Fig. 17, where the accuracies of the corresponding SU-SF and SU-NM versions are also provided for comparison. It is clear that for all three regions, SU-BR is more accurate than SU-SF and SU-NM when α takes a value between 0.2 and 0.8. For the heterogeneous region and region with land cover changes, the CCs are maximum when the balancing parameter is around 0.5, suggesting that the influences of the two terms are comparable after the magnitude adjustment by the magnitude regularization parameter *A*. Thus, for these two types of regions, the median is suggested as a preferable choice for α, as was done in the experiments in Sections 3.2 and 3.3. With respect to the homogeneous region, with an increase in α, the CC decreases very slightly (by only around 0.002 when α increases from 0.1 to 0.9). This is attributed mainly to the weak block effect for the homogeneous region. Therefore, since the magnitude of the data has been adjusted by the parameter *A*, the selection of α generally will not exert much influence on the prediction accuracy and the median could be a preferable choice in most cases.

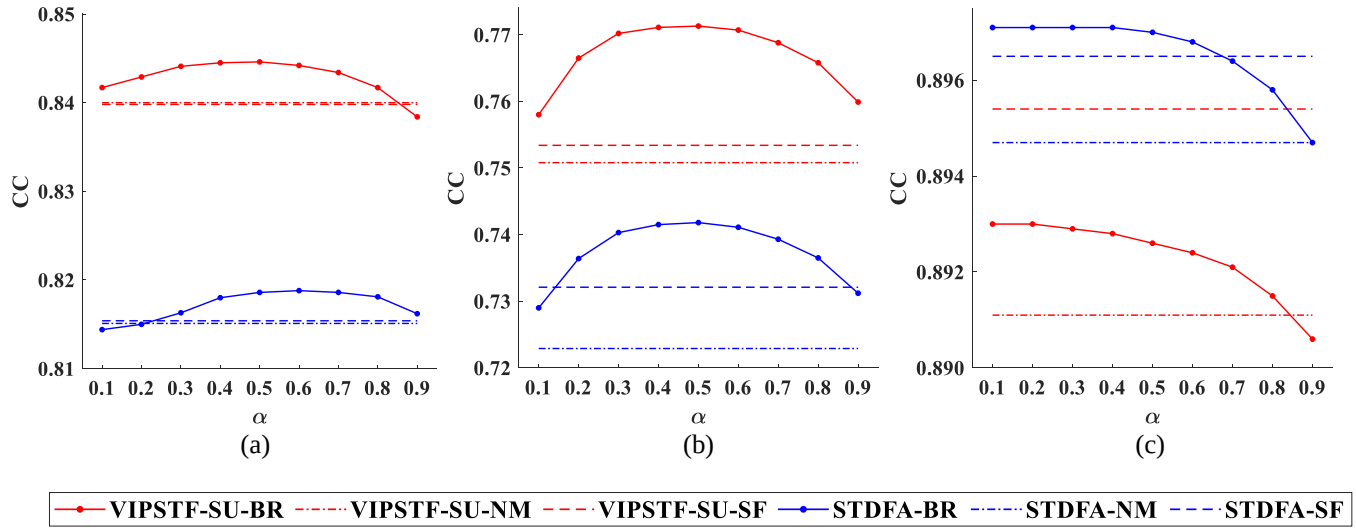


Fig. 17. The impact of α on the accuracy of STDFA-BR and VIPSTF-SU-BR. (a) Heterogeneous region. (b) Region with land cover changes. (c) Homogeneous region. The dotted line and dashed line represent the accuracies of SU-NM and SU-SF, respectively.

4. Discussion

4.1. Comparison between SU-NM, SU-SF and SU-BR

In the proposed SU-BR method, two terms are considered: the residual error in the unmixing model and the spatial continuity of class reflectance. The residual error term represents the data fidelity, which measures the ability to preserve the original coarse spatial resolution image at the prediction time. Meanwhile, the spatial continuity constraint is the key to removing blocks. For conventional spatial unmixing-based methods (i.e., UBDF, STDFA and VIPSTF-SU), the class reflectances are predicted by simply minimizing the residual error to ensure the greatest data fidelity. Due to the differences in sensors and acquisition conditions, however, a bias always exists in the coarse reflectance compared to the fine spatial resolution data (i.e., when the fine spatial resolution data are upscaled to the coarse spatial resolution, they are different from the observed coarse data) (Chen et al., 2020; Li et al., 2020b; Xie et al., 2018). It is obvious in Fig. 5 that the reflectance predicted

by UBDF (Fig. 5(a)) varies greatly from that of the reference (Fig. 5(e)). Thus, to consider merely the residual error may not result in an accurate prediction. To investigate the relation between the residual error and the fusion accuracy, the results of different blocks-removed methods based on UBDF for the heterogeneous region are listed in Table 6. Note that the residual error here is the average of errors of all coarse pixels in all bands.

