

Geographically Weighted Spatial Unmixing for Spatio-Temporal Fusion

Kaidi Peng, Qunming Wang, Yijie Tang, Xiaohua Tong, Peter M. Atkinson

Abstract—Spatio-temporal fusion is a technique applied to create images with both fine spatial and temporal resolutions by blending images with different spatial and temporal resolutions. Spatial unmixing is a widely used approach for spatio-temporal fusion, which requires only the minimum number of input images. However, ignorance of spatial variation in land cover between pixels is a common issue in existing spatial unmixing methods. For example, all coarse neighbors in a local window are treated equally in the unmixing model, which is inappropriate. Moreover, determination of the appropriate number of clusters in the known fine spatial resolution image remains a challenge. In this paper, a geographically weighted spatial unmixing (SU-GW) method was proposed to address the spatial variation in land cover and increase the accuracy of spatio-temporal fusion. SU-GW is a general model suitable for any spatial unmixing methods. Specifically, the existing regularized version and soft classification-based version were extended with the proposed geographically weighted scheme, producing 24 versions (i.e., 12 existing versions were extended to 12 corresponding geographically weighted versions) for spatial unmixing. Furthermore, the cluster validity index of Xie and Beni (XB) was introduced to determine automatically the number of clusters. A systematic comparison between the experimental results of the 24 versions indicated that SU-GW was effective in increasing the prediction accuracy. Importantly, all 12 existing methods were enhanced by integrating the SU-GW scheme. Moreover, the identified most accurate SU-GW enhanced version was demonstrated to outperform two prevailing spatio-temporal fusion approaches in a benchmark comparison. Therefore, it can be concluded that SU-GW provides a general solution for enhancing spatio-temporal fusion, which can be used to update existing methods as well as future potential versions.

Index Terms—Image fusion, spatio-temporal fusion, spatial unmixing, geographical weighting (GW).

NOMENCLATURE

BR	The blocks removed scheme for spatial unmixing
FCM	The fuzzy c -means scheme for spatial unmixing
FSDAF	The flexible spatio-temporal data fusion method
GW	The geographically weighted scheme for spatial unmixing

This work was supported by the National Natural Science Foundation of China under Grants 42171345 and 41971297, Fundamental Research Funds for the Central Universities under Grant 02502150021, and Tongji University under Grant 02502350047. (*Corresponding authors: Q. Wang*).

K. Peng, Q. Wang, Y. Tang and X. Tong are with the College of Surveying and Geo-Informatics, Tongji University, 1239 Siping Road, Shanghai 200092, China (e-mails: wqm1111@126.com).

P. M. Atkinson is with the Faculty of Science and Technology, Lancaster University, Lancaster LA1 4YR, UK and also with Geography and Environment, University of Southampton, Highfield, Southampton SO17 1BJ, UK.

STARFM	The spatial and temporal adaptive reflectance fusion model
STDFA	The spatial and temporal data fusion approach
SU	Spatial unmixing (i.e., one of the UBDF, STDFA and VIPSTF-SU choices)
SU-BR	Blocks-removed spatial unmixing
SU-B-G	Spatial unmixing integrating the BR and GW schemes
SU-FCM	FCM-based spatial unmixing
SU-F-B	Spatial unmixing integrating the FCM and BR schemes
SU-F-B-G	Spatial unmixing integrating the FCM, BR and GW schemes
SU-F-G	Spatial unmixing integrating the FCM and GW schemes
SU-GW	Geographically weighted spatial unmixing
UBDF	The unmixing-based data fusion method
VIPSTF-SU	Virtual image pair-based spatio-temporal fusion
XB	Cluster validity index of Xie and Beni

I. INTRODUCTION

Remote sensing satellites offer effective data sources for monitoring the Earth's surface. Free access to the data provided by a number of satellites (e.g., MODerate resolution Imaging Spectroradiometer (MODIS) and Landsat data) facilitates greatly repeated observation at the global scale [1], [2]. With an increasing demand for real-time and precise terrestrial monitoring, such data with both fine spatial and temporal resolutions are in great demand. Due to technical and budget limitations, however, existing single satellite sensors can provide remotely sensed images with either fine temporal resolution (e.g., daily MODIS images) or fine spatial resolution (e.g., 30 m Landsat images), but not both. To cope with this issue, spatio-temporal fusion methods have been developed over the decades [3]-[9]. Spatio-temporal fusion aims to create time-series images with both fine spatial and temporal resolutions, by blending images with fine spatial resolution, but coarse temporal resolution (e.g., 16-day, 30 m Landsat images) and images with fine temporal resolution, but coarse spatial resolution (e.g., daily, 500 m MODIS images). In recent years, spatio-temporal fusion has been developed rapidly and applied widely in a variety of domains, such as the monitoring of phenology dynamics [10]-[13], land surface temperature [14]-[18], suspended particulate matter [19], leaf area index [20]-[22], flooding [23], [24] and evapotranspiration [25].

Generally, three main groups of spatio-temporal fusion approaches can be identified: spatial unmixing-based methods

[26]-[32], spatial weighting-based methods [33]-[36] and machine learning-based methods [37]-[42]. Based on these categories, some hybrid methods have also been proposed [43]-[45]. These types of methods will be introduced briefly in the following.

A. Spatial unmixing-based methods

The concept of spatial unmixing is substantially different from the well-known spectral unmixing. The latter aims to estimate the proportion of each class within the coarse pixel and the class reflectances (also termed endmembers) are known, while the former aims to estimate the class reflectances and the class proportions are known (by referring to the fine spatial resolution data at the adjacent time). Spatial unmixing-based spatio-temporal fusion methods have been popular for their clear physical meaning and the requirement of minimum number of input images. Gevaert and Garc á-Haro [46] made a systematic comparison between the unmixing-based data fusion (UBDF) and the spatial and temporal adaptive reflectance fusion model (STARFM), which are typical exemplars of spatial unmixing- and spatial weighting-based models, respectively. The results indicate that spatial unmixing is more advantageous when very few fine spatial resolution images (fine images hereafter) are available. The multisensor multiresolution technique proposed by Zhukov *et al.* [47] is one of the earliest studies to present the spatial unmixing-based method, where the influence of selected parameters and sensitivity to sensor noise were discussed systematically. Zurita-Milla *et al.* [26] downscaled MEdium Resolution Imaging Spectrometer (MERIS) images to the Landsat-like spatial resolution using UBDF, with an analysis of the optimal number of clusters and window size. The spatial temporal data fusion approach (STDFA) proposed by Wu *et al.* [29] enhances the prediction by taking full advantage of the known fine image. The spatial unmixing-based virtual image pair-based spatio-temporal fusion (VIPSTF-SU) method proposed by Wang *et al.* [30] decreases the difference in feature space between images at the known and prediction times by introducing the concept of the virtual image pair. Zhou and Zhong [48] developed an image fusion model based on a Kalman Filter algorithm that can estimate the uncertainty in spatial unmixing.

B. Spatial weighting-based methods

Amongst the spatial weighting-based methods, STARFM [33] is one of the most widely applied methods, which employs spatial filtering to estimate the reflectance of the fine spatial resolution pixels based on the surrounding spectrally similar pixels. STARFM was adjusted to predict changes in vegetation by introducing a disturbance index to capture the changes [34]. The enhanced STARFM method proposed by Zhu *et al.* [35] further enhanced prediction in heterogeneous landscapes by using a conversion coefficient to model the differences in reflectance changes between the coarse and fine spatial resolutions. Fit-FC [36] was proposed to address the problem of strong temporal changes such as seasonal changes.

C. Machine learning-based methods

Machine learning-based methods have been developed for spatio-temporal fusion. Huang and Song [37] proposed to establish the relation between the fine and coarse data via sparse

representation based on two fine and coarse image pairs. Song and Huang [38] further extended the sparse representation method for the case with only one fine and coarse image pair. Wu *et al.* [39] utilized semi-coupled dictionary learning to fit the relationship between images at different spatial resolutions images with an L1/L2 mixed regularization sparse coding scheme. In [40], a nonlinear mapping convolutional neural network was proposed to replace the sparse representation method for relating MODIS and Landsat images. Moosavi *et al.* [41] proposed a hybrid wavelet-artificial intelligence fusion approach to predict fine spatio-temporal resolution land surface temperature data. Liu *et al.* [42] developed a two-stream convolutional neural network to consider the temporal dependence between fine images and formulate a temporal constraint to fully exploit the temporal information in the time-series.

D. Hybrid methods

Several hybrid methods have also been developed by combining the strategies of spatial weighting and spatial unmixing. The Flexible Spatiotemporal DAta Fusion (FSDAF) method [43] utilizes spatial unmixing to predict the temporal changes of each land cover class and spatial weighting to ensure spatial continuity. In the improved FSDAF method proposed by Liu *et al.* [44], the time-dependent changes predicted by spatial unmixing and space-dependent changes predicted by the thin plate spline interpolation are integrated. To address changes in land cover class, Li *et al.* [45] introduced an enhanced FSDAF that incorporates sub-pixel class proportion change information by employing a linear spectral unmixing strategy.

E. Extended versions of spatial unmixing

1) Regularized spatial unmixing

In spatial unmixing-based spatio-temporal fusion, prediction of reflectances may sometimes be challenging, due to the inherent uncertainty of the linear mixing model (mixtures of land cover in practice can be more complex than be represented by the model), collinearity between endmembers and complex spatial distribution of land cover classes [47], [49], [50]. Thus, a variety of extended versions have been developed to regularize the solution by adding different constraints to the spatial unmixing model. Zhukov *et al.* [47] and Amor ós-L ópez *et al.* [49] applied a preset class reflectance estimated from coarse images as a regularization term to reduce the uncertainty in the prediction of class reflectance. Xu *et al.* [32] incorporated the fine spatial resolution class spectral signature estimated using STARFM in advance. The constrained spatial unmixing model based on *prior* information, however, can be difficult to apply in practice since such information may be unavailable or imprecise. To address the common problem of the block effect which exists widely in spatial unmixing, Wang *et al.* [51] proposed the blocks-removed spatial unmixing (SU-BR) method. SU-BR constructs a spatial continuity term by assuming that the reflectances of the same class in spatially adjacent pixels are the same. The constraint is incorporated without the need of any ancillary information. It is regarded as a dynamic constraint and is updated iteratively to balance with the original data fidelity term, and to approach the optimal prediction gradually. SU-BR is, thus, a practical solution amongst the family of regularized versions.

2) Soft classification-based spatial unmixing

The spatial unmixing-based methods normally require a known fine spatial resolution map (usually obtained by land cover classification of a multispectral image) to characterize the distribution of land cover classes at the prediction time, based on the assumption of no land cover changes between the two times. Conventional spatial unmixing methods are implemented using the hard classified land cover map at the known time, and the predictions fail to reproduce the intra-class spectral variation within each coarse pixel. Considering the existence of mixed pixels at the known fine image and the uncertainty in land cover mapping, Amorós-López *et al.* [49] performed soft classification of the known fine image to recover the intra-class spectral variation. Soft classification is also applied in [45] to extract sub-pixel scale land cover information. The unsupervised fuzzy c -means (FCM) method applied in [49] is a widely acknowledged soft classification approach in spatial unmixing (called SU-FCM hereafter). FCM characterizes the possibility of a pixel belonging to each class through a continua ranging from 0 to 1.

F. The proposed geographically weighted spatial unmixing

Due to the existence of spatial heterogeneity, spatial variation in land cover is extremely common in remotely sensed images. In spatial unmixing, the neighboring coarse pixels are used to aid the prediction of the class reflectance in the center pixel, but all neighboring coarse pixels are treated equally (i.e., their weights are all a constant of one). Considering the heterogeneity of landscapes, however, different coarse neighbors should contribute differently to the center pixel [52]-[54]. This is a common problem for existing spatial unmixing methods, including the extended versions based on regularization (e.g., the representative SU-BR version) and soft classification (e.g., the commonly used SU-FCM version). According to the *Tobler's First Law of Geography* [55], the further the distance to the target pixel is, the less spatial association the neighboring pixel has with the target pixel. Mathematically, the weights assigned to the neighboring pixels should decrease as the distance increases. The law provides a direct solution to quantify the influence of each neighboring coarse pixel. Correspondingly, in this paper, a geographical weighting (GW)-based spatial unmixing scheme (called SU-GW hereafter) was proposed to account for the spatial variation and characterize the contributions of coarse neighbors more reliably.

G. Determination of number of clusters in spatial unmixing

The known fine spatial resolution land cover map plays an important role in spatial unmixing. Owing to the lack of training data which may require costly and laborious field investigation, especially for long time-series data covering a large area, unsupervised classification has been used widely in spatial unmixing as it is user-friendly and can be implemented automatically. In unsupervised classification, the number of clusters is a key parameter. Until now, however, estimation of the optimal number of clusters in spatial unmixing has relied on empirical or *prior* knowledge. Previous studies indicated that 3-to-6 classes can be an appropriate choice [43], [44], but this may not be universal for all cases, especially for areas with great heterogeneity. Thus, a more practical method is needed urgently. In [56], a cluster validity index was proposed by Xie and Beni

(XB hereafter) to determine the optimal number of clusters in unsupervised classification by quantifying the compactness and separation of partitions. The XB index has been validated to be applicable in subpixel mapping [57]. In this paper, the XB index was explored to demonstrate its potential in spatial unmixing-based spatio-temporal fusion.

