SFSDAF: an enhanced FSDAF that incorporates sub-pixel class fraction change information for spatio-temporal image fusion

Xiaodong Li[a], Giles M. Foody[b], Doreen S. Boyd[b], Yong Ge[c], Yihang Zhang[a], Yun Du[a], and Feng Ling[*a]

[a] *Key Laboratory for Environment and Disaster Monitoring and Evaluation, Hubei, Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan 430077, China*

[b] *School of Geography, University of Nottingham, University Park, Nottingham NG7 2RD, U.K.*

[c] *State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences & Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China*

E-mail address: lixiaodong@whigg.ac.cn (X. Li), lingf@whigg.ac.cn (F. Ling)

Abstract

Spatio-temporal image fusion methods have become a popular means to produce remotely sensed data sets that have both fine spatial and temporal resolution. Accurate prediction of reflectance change is difficult, especially when the change is caused by both phenological change and land cover class changes. Although several spatio-temporal fusion methods such as the Flexible Spatiotemporal DAta Fusion (FSDAF) directly derive land cover phenological change information (such as endmember change) at different dates, the direct derivation of land cover class change information is challenging. In this paper, an enhanced FSDAF that incorporates sub-pixel class fraction change information (SFSDAF) is proposed. By directly deriving the sub-pixel land cover class fraction change information the proposed method allows accurate prediction even for heterogeneous regions that undergo a land cover class change. In particular, SFSDAF directly derives fine spatial resolution endmember change and class fraction change at the date of the observed image pair and the date of prediction, which can help identify image reflectance change resulting from different sources. SFSDAF predicts a fine resolution image at the time of acquisition of coarse resolution images using only one prior coarse and fine resolution image pair, and accommodates variations in reflectance due to both natural fluctuations in class spectral response (e.g. due to phenology) and land cover class change. The method is illustrated using degraded and real images and compared against three established spatio-temporal methods. The results show that the SFSDAF produced the least blurred images and the most accurate predictions of fine resolution reflectance values, especially for regions of heterogeneous landscape and regions that undergo some land cover class change. Consequently, the SFSDAF has considerable potential in monitoring Earth surface dynamics.

Keywords: spatio-temporal image fusion, land cover class fraction, FSDAF.

1. Introduction

Land cover change is an important environmental variable having, for example, impacts greater than climate change (Chapin et al. 2000; Foley et al. 2005; Vitousek et al. 1997). Remote sensing as a key land cover data source is often constrained by factors such as the spatial and temporal resolution of the available imagery. Detailed land cover monitoring requires imagery with both fine spatial and temporal resolution. Unfortunately, there is often a trade-off between these resolutions. For example, remotely sensed imagery such as that acquired by the Moderate Resolution Imaging Spectroradiometer (MODIS) offer the potential to study land cover at a daily frequency but only at a 250+m spatial resolution. Remotely sensed imagery such as that acquired by Landsat sensors can be acquired at a finer spatial resolution (typically ~30m) but with a 16 days temporal resolution. One way to address this problem includes the direct combination of multiple satellites. However, a large proportion of medium spatial resolution imagery is contaminated by cloud (Ju and Roy 2008), and differences in sensor properties (e.g. spectral wavebands) may impact upon reflectance values and complicate analyses.

An alternative approach that makes use of available data sets is to fuse imagery that have a fine temporal but coarse spatial resolution with imagery that have a fine spatial but coarse temporal resolution based on spatio-temporal image fusion methods. Such approaches exploit the positive attributes of each data set to form the desired time series of fine resolution images. The aim of spatio-temporal image fusion is the generation of a fine spatial resolution (FR) image for the date represented by a coarse spatial resolution (CR) image, referred to here as the prediction date. This is achieved by integrating the spatial and temporal information in a pair of FR and CR images of the same region acquired at other dates (Zhu et al. 2018). A key challenge to the analysis is that the latent FR image to be predicted may differ from other FR images observed at other dates. These differences in image reflectance arise for two main

reasons. Firstly, reflectance changes may be associated with plant phenology or growth as well as issues such as seasonal changes in solar elevation. As a result, the typical spectral response of a class represented by the class endmember statistics may vary in time. This type of change is typically relatively gradual. Secondly, the reflectance of a location may change over time because of an alteration of the land cover class. This reflectance change caused by land cover class change is often relatively abrupt. In many situations, the reflectance may change in time due to both endmember change and land cover class change.