Table 6 The prediction accuracy (in terms of CC) and the residual error of the spatial unmixing methods for the heterogeneous region

	UBDF	UBDF-NM	UBDF-BR
Prediction accuracy	0.7220	0.7675	0.7874
Residual error	0.0196	0.0211	0.0229

It can be noticed that UBDF has the smallest residual error of 0.0196, but produces the smallest CC of 0.7220. On the contrary, although the residual error of UBDF-BR is the largest, it provides the greatest prediction accuracy. The residual error of UBDF-NM is smaller than that for UBDF-BR, but its performance in spatio-temporal fusion is inferior to UBDF-BR. This phenomenon is related to the mechanisms of the two types of blocks-removed methods. The original UBDF, STDFA and VIPSTF-SU methods simply consider minimizing the residual error as the objective, so that they are most likely to be influenced by the bias originating from the observed coarse data. As for the two blocks-removed methods used as benchmarks in this paper, SU-SF and SU-NM, they both apply a simple post-processing to the results of the original methods. The separate post-processing means that SU-SF and SU-NM are heavily dependent on the previous estimation. As a result, although the blocky artifacts can be removed by adapting these two methods, their ability to correct the reflectance misestimated by the original methods is limited. For the proposed SU-BR method, however, simultaneously the blocky artifacts are removed obviously and the prediction is closer to the reference (see Fig. 5(d)). The reason is that SU-BR performs unmixing by considering jointly the objective of minimizing the residual error and the constraint of the spatial continuity of land cover, and a balance is found between these two aspects through the iterative process. The constraint of spatial continuity allows the predicted reflectance to approach that of neighboring pixels gradually, producing greater possibility to reduce the influence of the

data fidelity where bias in the original observed coarse data can adversely affect the final prediction. Therefore, the prediction of SU-BR varies noticeably compared to the original method, SU-SF and SU-NM, and is closer to the reference.

In fact, SU-BR sacrifices data fidelity to a certain extent for a more accurate prediction. If the difference between the coarse and fine data is very small and the observed coarse data are sufficiently reliable (i.e., the data fidelity is sufficient reliable), the sacrifice of data fidelity may not necessarily lead to an increase in accuracy. In this case, post-processing such as the residual compensation strategy in Fit-FC (Wang and Atkinson, 2018) may be considered. In future research, it would be of great interest to consider a pre-processing step to reduce the difference between the coarse and fine data for more reliable spatio-temporal fusion.

4.2. Comparison between UBDF, STDFA and VIPSTF-SU and their BR versions

Three spatial unmixing-based spatio-temporal fusion methods, UBDF, STDFA and VIPSTF-SU, are considered in the application of SU-BR. For UBDF, only the classification map produced from the fine spatial resolution multispectral images at the known time is used instead of the original multispectral image. The classification map preserves the thematic class information, but ignores the intra-class spectral variation. Thus, UBDF fails to recover the intra-class spectral variation which is important to characterize the texture information in the fine spatial resolution image. STDFA and VIPSTF-SU are performed based on image pairs, which utilize one more input image (i.e., the coarse image at the known time) than UBDF. STDFA calculates the fine spatial resolution class reflectance change based on the changes between the coarse spatial resolution images at known and prediction times. VIPSTF-SU extends STDFA based on the virtual image pair constructed from the original image pair, which is closer to the images at the prediction time (Wang et al., 2020c). The application of virtual image pair decreases the uncertainty in unmixing and recovers the fine