H. Contributions

The main contributions of this paper are listed as follows.

- 1) SU-GW was proposed to quantify adaptively the influences of coarse neighbors in the spatial unmixing model, and further, increase the accuracy of spatio-temporal fusion.
- 2) SU-GW provides a general framework suitable for any spatial unmixing method. It can be integrated conveniently with the extended versions based on regularization and soft classification, as mentioned earlier. In this paper, 12 existing spatial unmixing methods (three typical original spatial unmixing methods (i.e., UBDF, STDFA and VIPSTF-SU) coupled with two representative extended versions (i.e., SU-BR and SU-FCM) or one of them or neither) were extended to their corresponding geographically weighted versions.
- 3) A systematic comparison between the 12 existing versions and the 12 proposed geographically weighted versions was conducted to investigate the benefit of each scheme and identify the most accurate version. The accuracy of the most accurate version (i.e., SU-F-B-G) was compared with the prevailing spatio-temporal fusion methods (i.e., STARFM and FSDAF) to demonstrate the advantages of the former.
- 4) The XB index was applied to determine automatically the optimal number of clusters in unsupervised classification of the known fine image in spatial unmixing. This is different from current strategies that determine the number of clusters empirically or based on *prior* knowledge.

The remainder of this paper is organized as follows. In Section II, the principles of two enhanced version (i.e., SU-BR and SU-FCM) and the proposed SU-GW are illustrated. Furthermore, the integrated versions based on the three schemes are also presented. In Section III, experimental results on four datasets are provided to make a systematic comparison between the 24 methods and demonstrate the advantage of SU-GW and its integrated extensions. Section IV discusses the results, characteristics of SU-GW and the problems that remain open. A conclusion is made in Section V.

II. METHOD

A. Original spatial unmixing (SU)

Generally, a unified framework can be used to summarize the mechanisms of spatial unmixing-based spatio-temporal fusion as

$$\hat{\mathbf{L}}_p = \lambda \mathbf{L}_k + f(\mathbf{M}_p - \lambda \mathbf{M}_k) = \lambda \mathbf{L}_k + f(\mathbf{Q}). \quad (1)$$

In Eq. (1), $\hat{\mathbf{L}}_p$ is the fine image (e.g., Landsat image) predicted at the target time. The prediction is divided into two parts: the known fine image \mathbf{L}_k weighted by a coefficient λ and the

increment related to coarse difference image $\mathbf{Q} = \mathbf{M}_p - \lambda \mathbf{M}_k$, where \mathbf{M}_p and \mathbf{M}_k are the coarse images (e.g., MODIS images) at the prediction and known times, respectively. f is a function to predict the fine spatial resolution increment (i.e., a downscaling process). For spatial unmixing, it is characterized by a linear unmixing model, which aims to estimate the reflectances of the land cover classes in each coarse pixel based on a local window. Its explicit mathematical expression can be written as

$$\begin{bmatrix} Q_1 \\ \vdots \\ Q_i \\ \vdots \\ Q_N \end{bmatrix} = \begin{bmatrix} p_{11} & \cdots & p_{1c} & \cdots & p_{1C} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ p_{i1} & \cdots & p_{ic} & \cdots & p_{iC} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ p_{N1} & \cdots & p_{Nc} & \cdots & p_{NC} \end{bmatrix} \times \begin{bmatrix} e_1 \\ \vdots \\ e_c \\ \vdots \\ e_C \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_N \end{bmatrix} \quad (2)$$

where N is the number of coarse pixels in the local window and C is the number of land cover classes in the image. ε_i is the i -th element of the residual error term ε . Q_i is the observed reflectance (in difference image) of the coarse pixel at \mathbf{X}_i composing an $N \times 1$ vector \mathbf{Q} . e_c is the reflectance for land cover class c constituting a $C \times 1$ vector \mathbf{E} that needs to be solved. p_{ic} denotes the coarse proportion of class c in the coarse pixel at \mathbf{X}_i , which is derived by upscaling a known, temporally neighboring fine spatial resolution classified land cover map. All the proportions constitute an $N \times C$ matrix \mathbf{P} . Generally, \mathbf{E} can be solved by the least squares rule based on the objective function

$$\hat{\mathbf{E}} = \arg \min_{\mathbf{E}} \|\varepsilon\|_2^2 = \arg \min_{\mathbf{E}} \|\mathbf{PE} - \mathbf{Q}\|_2^2. \quad (3)$$

The decomposed reflectances are assigned to the corresponding land cover classes of the fine pixels to construct the fused image.

Different choices of λ lead to different spatial unmixing-based methods. In this paper, three typical methods were considered, as illustrated briefly in the following.

1) UBDF ($\lambda = 0$)

For UBDF, the weight λ in Eq. (1) equals zero, which means the original fine image \mathbf{L}_k is not used. That is, the spatial unmixing process is operated directly on the observed coarse image \mathbf{M}_p .

2) STDFA ($\lambda = 1$)

To reproduce more intra-class spectral variation, STDFA was developed to incorporate the original fine image \mathbf{L}_k and the weight λ equals one. Specifically, the coarse difference image (i.e., $\mathbf{M}_p - \mathbf{M}_k$) is unmixed to obtain the reflectance change of each land cover class at the fine spatial resolution (i.e., the fine difference image), and the final prediction is the combination of the predicted fine difference image and the known fine image \mathbf{L}_k .

3) VIPSTF-SU (λ needs to be estimated)

In VIPSTF-SU, λ is a specific value calculated based on the defined concept of the virtual image pair. The created virtual image pair (i.e., $\lambda \mathbf{M}_k$ and $\lambda \mathbf{L}_k$) is closer to the data at the prediction time than the observed image pair (i.e., \mathbf{M}_k and \mathbf{L}_k)

in feature space to reduce the uncertainty in the unmixing-based downscaling process. The predicted virtual difference image at the fine spatial resolution will be combined with the virtual fine image $\lambda \mathbf{L}_k$ to obtain the final fused image.

Hereafter, the set of three spatial unmixing-based methods (i.e., UBDF, STDFA and VIPSTF-SU) is uniformly denoted as SU.

B. Blocks-removed spatial unmixing (SU-BR)

In the family of regularized spatial unmixing models, SU-BR proposed by Wang *et al.* [51] is a practical method without the need for any ancillary information. It was developed to remove the commonly existing blocky artifacts in spatial unmixing. In SU-BR, the spatial continuity is used as a constraint in addition to the objective function in Eq. (3) (i.e., maximizing data fidelity) and an iterative scheme is employed to balance data fidelity and spatial continuity. The new objective function of SU-BR is written as

$$\begin{aligned} \hat{\mathbf{E}}_i^{(t)} &= \arg \min_{\mathbf{E}_i^{(t)}} J_i = \alpha R_i^{(t)} + (1-\alpha) A D_i^{(t)} \\ &= \alpha \left\| \mathbf{PE}_i^{(t)} - \mathbf{Q} \right\|_2^2 + (1-\alpha) A \frac{\sum_{c=1}^C \sum_{j=1}^{N_0} \left[I_{i,j,c} (E_{i,c}^{(t)} - E_{j,c}^{(t-1)}) \right]^2}{\sum_{c=1}^C \sum_{j=1}^{N_0} I_{i,j,c}} \end{aligned} \quad (4)$$

where $I_{i,j,c}$ is an indicator function defined as

$$I_{i,j,c} = \begin{cases} 1, & \text{if pixels at } \mathbf{X}_i \text{ and } \mathbf{X}_j \text{ both contain class } c \\ 0, & \text{otherwise} \end{cases}. \quad (5)$$

In Eq. (4), R_i and D_i are the measurements of data fidelity and spatial continuity, respectively. $E_{i,c}$ 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. N_0 denotes the number of the neighbors in the local window. α is a balancing parameter taking a value between 0 and 1, A is a predefined magnitude regularization parameter and t is the iteration number. SU-BR is an optimization process, as the class reflectances of the neighboring pixels are updated one-by-one in each iteration, changing the constraint dynamically.

C. FCM-based spatial unmixing (SU-FCM)

Conventionally, unsupervised classification algorithms such as K-Means are used commonly to obtain the fine spatial resolution land cover map, following the assumption that the fine spatial resolution pixels are pure. With defined hard class labels, the estimated class reflectance is assigned to the corresponding fine pixels within each coarse pixel directly. That is, fine pixels belonging to the same class share exactly the same reflectance. Thus, the intra-class spectral variation is ignored inside each coarse pixel.

Generally, the signal of a pixel with a certain spatial scale in remotely sensed images can be a mixture of that for multiple classes, especially in highly heterogeneous regions, even if the spatial resolution is relatively fine [58]. To address the mixed pixel problem and characterize the intra-class spectral variation, fuzzy c -means (FCM), one of the representative soft classification methods, was developed alternatively for

interpretation of the known fine images [49], producing the SU-FCM method.

In SU-FCM, the degree to which a fine pixel belongs to one class is expressed not in terms of a binary 0 or 1, but alternatively by a continua that ranges between 0 and 1. The produced fine spatial resolution proportion map is degraded to the coarse spatial resolution to produce coarse proportions

$$p_{ic} = \frac{1}{M} \sum_{j=1}^M f_{jc}^{(i)} \quad (6)$$

where $f_{jc}^{(i)}$ denotes the FCM-estimated proportion of class c in the j -th fine pixel falling within the coarse pixel at \mathbf{X}_i , and M is the number of fine pixels within the coarse pixel. Following Eq. (2), the reflectance of each class can be estimated correspondingly. Afterwards, the reflectance of each fine pixel at the prediction time is the combination of the estimated class reflectances of all classes weighted by the corresponding proportions. The reflectance for the j -th fine pixel within \mathbf{X}_i , denoted as $q_j^{(i)}$, can be estimated as

$$q_j^{(i)} = \sum_{c=1}^C f_{jc}^{(i)} \times e_c. \quad (7)$$

As seen in Eq. (7), although the class reflectance e_c is fixed for a class, the FCM-estimated fine spatial resolution proportion $f_{jc}^{(i)}$ varies for each fine pixel, even for those belonging to the same class in the traditional non-FCM-based spatial unmixing. Thus, by SU-FCM, fine pixels of the same class in the non-FCM case normally present different reflectances, which is beneficial for reproducing intra-class spectral variation. It should be stressed that the FCM-based soft classification method is also applicable to SU-BR, leading to SU-F-B.

D. The criterion for defining the number of clusters

In spatial unmixing-based spatio-temporal fusion, the required fine spatial resolution land cover map is usually produced from the known fine images, following the basic assumption of a stable land cover distribution between the known and prediction times. Moreover, unsupervised classification is used widely, as it does not require training samples and is convenient to implement, especially for long time-series data. As one of the most crucial parameters in unsupervised classification, the number of clusters affects the proportion matrix \mathbf{P} in Eq. (3) directly as well as the stability of spatial unmixing. An overestimated number of clusters may generate classes with small-sized patches covered fully by only a single coarse pixel. In this case, the proportions of these classes in the neighboring coarse pixels are very likely to be zero; that is, there is no support from the neighbors for spatial unmixing, which will result in an unstable solution for these classes. On the other hand, an underestimated number of clusters may fail to characterize the spatial heterogeneity, as pixels with obviously different spectra can be grouped into the same class. Until now, however, the number of clusters in spatial unmixing has always been determined empirically or based on *prior* knowledge, which is costly or may be unreliable.

In this paper, the widely acknowledged cluster validity index of Xie and Beni (XB) [56], [57] was employed to choose the optimal number of clusters. It should be stressed that, the XB

value can only be calculated based on the known fine images instead of the target fine images which are actually unknown. Based on the assumption of no land cover changes, the number of clusters determined using the known fine image is rational for the fine image at the prediction time.

The expression of XB is written as

$$XB(C) = \frac{\sum_{i=1}^S \sum_{c=1}^C (u_{ci})^m \times \|\mathbf{v}_c - \mathbf{y}_i\|^2}{S \times \min_{c \neq k} \|\mathbf{v}_c - \mathbf{v}_k\|^2} \quad (8)$$

where \mathbf{y}_i is the spectra (vector) of the i -th fine pixel and S is the total number of pixels in the image. The fuzziness exponent m denotes the degree of class overlap and a value of two is widely used. \mathbf{v}_c or \mathbf{v}_k denotes the vector of the center of class c or k , and u_{ci} is the proportion of class c for the i -th pixel.

The number of clusters leading to the smallest XB value can be chosen as the optimal value. With the estimated number of clusters, the fine resolution classified map and the coarse proportion matrix \mathbf{P} in Eq. (3) are determined.