Addressing reflectance change effectively is a critical issue in spatio-temporal image fusion. Various algorithms have been proposed including the spatial and temporal adaptive reflectance fusion model (STARFM) (Gao et al. 2006) and its variants (Fu et al. 2013; Wang et al. 2017; Zhu et al. 2010), as well as the unmixing-based method (Zhukov et al. 1999) and its variants (Amoros-Lopez et al. 2013; Gevaert and Javier Garcia-Haro 2015; Wu et al. 2012; Zurita-Milla et al. 2008; Zurita-Milla et al. 2009). Although these fusion methods can accommodate endmember change and have been successfully applied in the fields such as forest phenology analysis (Gaertner et al. 2016; Schmidt et al. 2015; Walker et al. 2012; Zurita-Milla et al. 2009) and crop phenology analysis (Amoros-Lopez et al. 2013; Gao et al. 2017), the prediction of relatively abrupt reflectance changes caused by land cover class change is still challenging (Zhao et al. 2018; Zhu et al. 2018; Zhu et al. 2016).

One way to allow accurate reflectance prediction when an abrupt reflectance change due to a land cover conversion has occurred is to use two pairs of FR and CR images, one acquired before and the other after the date of prediction (sometimes referred to as the two-pairs case). The two-pairs case fusion can be divided into change detection-based fusion and learning-based fusion. The change detection-based fusion methods derive FR land cover class change information using the input FR image pairs which are supposed to contain the land cover change information at the prediction date which lies between them

and predict different reflectance changes separately (Amoros-Lopez et al. 2013; Hilker et al. 2009). In some situations, obtaining two FR images separated by a short period is difficult and the two-pairs case methods can become unsuitable if there is an abrupt land cover class change within the relevant timeframe. For instance, an area may be inundated by a flood on the prediction date but not be flooded in the two FR images used. In this situation, the two-pairs case methods that detect the FR land cover class change information by comparing the FR image pairs become unsuitable for use. Since the CR image at the prediction date may capture land cover change, it is used in comparison with the input CR images that pre- and post-date it to improve land cover change detection in the fusion (Huang and Zhang 2014; Zhong and Zhou 2019). The learning-based fusion methods do not distinguish between changed and unchanged land covers, and predict abrupt and gradual reflectance changes in a unified framework. The learning-based fusion methods learn the complex relationship between the CR-FR original or difference image pairs to predict the unknown FR image based on algorithms such as dictionary learning (Huang and Song 2012; Wu et al. 2015) and deep learning (Liu et al. 2019b; Song et al. 2018). However, the aforementioned fusion methods require two FR and CR image pairs, which are acquired before and after the date of prediction, and thus are unsuitable for near-real-time prediction. In practice, timely updating of FR images with fine temporal but coarse spatial resolution images plays a key role in many applications including those focused on time-sensitive issues such as flooding and wild-fire monitoring. However, the two-pairs case methods require a FR image that post-dates the CR images at the prediction time as input. As a result, they are unsuitable for near-real-time updating of an image data series with a fine spatial and temporal resolution. This greatly impedes the realization of the full potential of satellite remote sensing to provide up-to-date land cover information and limits their value in studies of contemporary land cover change.

Considering the limitations of two-pairs case methods, the one-pair case spatio-temporal fusion method is then often necessary to predict abrupt change. Similar to the two-pairs case fusion to predict abrupt change, the one-pair case fusion can also be divided into change detection-based fusion and learning-based fusion. For the change detection-based one-pair case fusion, since the FR image at the prediction time is unknown, the FR land cover change information at different times is detected by comparing the reflectance at the location of each FR pixel from the CR image at the prediction time and the FR image pre- or post-dates it (Chen et al. 2018; Wang and Huang 2017). The aforementioned one-pair case fusion methods are at the pixel level, and Zhao et al. (2018) proposed a robust adaptive spatial and temporal image fusion model (RASTRM) which developed the fusion to feature level (i.e., land cover classes of interest) for abrupt change. RASTFM first detected reflectance change at a medium spatial resolution (coarser than the input FR image and finer than the input CR image) to derive land cover shape change (such as expansion of urban and shrinking of lake) and non-shape change (such as phenological change or crop rotation), and then predicted the reflectance changes in shape change and non-shape change regions separately. However, the change detection-based one-pair case fusion requires a threshold to detect land cover change which may lead to false detection.

An alternative one-pair case fusion for abrupt change is the learning-based fusion method. Song and Huang (2013) proposed a dictionary-pair learning fusion for the one-pair case (DPL-One). Unlike the dictionary-pair learning fusion which requires two pairs of CR-FR image pairs to learn the dictionaries as with the previous study of Huang and Song (2012), DPL-One directly downscales the original CR image to FR scale based on sparse representation, and then fuses this image with the input FR data based on high-pass modulation. DPL-One predicts reflectance more accurately for a pixel undergoing land cover class change than STARFM, but this method may not accurately predict the object shape if the CR

and FR images have a large difference in pixel size. DPL-One was further improved by using similar pixels to reduce the blurring effects (Chen et al. 2017).