683 spatial resolution information more accurately. Among the three methods, VIPSTF-SU has the greatest
684 prediction accuracy. Inheriting this advantage, the corresponding SU-BR version of VIPSTF-SU (i.e.,
685 VIPSTF-SU-BR) is more accurate than the other two versions (UBDF-BR and STDFA-BR), as seen in the
686 experiments.
687
688 *4.3. The performance of SU-BR in different regions*
689
690 In this paper, the blocks-removed method was performed for three regions, including the heterogenous
691 region, the region with land cover changes and the homogeneous region. It can be noted that the performance
692 of SU-BR varies for different regions. Generally, SU-BR presents greater advantages in removing blocks and
693 recovering the reflectances in the heterogeneous region and the region with land cover changes. For the
694 homogeneous region, the prediction of the original methods does not present obvious blocky artifacts because
695 of the large similarity between neighboring pixels and the very small land cover change. Thus, the effect of
696 SU-BR is not obviously observed. For the other two regions, there exists great variation between neighboring
697 pixels, resulting in severe blocky artifacts, where there is a great need for SU-BR, as seen in the experiments.
698
699 *4.4. The applicability of SU-BR*
700
701 For the SU-BR method proposed in this paper, two aspects can be considered in regard to its applicability.
702 On the one hand, as validated in the experiments, SU-BR is applicable to different spatial unmixing methods,
703 including UBDF, STDFA and VIPSTF-SU. Therefore, SU-BR has the potential to solve effectively the
704 common problem of blocky artifacts in almost all spatial unmixing-based methods. On the other hand, SU-BR
705 provides a general framework for enhancing spatial unmixing-based methods, which can be summarized as
706 $$J = G + C \tag{8}$$

where *G* represents the goal (i.e., minimizing residual error) and *C* represents the constraint. The proposed SU-BR is fully compliant with this general framework, where the constraint *C* denotes the differences in reflectances of the same land cover class in spatially adjacent pixels. In some existing works, class reflectance of pure coarse pixels (e.g., MODIS pixels) (Xu et al., 2015) and prediction of some other spatio-temporal fusion methods such as STARFM (Gao et al., 2006) are used as constraints. The proposed SU-BR provides a flexible constraint that is compatible with any existing constraints. For example, the constraint provided by the pure coarse pixels can be added to the term *C* in Eq. (8) for possible enhancement, if such pure pixels exist widely in the observed coarse image at the prediction time. In future research, more potential constraints can be included in SU-BR to further enhance the performance of removing blocks and further, increase the accuracy of spatio-temporal fusion. The common choice for blending the multiple constraints would be linear combination. It would be a critical issue to determine reasonably the contributions of each constraint term.

4.5. The computational cost of SU-BR

As SU-BR requires a number of iterations, it is more time-consuming compared to the original spatial unmixing-based methods. Table 7 shows the computational cost of the three SU-BR versions and the corresponding original methods for the three regions. All experiments were carried out using MATLAB (R2019a) based on a laptop with an Intel(R) Core(TM) i7-8750H CPU at 2.20 GHz. The Landsat ETM+ images used in the heterogeneous region and the region with land cover changes have the same spatial size of 800 by 800 pixels, while the homogeneous region covers an area of 600 by 600 Landsat pixels. For the heterogeneous region and the region with land cover changes, by adopting SU-BR, the computing time increases from around 2 minutes to more than 44 minutes. As for the homogeneous region, the computational cost also increases significantly from less than 1 minute to more than 12 minutes.

It can be noted that the terminal condition of SU-BR involves two cases: 1) the pre-defined maximum number of iterations is achieved; 2) the difference between three consecutive realizations is smaller than the pre-defined threshold. Since the spatial unmixing is applied to each coarse pixel in each iteration, the computational cost is expensive when either the number of iterations or coarse pixels is large. Actually, the solution of some pixels remains stable after several iterations, especially for pixels located at the center of a large object or even in a homogeneous area. To reduce the redundant operation on these pixels, a pixel level terminal condition can be defined potentially. For example, when the change of predicted reflectance of a pixel reaches a threshold, this pixel will be marked and the result in this iteration will be recorded. Meanwhile, this pixel will not be updated in the next iterations, while its neighbors can be updated conditionally upon its static value. By adopting this strategy, the computational cost may be saved dramatically, especially for the homogeneous region where the block effect is relatively weak.

Table 7 The computational cost (in units of seconds)

Spatial size	Heterogeneous region		Region with land cover change		Homogeneous region	
	800×800 Landsat pixels		800×800 Landsat pixels		600×600 Landsat pixels	
	Original	SU-BR	Original	SU-BR	Original	SU-BR
UBDF	68.0	3117.0	245.8	7638.0	58.3	1764.0
STDFA	72.9	2678.1	120.6	3829.4	44.3	1127.0
VIPSTF-SU	62.6	2640.4	98.2	3975.3	50.1	767.7

4.6. The limitation of SU-BR

This paper aims at removing blocks in spatial unmixing-based spatio-temporal fusion methods. However, it can be seen from the visual presentation of the SU-BR prediction that the blocks still exist to a limited extent. Considering the mechanism of SU-BR, two main reasons may result in the incomplete removal of the blocky artifacts. First, it should be stressed that the implementation of SU-BR is based on the assumption that no land cover change occurs between images at the known and prediction times whereas, in fact, land cover change is