E. The proposed geographical weighting in spatial unmixing

In Eq. (2), each coarse pixel in the local window contributes equally to the center pixel (i.e., the weights are a constant of one), implying the inappropriate assumption that the reflectances of each land cover class do not show any spatial variation in the local window. However, according to *Tobler's First Law of Geography* [55], neighboring pixels nearer to the target pixel should have more influence on the analysis than those further away. The neglect of spatial variation in land cover class hampers the performance of the spatial unmixing-based methods inevitably. Thus, it is necessary to take spatial dependence into consideration. Alternatively, a weighting scheme (i.e., a geographical weighting) is incorporated into spatial unmixing to measure the influence of each neighboring pixel adaptively. As this is a common problem not only for the original spatial unmixing models, but also the extended versions (such as SU-BR and SU-FCM), the corresponding GW-based integrated versions are also developed in this section.

1) The proposed geographically weighted spatial unmixing (SU-GW)

In SU-GW, a weighting matrix with components calculated using a bi-square function [59] is incorporated into spatial unmixing. The bi-square function is written as

$$w_{ij} = \begin{cases} \left[1 - \left(d_{ij} / b \right)^2 \right]^2, & \text{if } d_{ij} < b \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

where d_{ij} is the spatial distance between the target pixel at \mathbf{X}_i and the neighboring pixel at \mathbf{X}_j and b is the bandwidth measuring the distance-decay. The weighting function provides a continuous, near-Gaussian distribution when the distance from the neighbor to the target pixel is less than b , and zero for any pixel beyond b . More precisely, the weight is the largest for the center pixel and decreases as the distance increases, as the spatial dependence decreases. When the distance is larger than b , the weight is zero, as the spatial dependence disappears. In this paper, b is set to half of the diagonal length of the local window in spatial unmixing, as the pixels beyond the local window are

not involved in the calculation for the center pixel. It is worth noting that the weighting function is not limited to Eq. (9), which is a practical choice. Other functions complying the *Tobler's First Law of Geography* are also encouraged, such as the Gaussian function.

Based on the quantification of contributions of spatial neighbors in Eq. (9), the goal in Eq. (3) becomes the minimization of the weighted residual; that is, Eq. (3) is extended by

$$\hat{\mathbf{E}} = \arg \min_{\mathbf{E}} \|\mathbf{W}^{1/2} \mathbf{e}\|_2^2 = \arg \min_{\mathbf{E}} \sum_{j=1}^N w_j (\mathbf{P}_j \mathbf{E} - Q_j)^2 \quad (10)$$

where \mathbf{W} is an $N \times N$ diagonal weighting matrix and the j -th diagonal element w_j is the weight assigned to the pixel at \mathbf{X}_j calculated based on Eq. (9). \mathbf{P}_j is the j -th row of proportion matrix \mathbf{P} , denoting the proportions of C land cover classes in the pixel at \mathbf{X}_j and Q_j is the corresponding observed reflectance for that pixel. As seen from Eq. (10), for a more distant neighbor the weight is smaller suggesting a reduced influence in the unmixing process. Fig. 1 depicts the weighting matrix graphically. For each pixel, the weight varies according to its location relative to the target center pixel.

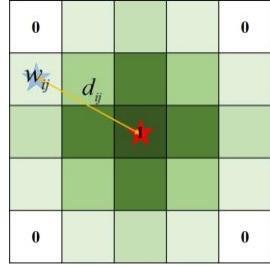


Fig. 1. Schematic diagram describing the weights assigned to neighboring pixels (with one of them marked by the blue star for example) in a local 5×5 pixel window based on Eq. (8). Darker green means a larger weight and *vice versa*. The weight of the target center pixel (marked by the red star) is one and the four pixels at the four corners contribute zero to the target pixel. All the weights vary between 0 and 1.

2) The proposed extended version of SU-B-G

The geographical weighting scheme can also be applied to the enhanced version of SU-BR described in Section II-B, producing the further extended version of SU-B-G. SU-B-G is developed by replacing the first term in Eq. (4) with Eq. (10). Thus, the new objective function is

$$\begin{aligned} \hat{\mathbf{E}}_i^{(t)} &= \arg \min_{\mathbf{E}_i^{(t)}} \alpha \sum_{j=1}^N w_j (\mathbf{P}_j \mathbf{E}_i^{(t)} - Q_j)^2 \\ &\quad + (1-\alpha) A \frac{\sum_{c=1}^C \sum_{j=1}^{N_0} [I_{i,j,c} (E_{i,c}^{(t)} - E_{j,c}^{(t-1)})]^2}{\sum_{c=1}^C \sum_{j=1}^{N_0} I_{i,j,c}} \end{aligned} \quad (11)$$

Likewise, the iterative process in [51] can be applied to optimize the objective function and predict the fine class reflectance.

3) The proposed extended versions of SU-F-G and SU-F-B-G

As mentioned in Section II-C, the BR and FCM schemes can also be integrated to produce an extended version SU-F-B. If we adopt the GW scheme for the SU-FCM-based versions, SU-FCM and SU-F-B can be further extended to SU-F-G and SU-F-B-G, respectively. In these two new methods, the

construction of the proportion matrix \mathbf{P} and the calculation of fine pixel reflectances are the same as for SU-FCM described in Section II-C. The SU-F-B-G version integrates all three schemes (i.e., BR, FCM and GW) into one model.

Fig. 2 presents the construction of the eight schemes (BR or not, FCM or not, and GW or not) applied to the three original SU methods (UBDF, STDFA and VIPSTF-SU), resulting in a total number of 12 new versions based on 12 existing non-GW versions. The relation between the different versions can be observed clearly based on the structure in Fig. 2. Mathematically, the main difference between the 12 proposed GW-based versions and the 12 original versions is that the original fidelity term in Eq. (3) is extended by including a weighting function (i.e., Eq. (10)). The performances of all 24 versions will be analyzed systematically in Section III.

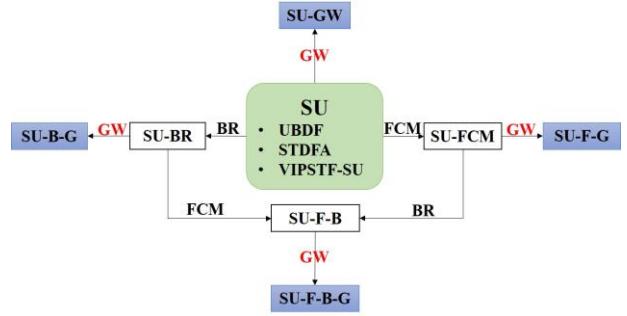


Fig. 2. The construction of SU-GW and its extended version coupled with FCM or BR or both (24 versions in all, including 12 existing versions and 12 corresponding geographically weighted versions proposed in this paper).

III. EXPERIMENTS

A. Data and experimental setup

Four datasets were used to examine the performances of SU-GW and its extended methods (i.e., the 24 versions in Fig. 2). The datasets cover two heterogeneous regions, one region with land cover changes and one homogeneous region.

The first heterogeneous region and the region with land cover changes are located in northern New South Wales, Australia, while the second heterogeneous region and the homogeneous region are located in southern New South Wales. For each region, two Landsat-MODIS image pairs were used, with the Landsat images acquired by TM (for the first heterogeneous region and the region with land cover changes) or ETM+ (for the second heterogeneous region and the homogeneous region) and the MODIS images acquired by Terra MOD09GA Collection 5. The spatial resolutions of the Landsat and MODIS images are 25 m and 300 m, respectively. The first heterogeneous region covers a spatial extent of 2 km by 2 km. The acquisition dates of the two image pairs are 14 February 2005 and 3 April 2005. The second heterogeneous region covers a 2 km by 2 km rice-based irrigation system. The two image pairs were acquired on 5 January 2002 and 13 February 2002. This area experienced significant phenological changes during the period. For the region with land cover changes, the two image pairs were acquired on 25 October 2004 and 26 November 2004. The flood inundation caused abrupt changes in the distribution of the water class. The homogeneous region covers a 1.8 km by 1.8 km area and the image pairs were acquired on 4 December 2001 and 5 January 2002.

The four pairs of Landsat-MODIS images are shown in Fig. 3. The task of spatio-temporal fusion for each region is to predict the latter Landsat image using the former Landsat-MODIS image pair and the latter MODIS image. The known latter Landsat image was used as a reference for accuracy assessment in each case. For quantitative assessment, five indices were used: the root mean square error (RMSE), relative global-dimensional synthesis error (ERGAS) [60], correlation coefficient (CC), universal image quality index (UIQI) [61] and spectral angle mapper (SAM). The prediction is more accurate when the RMSE, ERGAS and SAM values are smaller and CC and UIQI values are larger. They have been used widely for evaluation of image fusion methods, and have been introduced explicitly in a number of literature [62]. Thus, we did not introduce these indices again in this paper.

B. Validation of the estimated optimal number of clusters

The XB index was applied to determine the number of clusters for the four regions. Setting the range of number of clusters C as [3, 7], the calculated XB values for the first heterogeneous region, the region with land cover changes and the homogeneous

region are shown in Fig. 4(a)-(c), respectively. Meanwhile, the corresponding accuracies (in terms of CC) of SU-FCM (based on all three original SU versions) under enumerated number of clusters are also presented. As can be seen in Fig. 4, the CCs reach the peak while the XB value is the smallest in almost all cases. The predictions of the first heterogeneous region using STDFA-FCM with different number of clusters are listed in Fig. 5, where the optimal number of clusters is determined as 4 according to Fig. 4(a). Obvious spectral distortion (e.g., in the marked yellow circles in the sub-area) can be observed in Fig. 5(a) with an underestimated number of clusters. On the other hand, some small artifacts appear in Fig. 5(c)- Fig. 5(e) with overestimated number of clusters. Thus, both visual and quantitative assessments demonstrate the effectiveness of XB. Accordingly, in the following experiments, the number of clusters of the first heterogeneous region, the second heterogeneous region, the region with land cover changes and the homogeneous region are determined as 4, 5, 5 and 3, respectively.

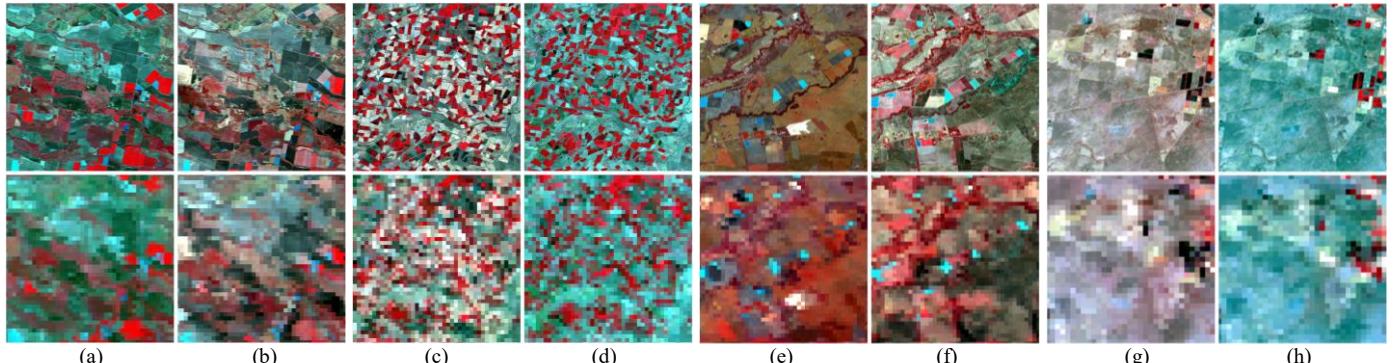


Fig. 3. (first line) Landsat and (second line) MODIS images for the first heterogeneous region acquired on (a) 14 February 2005 and (b) 3 April 2005, for the second heterogeneous region acquired on (c) 5 January 2002 and (d) 13 February 2002, for the region with land cover changes acquired on (e) 25 October 2004 and (f) 26 November 2004, and for the homogeneous region acquired on (g) 4 December 2001 and (h) 5 January 2002. All images use NIR-red-green as RGB.