751 inevitable. In SU-BR, if neighboring pixels belonging to different land cover classes with the center pixel at
752 the known time change to share the same land cover class at the prediction time, they will still be assumed to
753 belong to different classes and allocated with different reflectances in spatial unmixing. As a result, the blocky
754 artifacts will remain because these changed pixels are ignored. Second, the intra-class spectral variation can
755 also be an obstacle for complete elimination of blocky artifacts. As analyzed explicitly in Section 2.2, the
756 blocky artifacts reflect the intra-class spectral variation in fusion predictions at the coarse spatial resolution. It
757 means the block effect will exist as long as there is intra-class spectral variation for the observed data. No
758 matter how many iterations are taken in the SU-BR model, the difference between the estimated reflectance
759 for pixels of the same class remains, presenting the blocky artifacts. Except for the method to remove blocks
760 based on the spatial continuity of class reflectance, other post-processing strategies may be considered to
761 further eliminate the blocks. The application of these strategies may potentially enhance the performance in
762 removing the blocky artifacts. Nevertheless, it should be emphasized that the ultimate purpose of removing
763 blocks is to increase the accuracy of spatio-temporal fusion. It is still unclear whether the further removal of
764 blocks will necessarily benefit the prediction or increase the prediction accuracy.
765
766
767 **5. Conclusion**
768
769 The block effect is a long-standing issue in spatial unmixing-based spatio-temporal fusion, which influences
770 the prediction accuracy greatly. This paper proposed a SU-BR method to cope with the problem of blocky
771 artifacts in spatial unmixing predictions. Based on the assumption of spatial continuity, SU-BR removes the
772 blocky artifacts by minimizing the difference in reflectances of the same land cover class in spatially adjacent
773 pixels. SU-BR was applied to three typical spatial unmixing-based methods (i.e., UBDF, STDFA and
774 VIPSTF-SU), and was examined using datasets covering three different landscapes (one heterogeneous region,

775 one region experiencing land cover changes and one homogeneous region) in the experiments. The main
776 findings of this paper are summarized as follows.

777 1) SU-BR can remove the blocky artifacts effectively in spatial unmixing-based spatio-temporal fusion.
778 The blocky artifacts in the original UBDF, STDFA and VIPSTF-SU predictions are removed obviously
779 by applying SU-BR.
780 2) SU-BR can increase the prediction accuracy of spatio-temporal fusion. For the heterogeneous region,
781 the CCs of UBDF-BR, STDFA-SU-BR and VIPSTF-SU-BR are 0.0654, 0.0179 and 0.0265 larger than
782 the original methods.
783 3) SU-BR is more accurate than the other two potential benchmark methods for removing blocks, (i.e.,
784 SU-NM and SU-SF). For the region with land cover changes, the UIQI of STDFA-BR is 0.0201 and
785 0.0106 larger than STDFA-NM, STDFA-SF, respectively.
786 4) SU-BR also outperforms two state-of-the-art methods, that is, STARFM and FSDAF. STARFM and
787 FSDAF produce CCs of 0.8043 and 0.8314 in the heterogeneous region, while VIPSTF-SU-BR
788 produces a larger CC of 0.8446.
789 5) VIPSTF-SU-BR is a preferable choice in all three SU-BR versions. For the heterogeneous region, the
790 CC of VIPSTF-SU-BR is 0.0572 and 0.0260 larger than that of UBDF-BR and STDFA-BR. The UIQI
791 of VIPSTF-SU-BR is 0.7406 in the region with land cover change, which is 0.0958 and 0.0047 larger
792 than for UBDF-BR and STDFA-BR.
793 6) SU-BR is applicable to various regions dominated by different landscapes, and is more advantageous in
794 removing blocks for the heterogeneous region and the region experiencing land cover changes.
795
796
797 **Acknowledgment**
798

This work was supported by the National Natural Science Foundation of China under Grant 41971297, Fundamental Research Funds for the Central Universities under Grant 02502150021, and Tongji University under Grant 02502350047.

Appendix A

1) UBDF

For UBDF, **E** in Eqs. (1) and (2) denotes the sub-pixel level reflectances of all C classes. The reflectances of all N coarse pixels in the local window are arranged in an $N \times 1$ vector **Q**. The predicted fine spatial resolution reflectance of each class in **E** is assigned directly to the fine spatial resolution pixels in the center coarse pixel according to their class labels in the known fine resolution image.

2) STDFA

STDFA is performed on the changes in the coarse spatial resolution images between the known and prediction times, on the condition that the two coarse images can be observed. Accordingly, **E** represents the temporal change of the reflectances of land cover classes at the target fine spatial resolution and **Q** represents the temporal change of the reflectances of the coarse pixels in the local window. The predicted change of reflectance for each fine spatial resolution pixel is added to the known fine spatial resolution image to produce the final prediction. Compared to UBDF, STDFA makes fuller use of the fine spatial resolution image.

3) VIPSTF-SU

The VIPSTF approach proposed by Wang et al. (2020c) creates a virtual image pair to reduce the difference between the images at the known and prediction times to increase accuracy. VIPSTF-SU is performed by applying VIPSTF to the existing spatial unmixing-based STDFA method. Different from STDFA, VIPSTF-SU utilizes the virtual fine spatial resolution image to acquire the thematic map before upscaling it to synthesize

the coarse proportions. Moreover, **E** represents the temporal change of the reflectances of land cover classes between the virtual coarse image and the coarse image at the prediction time, and **Q** represents the corresponding temporal change of the reflectances of the coarse pixels in a local window. The final prediction is acquired by combining the predicted temporal change of the reflectance of land cover classes with the virtual fine spatial resolution image.

References

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," International Journal of Applied Earth Observation and Geoinformation. International Journal of Applied Earth Observation and Geoinformation 23, 132–141.

Belgiu, M., Stein, A., 2019. Spatiotemporal image fusion in remote sensing. Remote Sensing 11(7), 818.

Busetto, L., Meroni, M., Colombo, R., 2008. Combining medium and coarse spatial resolution satellite data to improve the estimation of sub-pixel NDVI time series. Remote Sensing of Environment 112(1), 118–131.

Chen, B., Huang, B., Xu, B., 2015. Comparison of spatiotemporal fusion models: A review. Remote Sensing 7(2), 1798–1835.

Chen, Y., Cao, R., Chen, J., Zhu, X., Zhou, J., Wang, G., Shen, M., Chen, X., Yang, W., 2020. A new cross-fusion method to automatically determine the optimal input image pairs for NDVI spatiotemporal data fusion. IEEE Transactions on Geoscience and Remote Sensing 58(7), 5179-5194.

Chiman, K., Bence, B., Feng, G., 2018. A hybrid color mapping approach to fusing MODIS and Landsat images for forward prediction. Remote Sensing 10(4), 520.

Das, M., Ghosh, S. K., 2016. Deep-STEP: A deep learning approach for spatiotemporal prediction of remote sensing data. IEEE Geoscience and Remote Sensing Letters 13, 1984–1988.

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 Transactions on Geoscience and Remote Sensing 44(8), 2207–2218.

Gevaert, C. M., Garcia-Haro, F. J., 2015. A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion. Remote Sensing of Environment 156, 34–44.

Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Sohlberg, R., 2000. Global land cover classification at the 1 km spatial resolution using a classification tree approach. International Journal of Remote Sensing 21, 1331–1364.