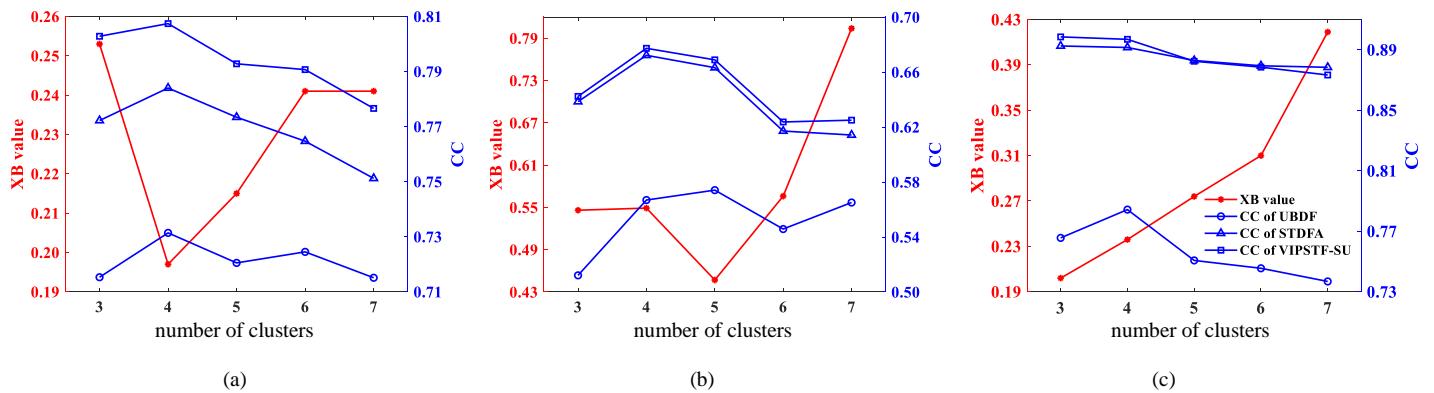


Fig. 4. (red line) XB and (blue lines) CC values under enumerated number of clusters. (a) The first heterogeneous region. (b) The region with land cover changes. (c) The homogeneous region. The number of clusters with the smallest XB was determined as the optimal one in the experiments.

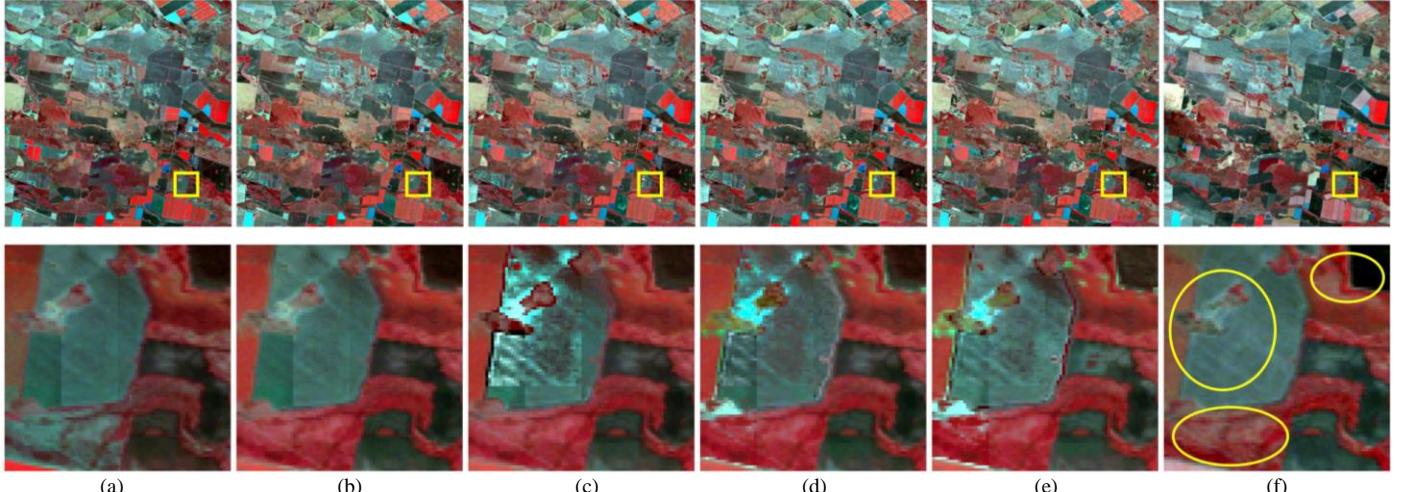


Fig. 5. The predictions of STDFA-FCM for the first heterogeneous region under different number of clusters. (a) 3. (b) 4. (c) 5. (d) 6. (e) 7. (f) Reference. The images in the second line are the corresponding zoomed predictions for the sub-area marked in yellow in the first line.

C. Comparison between GW and non-GW versions

1) SU and SU-GW

Figs. 6-9 display the results of the SU-based methods, including original SU, SU-BR, SU-FCM, SU-F-B and their corresponding GW-based versions for three regions. In Fig. 6, the results of all 24 versions are shown for systematical comparison. For clearer visual comparison between the results, a sub-area covering 60 by 60 Landsat pixels is shown for each case. It is observed clearly that the proposed GW-based methods can restore spectral information more accurately than the non-GW versions. For example, comparing Fig. 8(a) with Fig. 8(b), it is obvious that with the GW-based scheme, the color of the red patch is closer to the reference. Similar phenomenon can also be observed in many cases in Figs. 6-9. Another interesting observation is that the GW scheme can help to alleviate the block effect to some extent. This can be illustrated by comparing the zoomed parts in Fig. 6(a) and Fig. 6(b).

Fig. 10 shows the quantitative evaluation results for all four regions in terms of CC and RMSE, where the results of all 24 versions are included. Obviously, for all versions, larger CC and smaller RMSE are produced when the GW-based scheme is considered. Specifically, for the first heterogeneous region, the increases in CC are 0.0263, 0.0138 and 0.0147 from UBDF, STDFA and VIPSTF-SU to their corresponding GW versions, respectively. With respect to the region with land cover changes, the accuracies of the SU-GW are also considerably greater than SU, with remarkable gains in CC of 0.0469, 0.0331 and 0.0383 for UBDF, STDFA and VIPSTF-SU, respectively.

2) SU-BR and SU-B-G

Comparing the SU-BR prediction in Fig. 6(c) or Fig. 8(c) with the SU-B-G prediction in Fig. 6(d) or Fig. 8(d), we can see that the GW proposed scheme can further increase the accuracy of

the prediction, even though the employment of the BR scheme already enhanced the prediction.

The quantitative assessment results in Fig. 10 show that the GW scheme is beneficial to almost all the SU-BR methods. For the first heterogeneous region, the application of GW enhances the predictions with an increase in CC of 0.0280 for UBDF-BR, 0.0152 for STDFA-BR and 0.0160 for VIPSTF-SU-BR. Focusing on the region with land cover changes, the increases in CC are all above 0.0300 from SU-BR to SU-B-G.

3) SU-FCM and SU-F-G

With the GW scheme, the performances of the SU-FCM methods are also enhanced obviously. See, for example, the predictions in Fig. 7(e) and Fig. 7(f). From the quantitative assessment in Fig. 10, it is seen that for the first heterogeneous region, based on the choices of UBDF, STDFA and VIPSTF-SU, the CC values of SU-F-G are 0.0293, 0.0159 and 0.0181 larger than the corresponding non-GW versions. Regarding the region with land cover changes, the SU-F-G methods increase the CCs by 0.0392, 0.0189 and 0.0248 correspondingly.

4) SU-F-B and SU-F-B-G

Checking the results in Figs. 6 and 7, the color is restored more accurately by SU-F-B-G in comparison with SU-F-B. This can be observed clearly by comparing the VIPSTF-SU-F-B result in Fig. 7(g) to the VIPSTF-SU-F-B-G result in Fig. 7(h).

As can be seen from the CC and RMSE results presented in Fig. 10, the increase in CC and decrease in RMSE is prominent from SU-F-B to SU-F-B-G, especially for the region with land cover changes. More precisely, for this region, the increases in CC values are 0.0392, 0.0221 and 0.0303 for the three choices of UBDF, STDFA and VIPSTF-SU. Similarly, for the second heterogeneous region, with the GW scheme, the values of CC are 0.0097, 0.0038 and 0.0067 larger than corresponding SU-F-B methods.



Fig. 6. Landsat predictions for the first heterogeneous region based on the 24 methods. The first line presents the predictions of UBDF-based versions and the images in the second line are the corresponding zoomed predictions for the sub-area marked in yellow. The third line presents the predictions of STDFA-based versions and the images in the fourth line are the corresponding zoomed predictions. The fifth line presents the predictions of VIPSTF-SU-based versions and the images in the sixth line are the corresponding zoomed predictions. (a) SU. (b) SU-GW. (c) SU-BR. (d) SU-B-G. (e) SU-FCM. (f) SU-F-G. (g) SU-F-B. (h) SU-F-B-G. (i) Reference.

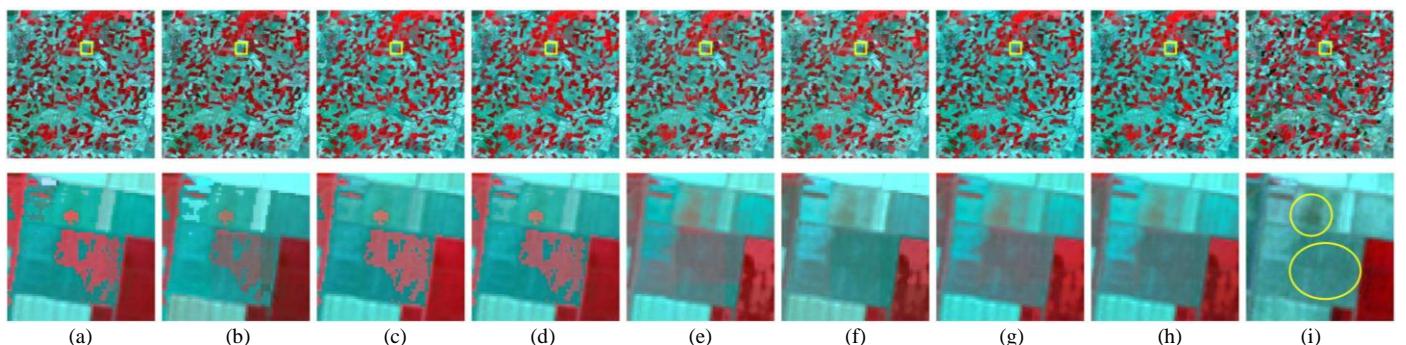


Fig. 7. Landsat predictions for the second heterogeneous region based on different enhanced VIPSTF-SU methods. (a) VIPSTF-SU. (b) VIPSTF-SU-GW. (c) VIPSTF-SU-BR. (d) VIPSTF-SU-B-G. (e) VIPSTF-SU-FCM. (f) VIPSTF-SU-F-G. (g) VIPSTF-SU-F-B. (h) VIPSTF-SU-F-B-G. (i) Reference. The images in the second line are the corresponding zoomed predictions for the sub-area marked in yellow in the first line.

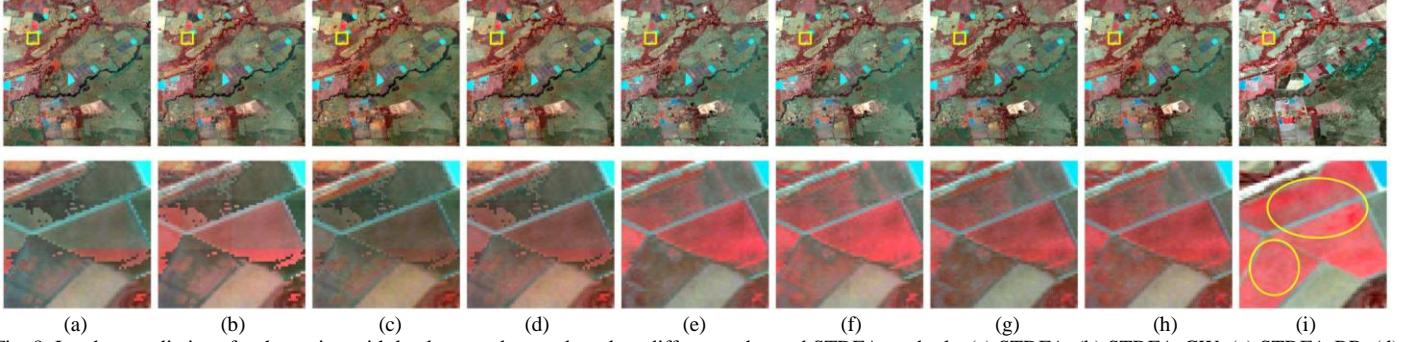


Fig. 8. Landsat predictions for the region with land cover changes based on different enhanced STDFA methods. (a) STDFA. (b) STDFA-GW. (c) STDFA-BR. (d) STDFA-B-G. (e) STDFA-FCM. (f) STDFA-F-G. (g) STDFA-F-B. (h) STDFA-F-B-G. (i) Reference. The images in the second line are the corresponding zoomed predictions for the sub-area marked in yellow in the first line.

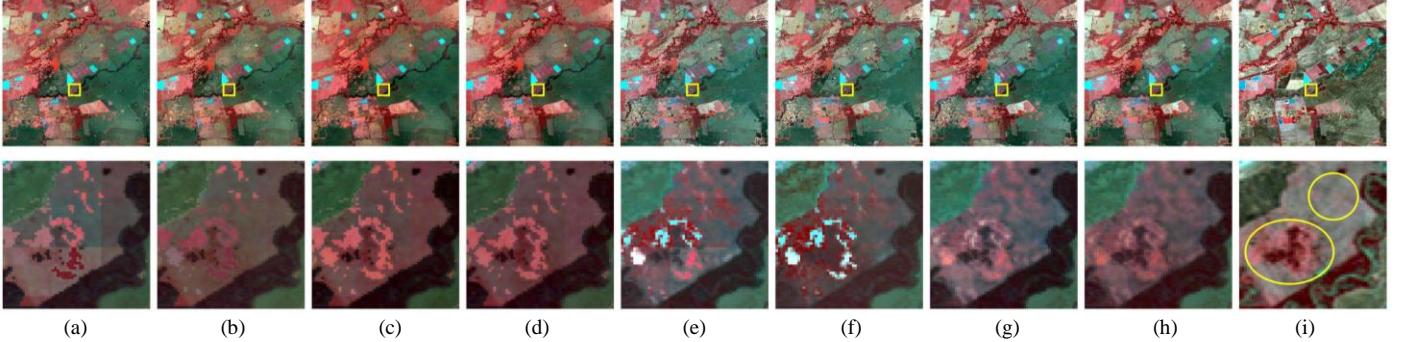


Fig. 9. Landsat predictions for the region with land cover changes based on different enhanced VIPSTF-SU methods. (a) VIPSTF-SU. (b) VIPSTF-SU-GW. (c) VIPSTF-SU-BR. (d) VIPSTF-SU-B-G. (e) VIPSTF-SU-FCM. (f) VIPSTF-SU-F-G. (g) VIPSTF-SU-F-B. (h) VIPSTF-SU-F-B-G. (i) Reference. The images in the second line are the corresponding zoomed predictions for the sub-area marked in yellow in the first line.

D. Systematic comparison between 24 versions

The strategies of FCM and BR are considered systematically in the proposed GW-based methods. In this section, we compare the performances between FCM and non-FCM, BR and non-BR, and F-B and non-F-B schemes.

First, the results predicted by the FCM scheme are all closer to the reference than the non-FCM-based results (e.g., see the red objects in the yellow circles in Fig. 7(e)-(h) and Fig. 7(a)-(d)). This is because FCM can restore the intra-class variation compared with the conventional *K*-Means-based spatial unmixing. Focusing again on the results in Fig. 10, for the second heterogeneous region, the increases in CC are larger than 0.0125 from SU to SU-FCM. With respect to the region with land cover changes, the predictions applying FCM have remarkable increases in CC of 0.0450, 0.0582 and 0.0614 for UBDF-, STDFA- and VIPSTF-SU-based methods, respectively, and the decreases in RMSE are 0.0011, 0.0021 and 0.0017 correspondingly.

Second, the blocks in predictions are removed noticeably with the BR scheme. Checking the results in Fig. 6, the dark green blocks in the sub-area are removed noticeably and the color is restored more accurately by the SU-BR-based methods than the original SU methods. Comparing the VIPSTF-SU-FCM result in

Fig. 9(e) to the VIPSTF-SU-F-B result in Fig. 9(g), we can see clearly the removal of blocky artifacts and more accurate restoration of spectral information when incorporating the BR scheme. In Fig. 10, the bars of the SU-BR-based methods are always larger in CC and smaller in RMSE compared with the corresponding non-BR-based methods. For the second heterogeneous region, the increases in CC values are over 0.0477, 0.0133 and 0.0189 with the employment of the BR scheme for UBDF, STDFA and VIPSTF-SU, respectively. Similarly, for the region with land cover changes, with BR applied, the values of CC are 0.0391, 0.0245 and 0.0338 larger for the corresponding non-BR-based methods.

Third, as seen from the results in Figs. 6-9, the performance of SU-F-B is more satisfactory than both SU-BR and SU-FCM. For example, when applying the FCM and BR schemes together, the VIPSTF-SU-F-B result in Fig. 9(g) restores the spectral information more accurately than Fig. 9(c) predicted by VIPSTF-SU-BR and Fig. 9(e) predicted by VIPSTF-SU-FCM. The same observation can be made when comparing Fig. 7(g) to Fig. 7(c) and Fig. 7(e). The CC and RMSE results in Fig. 10 also suggest that the combination of both the FCM and BR schemes lead to greater accuracy than the separate FCM or BR scheme.

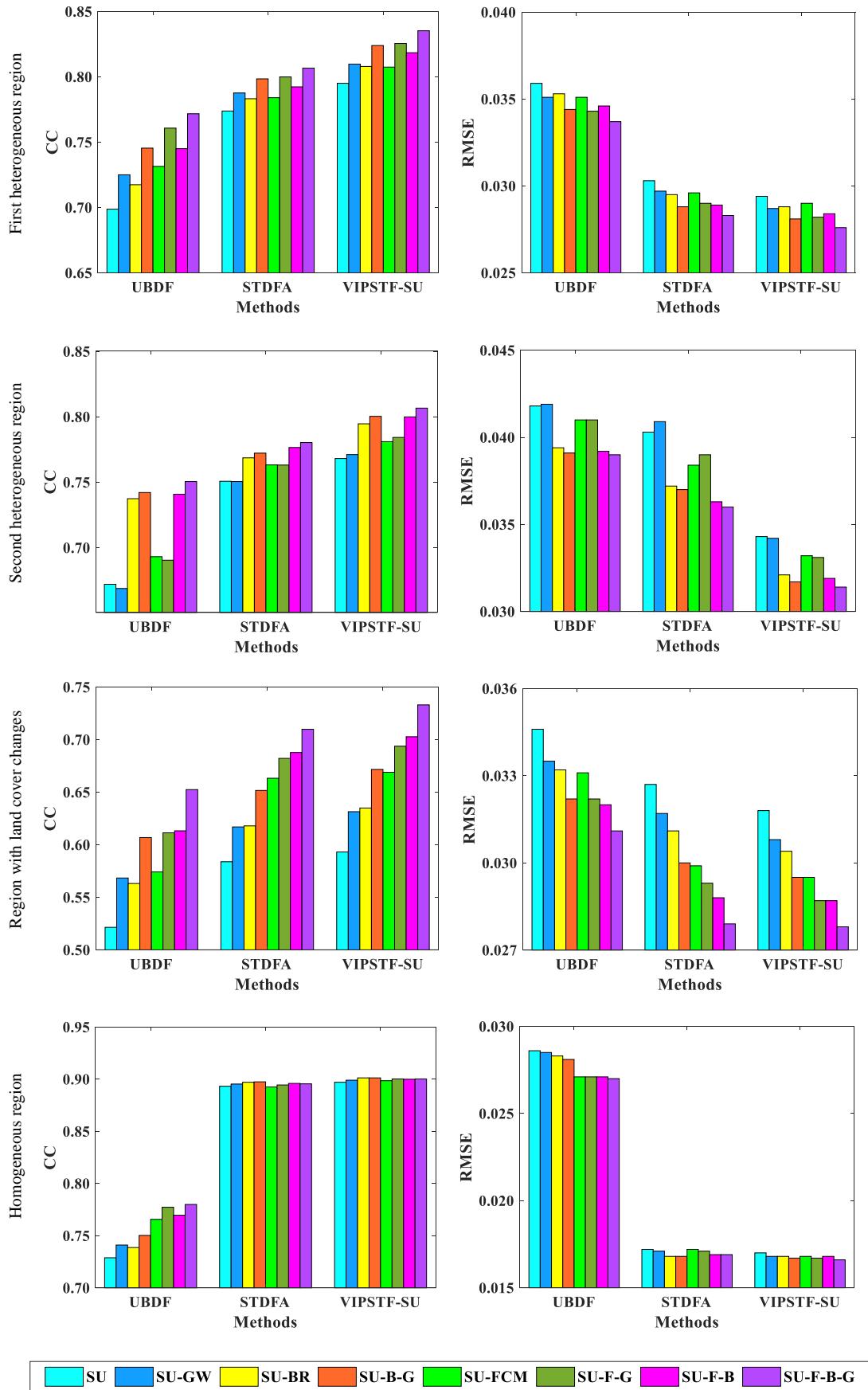


Fig. 10. The accuracy (in terms of CC and RMSE) of the 12 proposed GW-based and 12 corresponding non-GW-based spatial unmixing methods for different regions. SU represents one of the original spatial unmixing methods (i.e., UBDF, STDFA and VIPSTF-SU).

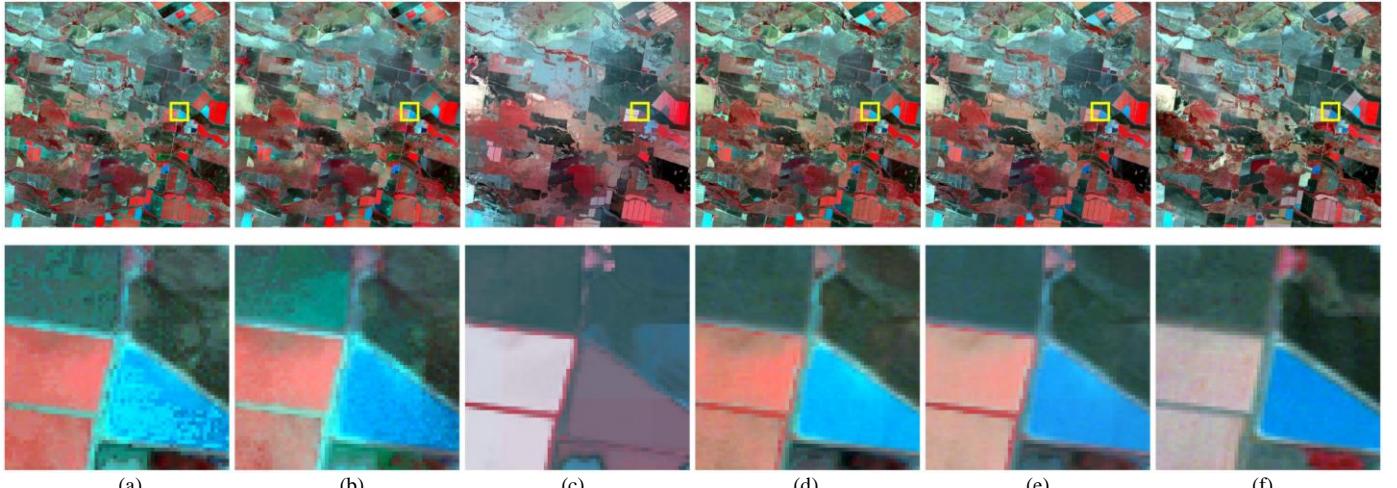


Fig. 11. Landsat predictions for the first heterogeneous region based on STARFM, FSDAF and SU-F-B-G. (a) STARFM. (b) FSDAF. (c) UBDF-F-B-G. (d) STDFA-F-B-G. (e) VIPSTF-SU-F-B-G. (f) Reference. The images in the second line are the corresponding zoomed predictions for the sub-area marked in yellow in the first line.

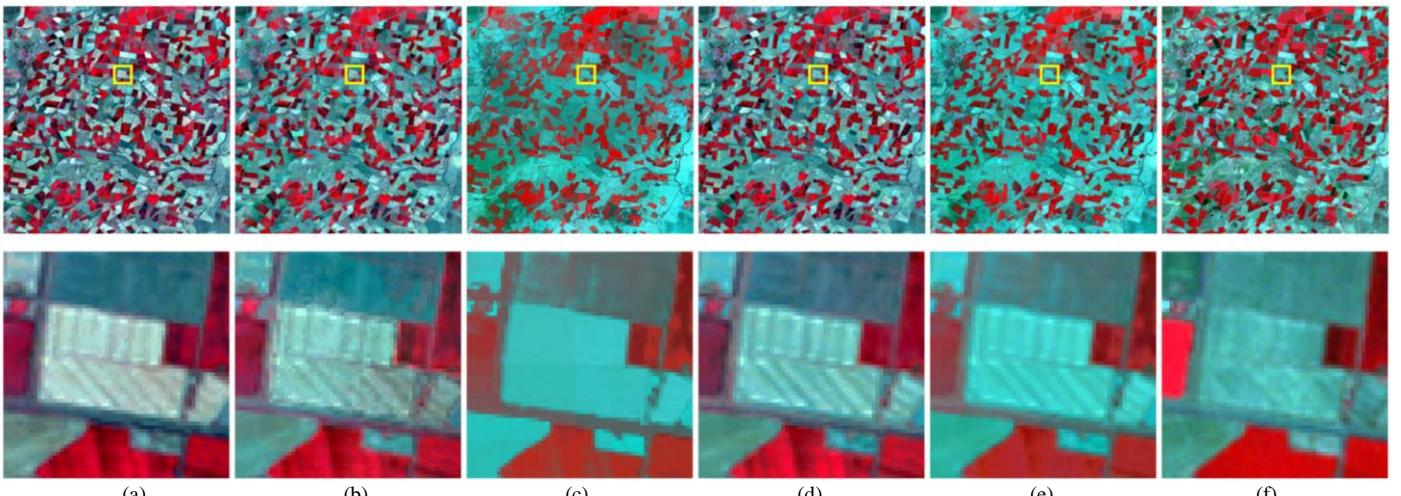


Fig. 12. Landsat predictions for the second heterogeneous region based on STARFM, FSDAF and SU-F-B-G. (a) STARFM. (b) FSDAF. (c) UBDF-F-B-G. (d) STDFA-F-B-G. (e) VIPSTF-SU-F-B-G. (f) Reference. The images in the second line are the corresponding zoomed predictions for the sub-area marked in yellow in the first line.