851 Hilker, T., Wulder, M. A., 2009. A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance
852 based on Landsat and MODIS. Remote Sensing of Environment 113(8), 1613–1627.
853 Houborg, R., McCabe, M. F., Gao, F., 2016. A spatio-temporal enhancement method for medium resolution LAI (STEM-LAI).
854 International Journal of Applied Earth Observation and Geoinformation 47, 15–29.
855 Huang, B., Song, H., 2012. Spatiotemporal reflectance fusion via sparse representation. IEEE Transactions on Geoscience and
856 Remote Sensing 50, 3707–3716.
857 Huang, B., Wang, J., Song, H., Fu, D., Wong, K., 2013. Generating high spatiotemporal resolution land surface temperature for
858 urban heat island monitoring. IEEE Geoscience and Remote Sensing Letters 10(5), 1011–1015.
859 Johnson, M. D., Hsieh, W. W., Cannon, A. J., 2016. Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation
860 indices and machine learning methods. Agricultural and Forest Meteorology 218–219, 74–84.
861 Ju, J., Roy, D. P., 2008. The availability of cloud-free Landsat ETM plus data over the conterminous United States and globally.
862 Remote Sensing of Environment 112, 1196–1211.
863 Lees, K. J., Quaife, T., Artz, R. R. E., 2018. Potential for using remote sensing to estimate carbon fluxes across northern peatlands–
864 A review. Science of The Total Environment 615, 857–874.
865 Li, A., Bo, Y., Zhu, Y., Guo, P., Bi, J., He, Y., 2013. Blending multi-resolution satellite sea surface temperature (SST) products
866 using Bayesian maximum entropy method. Remote Sensing of Environment 135, 52–63.
867 Li, X., Foody, G. M., Boyd, D. S., Ge, Y., Zhang, Y., Du, Y., Ling, F., 2020a. SFSDAF: An enhanced FSDAF that incorporates
868 sub-pixel class fraction change information for spatio-temporal image fusion. Remote Sensing of Environment 237, 111537.
869 Li, Y., Li, J., Lin, H., Jin, C., Antonio, P., 2020b. A new sensor bias-driven spatio-temporal fusion model based on convolutional
870 neural networks. SCIENCE CHINA Information Sciences 63(4), 140302.
871 Liu, M., Yang, W., Zhu, X., Chen, J., Chen, X., Yang, L., Helmer, E. H., 2019. An Improved Flexible Spatiotemporal DAta Fusion
872 (IFSDAF) method for producing high spatiotemporal resolution normalized difference vegetation index time series. Remote
873 Sensing of Environment 227, 74–89.
874 Liu, W., Zeng, Y., Li, S., Huang, W., 2020. Spectral unmixing based spatiotemporal downscaling fusion approach. International
875 Journal of Applied Earth Observation and Geoinformation 88, 102054.
876 Liu, X., Deng, C., Wang, S., Huang, G., Zhao, B., Lauren, P., 2016. Fast and accurate spatiotemporal fusion based upon extreme
877 learning machine. IEEE Geoscience and Remote Sensing Letters 13, 2039–2043.
878 Luo, Y., Guan, K., Peng, J., 2018. A generic and fully-automated method to fuse multiple sources of optical satellite data to generate
879 a high-resolution, daily and cloud-/gap-free surface reflectance product 214, 87–99.

880 Ma, J., Zhang, W., Marinoni, A., Gao, L., Zhang, B., 2018. An improved spatial and temporal reflectance unmixing model to
881 synthesize time series of Landsat-like images. Remote Sensing 10, 1388.
882 Meng, J. H., Du, X., Wu, B. F., 2013. Generation of high spatial and temporal resolution NDVI and its application in crop biomass
883 estimation. International Journal of Digital Earth 6, 203–218.
884 Ranchin, T., Wald, L., 2000. Fusion of high spatial and spectral resolution images: The ARSIS concept and its implementation.
885 Photogrammetric Engineering and Remote Sensing 66, 49–61.
886 Shen, H., Meng, X., Zhang, L., 2016. An integrated framework for the spatio-temporal-spectral fusion of remote sensing images.
887 IEEE Transactions on Geoscience and Remote Sensing 54, 7135–7148.
888 Song, H., Huang, B., 2013. Spatiotemporal satellite image fusion through one-pair image learning. IEEE Transactions on
889 Geoscience and Remote Sensing 51, 1883–1896.
890 Tang, Y., Wang, Q., Zhang, K., Atkinson, P. M., 2020. Quantifying the effect of registration error on spatio-temporal fusion. IEEE
891 Journal of Selected Topics in Applied Earth Observations and Remote Sensing 13, 487–503.
892 Tewes, A., Thonfeld, F., Schmidt, M., 2015. Using RapidEye and MODIS data fusion to monitor vegetation dynamics in semi-arid
893 rangelands in South Africa. Remote Sensing 7, 6510–6534.
894 Tobler, W. R., 1970. A computer movie simulating urban growth in the Detroit region. Economic Geography 46(2), 234-240.
895 Wang, J., Schmitz, O., Lu, M., Karssenberg, D., 2020a. Thermal unmixing based downscaling for fine resolution diurnal land
896 surface temperature analysis. ISPRS Journal of Photogrammetry and Remote Sensing 161, 76–89.
897 Wang, L., Wang, X., Wang, Q., 2020b. Using 250-m MODIS data for enhancing spatiotemporal fusion by sparse representation.
898 Photogrammetric Engineering and Remote Sensing 86(6), 383–392.
899 Wang, Q., Atkinson, P. M., 2018. Spatio-temporal fusion for daily Sentinel-2 images. Remote Sensing of Environment 204, 31–42.
900 Wang, Q., Tang, Y., Tong, X., Atkinson, P. M., 2020c. Virtual image pair-based spatio-temporal fusion. Remote Sensing of
901 Environment 249, 112009.
902 Wang, Q., Shi, W., Atkinson, P. M., 2020d. Information loss-guided multi-resolution image fusion. IEEE Transactions on
903 Geoscience and Remote Sensing 58(1), 45–57.
904 Wang, Z., Bovik, A. C., 2002. A universal image quality index. IEEE Signal Processing Letters 9, 81–84.
905 Weng, Q., Peng, F., Feng, G., 2014. Generating daily land surface temperature at Landsat resolution by fusing Landsat and MODIS
906 data. Remote Sensing of Environment 145(8), 55–67.