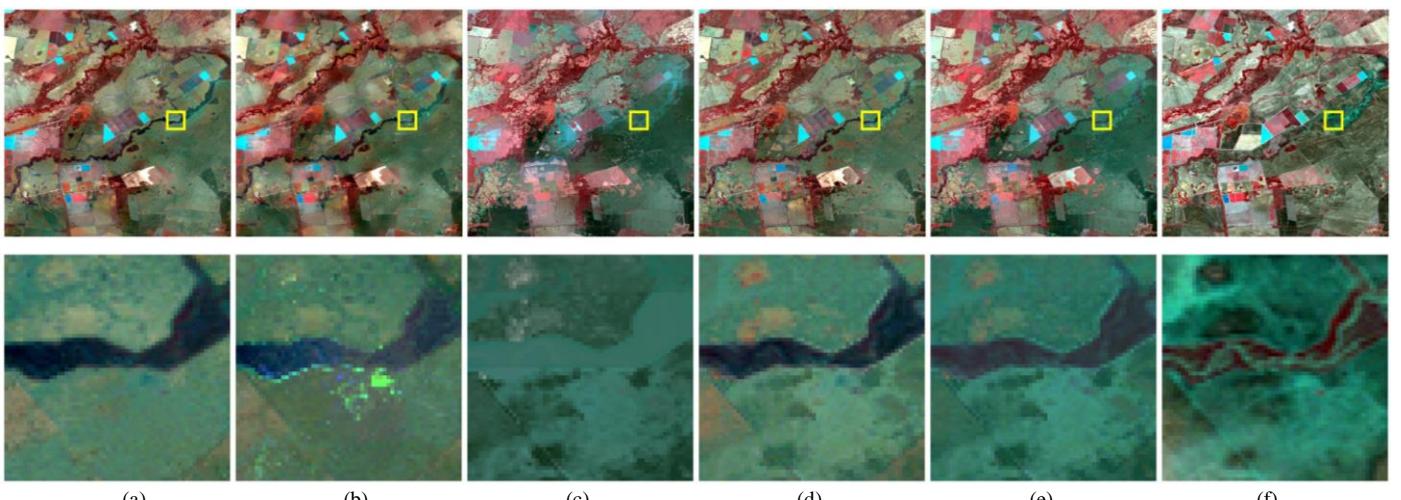


Fig. 13. Landsat predictions for the region with land cover changes based on STARFM, FSDAF and SU-F-B-G. (a) STARFM. (b) FSDAF. (c) UBDF-F-B-G. (d) STDFA-F-B-G. (e) VIPSTF-SU-F-B-G. (f) Reference. The images in the second line are the corresponding zoomed predictions for the sub-area marked in yellow in the first line.

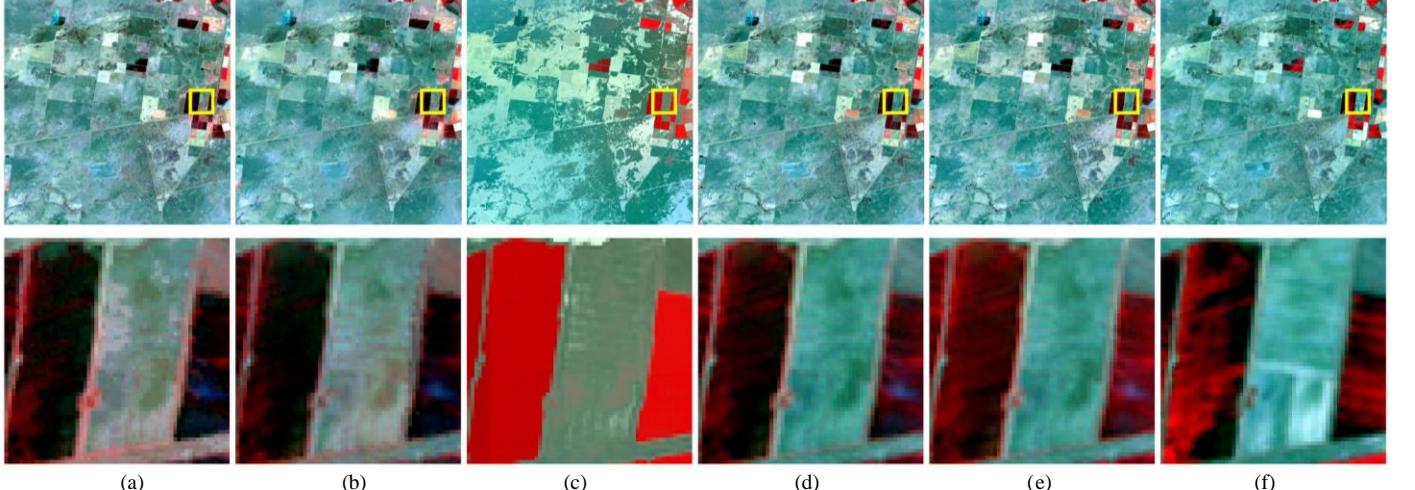


Fig. 14. Landsat predictions for the homogeneous region based on STARFM, FSDAF and SU-F-B-G. (a) STARFM. (b) FSDAF. (c) UBDF-F-B-G. (d) STDFA-F-B-G. (e) VIPSTF-SU-F-B-G. (f) Reference. The images in the second line are the corresponding zoomed predictions for the sub-area marked in yellow in the first line.

Table 1 Quantitative assessment of STARFM, FSDAF and SU-F-B-G

	Ideal	CC	RMSE	ERGAS	SAM	UIQI
First Heterogeneous region	STARFM	0.8037	0.0293	0.9608	0.1294	0.7943
	FSDAF	0.8155	0.0276	0.9088	0.1089	0.8025
	UBDF-F-B-G	0.7717	0.0337	1.1917	0.1217	0.7010
	STDFA-F-B-G	0.8066	0.0283	0.9312	0.1129	0.7965
Second Heterogeneous region	VIPSTF-SU-F-B-G	0.8352	0.0276	0.9454	0.0858	0.8025
	STARFM	0.8043	0.0411	1.6696	0.1758	0.7753
	FSDAF	0.8314	0.0357	1.4137	0.1522	0.8169
	UBDF-F-B-G	0.8005	0.0390	1.5070	0.2062	0.6659
Region with land cover changes	STDFA-F-B-G	0.8303	0.0360	1.4484	0.1533	0.8118
	VIPSTF-SU-F-B-G	0.8566	0.0314	1.1964	0.1450	0.8198
	STARFM	0.7329	0.0274	0.7920	0.1256	0.7136
	FSDAF	0.7213	0.0275	0.8025	0.1258	0.6937
Homogeneous region	UBDF-F-B-G	0.6524	0.0311	0.9070	0.1536	0.5761
	STDFA-F-B-G	0.7098	0.0279	0.8154	0.1211	0.6883
	VIPSTF-SU-F-B-G	0.7330	0.0278	0.7927	0.1315	0.6667
	STARFM	0.8897	0.0180	0.4228	0.0676	0.8876
	FSDAF	0.8940	0.0172	0.4144	0.0575	0.8852
	UBDF-F-B-G	0.7799	0.0270	0.6400	0.0882	0.6829
	STDFA-F-B-G	0.8955	0.0169	0.4113	0.0510	0.8913
	VIPSTF-SU-F-B-G	0.9002	0.0166	0.4054	0.0502	0.8852

E. Comparison with other spatio-temporal fusion methods

As concluded above, SU-F-B-G is the most accurate version for spatial unmixing. To confirm this advantage, it is necessary to compare this version with other prevailing methods. In this paper, the well-known STARFM and FSDAF methods were used as benchmark methods. For visual comparison, the predictions of all four regions are provided, as shown in Figs. 11-14, with one zoomed sub-area in each case to facilitate the comparison. In Fig. 11, the results predicted by STARFM and FSDAF present obvious speckle noise and the light pink object in the left part of the sub-area is inappropriately predicted as dark orange. With respect to UBDF-F-B-G, the predicted light pink object is the closest to the reference, but the blue object in the right part of the sub-area in the reference is predicted incorrectly. Amongst all five methods, VIPSTF-SU-F-B-G performs most satisfactorily in restoring the Landsat image. For predictions of the second heterogeneous region shown in Fig. 12, the hue is darker in the STARFM and FSDAF results compared to the reference as a whole, which can also be observed clearly from the green objects in the center of the zoomed sub-area. The prediction in Fig. 12(c) fails to reproduce the spatial variation.

Fig. 12(d) presents more spatial variation but, simultaneously, obvious spectral deviation. Fig. 12(e) shows the most accurate result. For the region with land cover changes shown in Fig. 13, the color and spatial texture of the flood in Fig. 13(e) are closer to the reference in Fig. 13(f) than the results of the other four methods. With respect to the predictions of the homogeneous region, UBDF-F-B-G produces less accurate prediction than the other four methods. Both STDFA-F-B-G and VPSTF-SU-F-B-G reproduce the green object more satisfactorily, as shown in the results of the sub-area.

Quantitative assessment of the five methods is shown in Table 1. Overall, VIPSTF-SU-F-B-G is always the most accurate version amongst all three SU-F-B-G versions. Moreover, its accuracy is greater than STARFM and FSDAF, especially for the two heterogeneous regions. In the first heterogeneous region, the CC value of VIPSTF-SU-F-B-G is 0.0315 larger than STARFM and 0.0197 larger than FSDAF. Compared with STDFA-F-B-G, the increase in CC is 0.0286. With respect to the second heterogeneous region, for VIPSTF-SU-F-B-G, its CC value is 0.0523 and 0.0252 larger than STARFM and FSDAF, respectively. Meanwhile, the corresponding RMSE value is 0.0097 and 0.0043 smaller. For

the region with land cover changes, VIPSTF-SU-F-B-G and STARFM produce comparable results with greater accuracy than the other three methods. This is because spatial unmixing-based methods assume the stable changes in land cover and becomes more challenging when the land cover in the region changes temporally. However, the large increase in accuracy over the original SU versions in this case suggests that VIPSTF-SU-F-B-G overcomes the limitation of spatial unmixing in predicting land cover changes and reflects its potential in broader applications. In the homogeneous region, except UBDF-F-B-G, the accuracy of all methods is very close and VIPSTF-SU-F-B-G is slightly more accurate than the other

methods. This is because spatio-temporal fusion involves less uncertainty for homogeneous regions, and the accuracies of most existing methods are great.

F. The computational cost of the methods

Table 2 lists the computational costs of all the methods for the first heterogeneous region and region with land cover changes (in units of seconds)

	First heterogeneous region			Region with land cover changes		
	UBDF-based	STDFA-based	VIPSTF-SU-based	UBDF-based	STDFA-based	VIPSTF-SU-based
SU	81.1	67.6	74.2	134.4	97.1	92.3
SU-GW	113.7	115.6	137.7	178.5	160.7	130.1
SU-BR	4127.7	2471.4	2494.0	7298.3	5943.7	5548.3
SU-B-G	5319.3	2511.7	2494.5	7264.9	3531.3	3889.6
SU-FCM	107.0	122.8	96.2	113.7	124.4	122.1
SU-F-G	125.1	106.6	105.5	143.4	110.9	124.9
SU-F-B	840.2	567.8	465.4	1273.7	496.2	1272.1
SU-F-B-G	735.9	478.9	428.1	1824.8	1187.4	1435.3

IV. DISCUSSION

A. Applicability of the GW scheme

The ability of the GW scheme to restore more reliable spectral information and alleviate the block effect was demonstrated in the experiments, as reported in Section III-C. In this section, we explain the mechanism behind this result further. When solving Eq. (2), a local window is used for estimation of the class reflectance, based on the assumption that the reflectance of the same class is also the same. This can be challenging for neighbors further away from the target center pixel (e.g., pixels at the border of the local window) for two reasons. First, the spatial association between them and the center pixel is weaker as the reflectance of the same class can differ. Second, some of the pixels at the border of the current local window are excluded in the local window for the next target center pixel. The difference between the two local windows for the two adjacent target center pixels can lead to obvious deviations in reflectance, that is, the block effect [51]. Using the GW scheme proposed in this paper, smaller weights are assigned to these spatially distant neighbors to weaken their influence in spatial unmixing. Meanwhile, the weights of neighbors closer to the center pixel (e.g., common pixels of the two local windows for the two adjacent target pixels) are larger, emphasizing the relationship of the two adjacent target pixels. This is beneficial for alleviating the block effect, and further, reproducing more reliable spectral information for spatial unmixing-based spatio-temporal fusion. It should be stressed, however, this GW-based blocks-removed scheme is totally different from SU-BR introduced in Section II-B. In SU-BR, a new term is included to account for the spatial continuity explicitly. The two schemes are not in conflict, and can be integrated for more reliable spatial unmixing, as is done by the SU-B-G version.

The framework provided by the GW scheme is applicable to and can be integrated with the FCM and BR schemes. Generally,

it can also be applied to any regularized spatial unmixing model without the need for any *prior* information, as the GW scheme modifies only the date fidelity term and is compatible with any additional constraint terms. Thus, GW can be regarded as a general solution to enhance the regularized spatial unmixing models. The integration of GW and the potential new constraints would be an interesting avenue for future research.

B. Influence of heterogeneity on the GW scheme

The experimental results reflect the sensitivity of the GW scheme to regions with different heterogeneity. This is especially obvious for the homogeneous region, where increase in accuracy of the GW scheme is limited. Moreover, the increase in accuracy for the second heterogeneous region (Region 2 hereafter) is obviously smaller than for the first heterogeneous region (Region 1 hereafter). To reveal the influence of heterogeneity quantitatively, an index is developed and the two heterogeneous regions are considered here.

As spatial unmixing is performed for coarse pixels, we need to quantify the heterogeneity at the corresponding spatial resolution. Here, we introduce the heterogeneity index, denoted as H , which was proposed by Wang *et al.* [51] to evaluate the deviation in class reflectances between the target coarse pixel and the neighboring pixels. It is calculated with the aid of the reference fine image. The size of the neighborhood is set to the same as the local window size in the unmixing model in Eq. (2). The heterogeneity index is calculated for each coarse pixel as

$$H_i = \frac{\sum_{c=1}^C \sum_{j=1}^N \left[I_{i,j,c} (\bar{E}_{i,c} - \bar{E}_{j,c}) \right]^2}{\sum_{c=1}^C \sum_{j=1}^N I_{i,j,c}} \quad (12)$$

where

$$\bar{E}_{i,c} = \sum_{m=1}^{M_{i,c}} e_{m,c}^{(i)} / M_{i,c} \quad (13)$$

In Eqs. (12) and (13), $M_{i,c}$ is the number of fine pixels belonging to class c in the coarse pixel at \mathbf{X}_i and $e_{m,c}^{(i)}$ denotes the reflectance of the m -th fine pixel for class c in the coarse pixel. $\bar{E}_{i,c}$ and $\bar{E}_{j,c}$ are the simulated reflectances of class c for the center coarse pixel at \mathbf{X}_i and neighboring coarse pixel at \mathbf{X}_j , which are generated by averaging all $e_{m,c}^{(i)}$ or $e_{m,c}^{(j)}$ within the corresponding coarse pixels. The meaning of the other variables is the same as in Eq. (4). For each band, the values for all coarse pixels calculated based on Eq. (12) are averaged to produce the statistical index H . A larger H indicates greater heterogeneity and *vice versa*. The H values and corresponding increases in CC from VIPSTF-SU to VIPSTF-SU-GW for the two heterogeneous regions are listed in Table 3.

Table 3 The values of the heterogeneity index H and corresponding increases in CC from VIPSTF-SU to VIPSTF-SU-GW for each band

Band	Region 1		Region 2	
	$H(\times 10^{-4})$	Increase in CC	$H(\times 10^{-4})$	Increase in CC
Blue	1.13	0.0191	0.87	0.0019
Green	1.84	0.0179	1.75	0.0043
Red	2.66	0.0208	2.67	0.0009
NIR	10	0.0067	21	0.0157
SWIR1	7.34	0.0133	8.08	-0.0015
SWIR2	8.93	0.0106	5.21	-0.0038

For the blue, green and SWIR2 bands, the H values of Region 1 are larger than those of Region 2. Correspondingly, the increases in CC are larger than those for Region 2. In addition, focusing on the NIR band, the H value of Region 1 is considerably smaller than for Region 2 and the increase in CC is smaller accordingly. Overall, Region 1 presents greater heterogeneity and the increase in accuracy is larger. Thus, it can be concluded that the proposed GW scheme is more beneficial for regions with greater heterogeneity.

C. Difference between GW and GWR

The concept of GW is originated from the well-known geographical weighting regression (GWR) in spatial statistics. SU-GW and GWR both aim to identify a fitting model between dependent and independent variables at the same geographical locations based on the available observations. Moreover, they both use a weighting function to quantify the influence of independent variables in a local window. However, the GW scheme in this paper is a specific case of GWR and differs from GWR mainly in two aspects. First, GWR is a typical fitting-then-prediction process. The parameters estimated in the fitting process are used to predict dependent variables at other locations based on the corresponding observed independent variables. In SU-GW, however, the estimated reflectance is our final goal in spatial unmixing. Second, SU-GW aims to estimate the parameters in terms of reflectance in this paper, whose value is constrained to be within (0, 1) to guarantee the physical meaning. The estimation in GWR is generally not constrained.

D. Comparison between the FCM and BR schemes

The advantage of the FCM scheme to recover intra-class variation was validated consistently in the experiments. In the FCM-based spatial unmixing methods, the prediction of each fine pixel is composed of the class reflectance of more than one

class with corresponding membership. This scheme is more flexible in relation to handling the error propagated from classification than the non-FCM scheme that relates each fine pixel strictly to only one class. The basic assumption of SU-BR is that neighboring coarse pixels share similar reflectances for the same class. The spatial unmixing results are also enhanced noticeably. In most cases in the experiments, the enhancement by FCM tends to be more obvious than for BR. This can be seen clearly from the visual inspection in Figs. 6-8. The accuracy increase for FCM is generally larger, as shown in Fig. 10. However, we can also observe clearly that both FCM and BR can compensate for each other, and the accuracy is further increased by integrating both aspects (i.e., the SU-F-B version).

E. Uncertainty introduced by abrupt changes in defining the optimal number of clusters

We proposed to apply the XB index to define the optimal number of clusters. As there is normally very limited fine spatial resolution information for the prediction time, the XB index was calculated based on the known fine images following the conventional assumption in spatial unmixing (i.e., there is no land cover change between the known and prediction times). For the case with abrupt changes, such as the flood inundation in the region with land cover changes, uncertainty exists inevitably in the number of clusters determination, as new classes are involved relative to the known fine image. On the other hand, the coarse proportion synthesized from the classified land cover map is not reliable, no matter how many clusters are defined. That is, the uncertainty caused by land cover changes cannot be eliminated. Nevertheless, it should be pointed out that abrupt changes occur only for a small-sized area in the entire study region. In spatial unmixing, each coarse pixel is decomposed separately based on a local proportion matrix and the errors for the pixels in the small-sized changed sub-area do not propagate to the others. That is, the prediction for the remaining unchanged area is not affected. Therefore, the XB method for defining the number of clusters directly from the known fine images is an acceptable choice in most cases, especially when the abrupt changes are sufficiently limited.

V. CONCLUSION

In this paper, a geographically weighted spatial unmixing model (SU-GW) was proposed to quantify the influence of coarse neighbors in spatial unmixing more accurately. Moreover, the XB index was applied to determine automatically the optimal number of clusters for classification of the known fine spatial resolution image. SU-GW can be integrated conveniently with existing spatial unmixing models, and the extended versions based on SU-BR and SU-FCM were considered in the paper. By integrating one or more schemes in SU-GW, SU-BR and SU-FCM with three typical spatial unmixing methods (i.e., UBDF, STFDA, and VIPSTF-SU), a total number of 24 schemes (12 existing non-GW versions and 12 proposed GW-based versions) were generated. Experiments on two heterogeneous regions, one region with land cover changes and one homogeneous region were performed and the 24 versions were compared systematically. The key findings are summarized as follows.

- 1) SU-GW is an effective spatial unmixing method to increase the accuracy of spatio-temporal fusion, and the increase in accuracy is influenced by the heterogeneity of the landscape.
- 2) The SU-GW scheme is complementary with the other two schemes (i.e., SU-FCM and SU-BR). The accuracy of the latter can be increased by integrating SU-GW.
- 3) All 12 existing non-GW-based versions can be enhanced by integrating the SU-GW scheme.
- 4) The accuracy of VIPSTF-SU-F-B-G is the greatest amongst all 24 methods, and it is also comparable to or even greater than the prevailing STARFM and FSDAF methods, especially for the heterogeneous regions.
- 5) The XB index is effective in determining the number of clusters.

REFERENCES

- [1] J. B. Campbell, R. H. Wynne, *Introduction to remote sensing*, The Guilford Press, 2011.
- [2] B. Chen, B. Huang, B. Xu, "Multi-source remotely sensed data fusion for improving land cover classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 124, pp. 27-39, 2017.
- [3] M. Belgiu, A. Stein, "Spatiotemporal image fusion in remote sensing," *Remote Sensing*, vol. 11, no. 7, pp. 818, 2019.
- [4] B. Chen, B. Huang, "Comparison of spatiotemporal fusion models: a review," *Remote Sensing*, vol. 7, no. 2, pp. 1798-1835, 2015.
- [5] F. Gao, T. Hilker, X. Zhu, M. Anderson , J. Masek, P. Wang, Y. Yang, "Fusing Landsat and MODIS data for vegetation monitoring," *IEEE Geoscience and Remote Sensing Magazine*, vol. 3, pp. 47-60, 2015.
- [6] H. Zhang , B. Huang, M. Zhang, K. Cao, L. Yu, "A generalization of spatial and temporal fusion methods for remotely sensed surface parameters," *International Journal of Remote Sensing*, vol. 36, pp. 4411-4445, 2015.
- [7] J. Zhou, J. Chen, X. Chen, X. Zhu, Y. Qiu, H. Song, Y. Rao, C. Zhang, X. Cao, X. Cui, "Sensitivity of six typical spatiotemporal fusion methods to different influential factors: A comparative study for a normalized difference vegetation index time series reconstruction," *Remote Sensing of Environment*, vol. 252, pp. 112130, 2021.
- [8] X. Zhu, F. Cai, J. Tian, T.K.-A. Williams, "Spatiotemporal fusion of multisource remote sensing data: literature survey, taxonomy, principles, applications, and future directions," *Remote Sensing*, vol. 10, no. 4, pp. 527, 2018.
- [9] Y. Tang, Q. Wang, K. Zhang, P. M. Atkinson, "Quantifying the effect of registration error on spatio-temporal fusion," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 487–503, 2020.
- [10] S. Delalieux, P. Zarco-Tejada, L. Tits, M. Jiménez-Bello, D. Intrigliolo, B. Somers, "Unmixing-based fusion of hyperspatial and hyperspectral airborne imagery for early detection of vegetation stress," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, pp. 2571-2582, 2014.
- [11] K. Jia, S. Liang, Z. Lei, X. Wei, Y. Yao, X. Xie, "Forest cover classification using Landsat ETM plus data and time series MODIS NDVI data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 33, pp. 32-38, 2014.
- [12] J. Walker, K. de Beurs, R. Wynne, F. Gao, "Evaluation of Landsat and MODIS data fusion products for analysis of dryland forest phenology," *Remote Sensing of Environment*, vol. 117, pp. 381-393, 2011.
- [13] R. Zurita-Milla, G. Kaiser, J.G.P.W. Clevers, W. Schneider, M. E. Schaepman, "Downscaling time series of MERIS full resolution data to monitor vegetation seasonal dynamics," *Remote Sensing of Environment*, vol. 113, pp. 1874-1885, 2009.
- [14] B. Huang, J. Wang, H. Song, D. Fu, K. Wong, "Generating high spatiotemporal resolution land surface temperature for urban heat island monitoring," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 5, pp. 1011-1015, 2013.
- [15] J. Wang, O. Schmitz, M. Lu, D. Karssenberg, "Thermal unmixing based downscaling for fine resolution diurnal land surface temperature analysis," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 161, pp. 76–89, 2020.
- [16] H. Shen, L. Huang, L. Zhang, W. Penghai, C. Zeng, "Long-term and fine-scale satellite monitoring of the urban heat island effect by the fusion of multi-temporal and multi-sensor remote sensed data: a 26-year case study of the city of Wuhan in China," *Remote Sensing of Environment*, vol. 172, pp.109-125, 2016.
- [17] Q. Weng, P. Fu, F. Gao, "Generating daily land surface temperature at Landsat resolution by fusing Landsat and MODIS data," *Remote Sensing of Environment*, vol. 145, pp. 55-67, 2014.
- [18] P. Wu, H. Shen, L. Zhang, F.-M. Göttsche, "Integrated fusion of multi-scale polar-orbiting and geostationary satellite observations for the mapping of high spatial and temporal resolution land surface temperature," *Remote Sensing of Environment*, vol. 156, pp. 169-181, 2015.
- [19] P. Li, Y. Ke, D. Wang, H. Ji, S. Chen, M. Chen, M. Lyu, D. Zhou, "Human impact on suspended particulate matter in the Yellow River Estuary, China: evidence from remote sensing data fusion using an improved spatiotemporal fusion method," *Science of The Total Environment*, vol. 750, pp. 141612, 2020.
- [20] A. Tewes, F. Thonfeld, M. Schmidt, R. J. Oomen, X. Zhu, O. Dubovik, G. Menz, J. Schellberg, "Using RapidEye and MODIS data fusion to monitor vegetation dynamics in semi-arid rangelands in South Africa," *Remote Sensing*, vol. 7, no. 6, pp. 6510-6534, 2015.
- [21] R. Houborg, M. McCabe, F. Gao, "A spatio-temporal enhancement method for medium resolution LAI (STEM-LAI)," *International Journal of Applied Earth Observation and Geoinformation*, vol. 47, pp. 15-29, 2016.
- [22] J. Meng, X. Du, B. Wu, "Generation of high spatial and temporal resolution NDVI and its application in crop biomass estimation," *International Journal of Digital Earth*, vol. 6, no. 3, pp. 1-16, 2011.
- [23] B. Chen, L. Chen, B. Huang, R. Michishita, B. Xu, "Dynamic monitoring of the poyang lake wetland by integrating landsat and modis observations," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 139, pp. 75-87, 2018.
- [24] Zhang, J. Zhu, Liu, "Blending MODIS and Landsat images for urban flood mapping," *International Journal of Remote Sensing*, vol. 35, no. 9, pp. 3237-3253, 2014.
- [25] N. Bhattacharai, L. Quackenbush, M. Dougherty, L. Marzen, "A simple Landsat-MODIS fusion approach for monitoring seasonal evapotranspiration at 30 m spatial resolution," *International Journal of Remote Sensing*, vol. 36, no. 1, pp. 115-143, 2015.
- [26] R., Zurita-Milla, J. Clevers, M. E. Schaepman, "Unmixing-based Landsat TM and MERIS FR data fusion," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 3, pp. 453-457, 2008.
- [27] L. Busetto, M. Meroni, R. Colombo, "Combining medium and coarse spatial resolution satellite data to improve the estimation of sub-pixel NDVI time series," *Remote Sensing of Environment*, vol. 112, no. 1, pp. 118-131, 2008.
- [28] F. Maselli, "Definition of spatially variable spectral endmembers by locally calibrated multivariate regression analysis," *Remote Sensing of Environment*, vol. 75, pp. 29-38, 2001.
- [29] M. Wu, Z. Niu, C. Wang, C. Wu, L. Wang, "Use of MODIS and Landsat time series data to generate high-resolution temporal synthetic Landsat data using a spatial and temporal reflectance fusion model," *Journal of Applied Remote Sensing*, vol. 6, no. 1, pp. 063507, 2012.
- [30] Q. Wang, Y. Tang, X. Tong, P. M. Atkinson, "Virtual image pair-based spatio-temporal fusion," *Remote Sensing of Environment*, vol. 249, 112009, 2020.
- [31] M. Wu, C. Wu, W. Huang, Z. Niu, C. Wang, W. Li, P. Hao, "An improved high spatial and temporal data fusion approach for combining Landsat and MODIS data to generate daily synthetic Landsat imagery," *Information Fusion*, vol. 31, pp. 14-25, 2015.
- [32] Y. Xu, B. Huang, Y. Dr, K. Cao, C. Guo, D. Meng, "Spatial and temporal image fusion via regularized spatial unmixing," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 6, pp. 1362-1366, 2015.
- [33] F. Gao, J. Masek, M. Schwaller, F. Hall, "On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 8, pp. 2207-2218, 2006.
- [34] T. Hilker, M. A. Wulder, N. C. Coops, J. Linke, G. McDermid, J. G. Masek, F. Gao, J. C. White, "A new data fusion model for high spatial-and temporal-resolution mapping of forest disturbance based on Landsat and MODIS," *Remote Sensing of Environment*, vol. 113, no. 8, pp. 1613-1627, 2009.