907 Wu, M., Niu, Z., Wang, C., Wu, C., Wang, L., 2012. Use of MODIS and Landsat time series data to generate high-resolution
908 temporal synthetic Landsat data using a spatial and temporal reflectance fusion model. Journal of Applied Remote Sensing
909 6(13), 063507.
910 Wu, P. H., Shen, H. F., Zhang, L. P., 2015. Integrated fusion of multi-scale polar-orbiting and geostationary satellite observations for
911 the mapping of high spatial and temporal resolution land surface temperature. Remote Sensing of Environment 156, 169–181.
912 Xie, D., Gao, F., Sun, L., Anderson, M., 2018. Improving spatial-temporal data fusion by choosing optimal input image pairs.
913 Remote Sensing 10(7), 1142.
914 Xu, Y., Huang, Y., Xu, Y., Cao, K., Guo, C., Meng, D., 2015. Spatial and temporal image fusion via regularized spatial unmixing.
915 IEEE Geoscience and Remote Sensing Letters 12(6), 1362–1366.
916 Xue, J., Leung, Y., Fung, T., 2017. A Bayesian data fusion approach to spatio-temporal fusion of remotely sensed images. Remote
917 Sensing 9, 1310.
918 Zhang, H. K., Chen, J. M., Huang, B., 2014. Reconstructing seasonal variation of Landsat vegetation index related to leaf area index
919 by fusing with MODIS data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7, 950–960.
920 Zhang, X., Jayavelu, S., Liu, L., Friedl, M. A., Henebry, G. M., Liu, Y., 2018. Evaluation of land surface phenology from VIIRS data
921 using time series of PhenoCam imagery. Agricultural and Forest Meteorology 256, 137–149.
922 Zhu, X., Chen, J., Gao, F., 2010. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous
923 regions. Remote Sensing of Environment 114(11), 2610–2623.
924 Zhu, X., Helmer, E. H., Gao, F., Liu, D., Chen, J., Lefsky, M. A., 2016. A flexible spatiotemporal method for fusing satellite images
925 with different resolutions. Remote Sensing of Environment 172, 165–177.
926 Zhu, X., Cai, F., Tian, J., 2018. Spatiotemporal fusion of multisource remote sensing data literature survey, taxonomy, principles,
927 applications, and future directions. Remote Sensing 10(4), 527.
928 Zhukov, B., Oertel, D., Lanzl, F., 1999. Unmixing-based multisensor multiresolution image fusion. IEEE Transactions on
929 Geoscience and Remote Sensing 37(3), 1212–1226.
930 Zurita-Milla, R., Clevers, J. G. P. W., Schaepman, M. E., 2008. Unmixing-based Landsat TM and MERIS FR data fusion. IEEE
931 Geoscience and Remote Sensing Letters 5(3), 453–457.
932 Zurita-Milla, R., Kaiser, G., Clevers, J. G. P. W., Schneider, W., Schaepman, M. E., 2009. Downscaling time series of MERIS full
933 resolution data to monitor vegetation seasonal dynamics. Remote Sensing of Environment 113, 1874–1885.

934 Zurita-Milla, R., Gómez-Chova, L., Guanter, L., Clevers, J. G. P. W., Camps-Valls, G., 2011. Multitemporal unmixing of
935 medium-spatial-resolution satellite images: A case study using MERIS images for land-cover mapping. IEEE Transactions on
936 Geoscience and Remote Sensing 49(11), 4308–4317.
937

Fig. 1. An example for illustration of two adjacent cases (w=3). (a) and (b) represent the side- and vertex-adjacent cases, respectively. The pixels covered by diagonals at minus 45° represent distinct coarse pixels in a 3×3 window centered at the pixel marked by the red solid star. The pixels covered by diagonals at 45° represent distinct coarse pixels in a 3×3 window centered at the pixel marked by the red hollow star. The pixels covered by checks represent shared coarse pixels of the two local widows.

Fig. 2. An example for illustration of the block effect. The trapezoid represents an object shared by neighboring coarse pixels. (a) is a prediction in which each part displays different colors. (b) is the reference image with fixed color.

Fig. 3. Flowchart of the proposed SU-BR method. All bands of the image follow this scheme one by one.