- [35] X. Zhu, J. Chen, F. Gao, X. Chen, J. G. Masek, "An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions," *Remote Sensing of Environment*, vol. 114, no. 11, pp. 2610-2623, 2010.
- [36] Q. Wang, P. M. Atkinson, "Spatio-temporal fusion for daily Sentinel-2 images," *Remote Sensing of Environment*, vol. 204, pp. 31-42, 2018.
- [37] B. Huang and H. Song, "Spatiotemporal reflectance fusion via sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 10, pp. 3707-3716, 2012.
- [38] H. Song and B. Huang, "Spatiotemporal satellite image fusion through one-pair image learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 4, pp. 1883-1896, 2013.
- [39] B. Wu, B. Huang, L. Zhang, "An error-bound-regularized sparse coding for spatiotemporal reflectance fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 12, pp. 6791-6803, 2015.
- [40] H. Song, Q. Liu, G. Wang, R. Hang and B. Huang, "Spatiotemporal satellite image fusion using deep convolutional neural networks," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 3, pp. 821-829, 2018.
- [41] V. Moosavi, A. Talebi, M. H. Mokhtari, S. R. F. Shamsi, Y. Niazi, "A wavelet-artificial intelligence fusion approach (WAIFA) for blending Landsat and MODIS surface temperature," *Remote Sensing of Environment*, vol. 169, pp. 243-254, 2015.
- [42] X. Liu, C. Deng, J. Chanussot, D. Hong and B. Zhao, "StfNet: a two-stream convolutional neural network for spatiotemporal image fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 6552-6564, 2019.
- [43] X. Zhu, E. H. Helmer, F. Gao, D. Liu, J. Chen, M. A. Lefsky, "A flexible spatiotemporal method for fusing satellite images with different resolutions," *Remote Sensing of Environment*, vol. 172, pp. 165-177, 2016.
- [44] M. Liu, W. Yang, X. Zhu, J. Chen, X. Chen, L. Yang, E. H. Helmer, "An Improved Flexible Spatiotemporal DAta Fusion (IFSDAF) method for producing high spatiotemporal resolution normalized difference vegetation index time series," *Remote Sensing of Environment*, vol. 227, pp. 74-89, 2019.
- [45] X. Li, G. M. Foody, D. S. Boyd, Y. Ge, Y. Zhang, Y. Du, F. Ling, "SFSDAF: An enhanced FSDAF that incorporates sub-pixel class fraction change information for spatio-temporal image fusion," *Remote Sensing of Environment*, vol. 237, pp. 111537, 2020.
- [46] C. Gevaert, F. Garc ía-Haro, "A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion," *Remote Sensing of Environment*, vol. 156, pp. 34-44, 2015.
- [47] B. Zhukov, D. Oertel, F. Lanzl, G. Reinhard, "Unmixing-based multisensor multiresolution image fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3, pp. 1212-1226, 1999.
- [48] F. Zhou, D. Zhong, "Kalman filter method for generating time-series synthetic Landsat images and their uncertainty from Landsat and MODIS observations," *Remote Sensing of Environment*, vol. 239, pp. 111628, 2020.
- [49] J. Amorós-López, L. Gómez-Chova, L. Alonso, L. Guanter, R. Zurita-Milla, J. Moreno, G. Camps-Valls, "Multitemporal fusion of Landsat/TM and ENVISAT/MERIS for crop monitoring," *International Journal of Applied Earth Observation and Geoinformation*, vol. 23, pp. 132-141, 2013.
- [50] F. D. Van der Meer, X. Jia, "Collinearity and orthogonality of endmembers in linear spectral unmixing," *International Journal of Applied Earth Observation and Geoinformation*, vol. 18, no. 1, pp. 491-503, 2012.
- [51] Q. Wang, K. Peng, Y. Tang, X. Tong, P. M. Atkinson, "Blocks-removed spatial unmixing for downscaling MODIS images," *Remote Sensing of Environment*, vol. 256, pp. 112325, 2021.
- [52] A. Kashki, M. Karimi, R. Zandi, Z. Roki, "Evaluation of the effect of geographical parameters on the formation of the land surface temperature by applying OLS and GWR, A case study Shiraz City, Iran," *Urban Climate*, vol. 37, pp. 100832, 2021.
- [53] H. Yu, H. Gong, B. Chen, K. Liu, M. Gao, "Analysis of the influence of groundwater on land subsidence in Beijing based on the geographical weighted regression (GWR) model," *Science of The Total Environment*, vol. 738, pp. 139405, 2020.
- [54] Q. Wang, L. Wang, C. Wei, Y. Jin, Z. Li, X. Tong, P. M. Atkinson, "Filling gaps in Landsat ETM+ SLC-off images with Sentinel-2 images," *International Journal of Applied Earth Observation and Geoinformation*, vol. 101, pp. 102365, 2021.
- [55] W. Tobler, "A computer movie simulating urban growth in the Detroit region," *Economic Geography*, vol. 46, pp. 234-240, 1970.
- [56] X. Xie, G. Beni, "A validity measure for fuzzy clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, pp. 841-847, 1991.
- [57] Y. Zhang, Y. Du, X. Li, S. Fang, F. Ling, "Unsupervised subpixel mapping of remotely sensed imagery based on fuzzy c-means clustering approach," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, pp. 1024-1028, 2014.
- [58] Q. Wang, X. Ding, X. Tong, P. M. Atkinson, "Spatio-temporal spectral unmixing of time-series images," *Remote Sensing of Environment*, vol. 259, pp. 112407, 2021.
- [59] G. M. Foody, "Geographical weighting as a further refinement to regression modelling: an example focused on the NDVI-rainfall relationship," *Remote Sensing of Environment*, vol. 88, pp. 283-293, 2003.
- [60] T. Ranchin, L. Wald, "Fusion of high spatial and spectral resolution images: the ARSIS concept and its implementation," *Photogrammetric Engineering & Remote Sensing*, vol. 66, pp. 49-61, 2000.
- [61] Z. Wang, A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81-84, 2002.
- [62] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G. A. Licciardi, R. Restaino, and L. Wald, "A critical comparison among pansharpening algorithms," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2565-2586, 2015.



Kaidi Peng is currently pursuing the B.S. degree with Tongji University, Shanghai, China. Her research interests focus on remote sensing image fusion.



Qunming Wang received the Ph.D. degree from the Hong Kong Polytechnic University, Hong Kong, in 2015.

He is currently a Professor with the College of Surveying and Geo-Informatics, Tongji University, Shanghai, China. He was a Lecturer (Assistant Professor) with Lancaster Environment Centre, Lancaster University, Lancaster, U.K., from 2017 to 2018. His 3-year Ph.D. study was supported by the hypercompetitive Hong Kong Ph.D. Fellowship and his Ph.D. thesis was awarded as the Outstanding Thesis in the Faculty. He has authored or coauthored 60 peer-reviewed articles in international journals such as *Remote Sensing of Environment*, *IEEE Transactions on Geoscience and Remote Sensing* and *ISPRS Journal of Photogrammetry and Remote Sensing*. His research interests include remote sensing, image processing and geostatistics.

Dr. Wang serves as Associate Editor for *Science of Remote Sensing* (sister journal of *Remote Sensing of Environment*) and *Photogrammetric Engineering & Remote Sensing*, and was Associate Editor for *Computers and Geosciences* (2017–2020).



Yijie Tang received the B.S. degree from Nanjing Normal University, Nanjing, China, in 2019. She is currently pursuing the Ph.D. degree in Tongji University, Shanghai, China. Her research interests focus on remote sensing image fusion.



Xiaohua Tong received the Ph.D. degree in traffic engineering from Tongji University, Shanghai, China, in 1999.

He is currently a Professor with the College of Surveying and Geoinformatics, Tongji University. He was a Research Fellow with Hong Kong Polytechnic University, Hong Kong, in 2006, and a Visiting Scholar with the University of California, Santa Barbara, CA, USA, between 2008 and 2009. His research interests include remote sensing, geographic information system, uncertainty and spatial data quality, and image processing for high-resolution and hyperspectral images.



Peter M. Atkinson received the Ph.D. degree from the University of Sheffield (NERC CASE award with Rothamsted Experimental Station) in 1990. More recently, he received the MBA degree from the University of Southampton in 2012.

He is currently Distinguished Professor of Spatial Data Science and Dean of the Faculty of Science and Technology at Lancaster University, UK. He was previously Professor of Geography at the University Southampton, where he is currently Visiting Professor. He is also Visiting Professor at the Chinese Academy of Sciences, Beijing. He previously held the Belle van Zuylen Chair at Utrecht University, the Netherlands, is a recipient of the Peter Burrough Award of the International Spatial Accuracy Research Association and is a Fellow of the Learned Society of Wales. The main focus of his research is in remote sensing, geographical information science and spatial (and space-time) statistics applied to a range of environmental science and socio-economic problems. He has published over 300 peer-reviewed articles in international scientific journals and around 50 refereed book chapters. He has also edited nine journal special issues and eight books.

Professor Atkinson is Editor-in-Chief of *Science of Remote Sensing*, a sister journal of *Remote Sensing of Environment*. He also sits on the editorial boards of several further journals including *Geographical Analysis*, *Spatial Statistics*, *International Journal of Applied Earth Observation and Geoinformation*, and *Environmental Informatics*.