Fig. 4. Landsat (first line) and MODIS (second line) images for the heterogeneous region acquired on (a) 5 January 2002 and (b) 13 February 2002, for the region with land cover changes acquired on (c) 14 February 2005 and (d) 3 April 2005, and for the homogeneous region acquired on (e) 4 December 2001 and (f) 5 January 2002. All images use NIR-red-green as RGB.

Fig. 5. Predictions for the heterogeneous region based on UBDF coupled with different blocks-removed methods. (a) UBDF. (b) UBDF-NM. (c) UBDF-SF. (d) UBDF-BR. (e) Reference. The images in the second-to-fourth lines are the corresponding predictions for the three sub-areas marked in yellow in the first line.

Fig. 6. Predictions for the heterogeneous region based on STDFA coupled with different blocks-removed methods. (a) STDFA. (b) STDFA-NM. (c) STDFA-SF. (d) STDFA-BR. (e) Reference. The images in the second-to-fourth lines are the corresponding predictions for the three sub-areas marked in yellow in the first line.

Fig. 7. Predictions for the heterogeneous region based on VIPSTF-SU coupled with different blocks-removed methods. (a) VIPSTF-SU. (b) VIPSTF-SU-NM. (c) VIPSTF-SU-SF. (d) VIPSTF-SU-BR. (e) Reference. The images in the second-to-fourth lines are the corresponding predictions for the three sub-areas marked in yellow in the first line.

Fig. 8. Blocks-removed temporal change images for the original spatial unmixing and SU-BR methods. (a) STDFA (left) and STDFA-BR (right) predictions for the red band. (b) VIPSTF-SU (left) and VIPSTF-SU-BR (right) predictions for the red band. The images in the second line are the corresponding predictions for the sub-area marked in black in the first line.

Fig. 9. Predictions for the region with land cover changes based on UBDF coupled with different blocks-removed methods. (a) Landsat at the known time. (b) UBDF. (c) UBDF-NM. (d) UBDF-SF. (e) UBDF-BR. (f) Reference. The images in the second line are the corresponding predictions for the sub-area marked in yellow in the first line.

Fig. 10. Predictions for the region with land cover changes based on STDFA coupled with different blocks-removed methods. (a) Landsat at the known time. (b) STDFA. (c) STDFA-NM. (d) STDFA-SF. (e) STDFA-BR. (f) Reference. The images in the second line are the corresponding predictions for the sub-area marked in yellow in the first line.

Fig. 11. Predictions for the region with land cover changes based on VIPSTF-SU coupled with different blocks-removed methods. (a) Landsat at the known time. (b) VIPSTF-SU. (c) VIPSTF-SU-NM. (d) VIPSTF-SU-SF. (e) VIPSTF-SU-BR. (f) Reference. The images in the second line are the corresponding predictions for the sub-area marked in yellow in the first line.

Fig. 12. Predictions for the homogeneous region based on UBDF coupled with different blocks-removed methods. (a) UBDF. (b) UBDF-NM. (c) UBDF-SF. (d) UBDF-BR. (e) Reference. The images in the second line are the corresponding predictions for the sub-area marked in yellow in the first line.

Fig. 13. Predictions for the heterogeneous region using different methods. (a) STARFM. (b) FSDAF. (c) UBDF-BR. (d) STDFA-BR. (e) VIPSTF-SU-BR. (f) Reference. The images in the second line are the corresponding predictions for the sub-area marked in yellow in the first line.

Fig. 14. Predictions for the region with land cover changes using different methods. (a) STARFM. (b) FSDAF. (c) UBDF-BR. (d) STDFA-BR. (e) VIPSTF-SU-BR. (f) Reference. The images in the second line are the corresponding predictions for the sub-area marked in yellow in the first line.

Fig. 15. Predictions for the homogeneous region using different methods. (a) STARFM. (b) FSDAF. (c) UBDF-BR. (d) STDFA-BR. (e) VIPSTF-SU-BR. (f) Reference. The images in the second line are the corresponding predictions for the sub-area marked in yellow in the first line.

Fig. 16. Histograms of the data fidelity term *R* (left) and the spatial continuity constraint term *D* (right) produced based on the predictions of the original STDFA method. (a) and (b) heterogeneous region. (c) and (d) region with land cover changes. (e) and (f) homogeneous region.

Fig. 17. The impact of α on the accuracy of STDFA-BR and VIPSTF-SU-BR. (a) Heterogeneous region. (b) Region with land cover changes. (c) Homogeneous region. The dotted line and dashed line represent the accuracies of SU-NM and SU-SF, respectively.