The aforementioned one-pair fusion methods could predict abrupt land surface change, but they do not derive land cover information such as the land cover class or fractions present in the image, and do not explore how land cover phenological change and land cover class change affect surface reflectance change. Zhu et al. (2016) developed a Flexible Spatiotemporal DAta Fusion (FSDAF) method. FSFAD first directly estimate endmember changes to account for the land cover phenological change in a temporal prediction step. Then FSDAF spatially interpolates the CR at the prediction date, which could contain land cover change information, to FR scale in a spatial prediction step. Finally, FSDAF combines the temporal prediction image which contains land cover phenological change information and spatial prediction image which contains land cover class change information for a final prediction. FSDAF improves the fusion of reflectance change caused by both endmember and land cover class change compared with STARFM and unmixing-based methods. However, the FSDAF temporal prediction assumes that the land cover is unchanged in the temporal prediction step. Liu et al. (Liu et al. 2019a) proposed an improved FSDAF (IFSDAF) which generated more spatially continuous Normalized Difference Vegetation Index (NDVI) images by considering the spatial autocorrelation of NDVI. The land cover class change information contained in the reflectance imagery is not directly derived and used in IFSDAF.

Dealing with land cover class change effectively is the largest problem for the one-pair case spatio-temporal image fusion. The problem in the accommodation for land cover class change within spatio-temporal fusion methods is that the FR land cover information at the prediction time is, of course, unavailable. If accurate FR land cover information at the prediction date could be obtained, the FR land

cover change trajectory could be utilized in the fusion methods. Fortunately, there has been recently great progress in the FR land cover information derivation from CR imagery, especially with the spatio-temporal super-resolution land cover mapping (STSRLCM), which can predict fine spatial resolution land cover maps time series based on multiple CR images and one (Ling et al. 2011; Xu and Huang 2014) or a few (Li et al. 2017) FR land cover maps. Although STSRLCM has been successfully used in land cover change detection (Li et al. 2016) and mapping (Li et al. 2015; Wang et al. 2016), it assumes FR pixels are pure and produces thematic maps rather than reflectance images with continuous fields. The mixed pixel problem, in which a pixel may represent an area of multiple classes, also needs to be addressed at the FR scale to achieve the full potential of spatio-temporal image fusion especially for highly fragmented landscapes or because of issues of intra-class mixing (Keshava and Mustard 2002).

In this paper, a new spatio-temporal fusion method that accommodates sub-pixel class fraction change is proposed to address the combined effect of both endmember change and land cover class change. The proposed method is based on FSDAF considering its extensibility and wide usage in the fusion of spectral reflectance (Alves et al. 2018; Chen et al. 2017; Sun et al. 2019), NDVI (Chen et al. 2018; Liao et al. 2017; Liu et al. 2019a; Maselli et al. 2019), land surface temperature (Zhang et al. 2017), as well as land cover class fractions (Zhang et al. 2018). The proposed enhanced FSDAF that considered sub-pixel class fraction change information (SFSDAF) aims to extend FSDAF by not only directly deriving endmember change to represent phenological change, but also directly deriving sub-pixel land cover class fraction changes in the FR pixels to accommodate land cover class change and the presence of mixed pixels at the FR scale. Furthermore, SFSDAF is designed as a one-pair case spatio-temporal image fusion method, and thus has border applicability than two-pairs case fusion methods (e.g. it is appropriate for use in near-real-time applications). The proposed SFSDAF method is compared here with

a set of contemporary state-of-the-art spatio-temporal image fusion methods, STARFM, the unmixing-based data fusion (UBDF) (Zurita-Milla et al. 2008), and FSDAF, and the accuracy is assessed using degraded as well as real satellite sensor images to aid understanding of the method and also allow a rigorous assessment of its performance.

2. Methods

SFSDAF aims to estimate the FR image at the prediction date t_p using a FR image at date t_0, a CR image at t_0 and a CR image at t_p as input. The method has four steps that are highlighted in blue in Fig. 1. The first step is the estimation of the FR endmembers and class fractions at t_0. The second step is the estimation of the FR endmembers and class fractions at t_p. The third step is the temporal prediction of the FR image at t_p considering both endmember change and sub-pixel land cover class fraction change information from t_0 to t_p. The final step is the refinement of the temporal prediction image by using a spatial interpolation approach to predict the final FR prediction image at t_p. A flowchart of the proposed method is provided in Fig. 1 and the key details are introduced in the following sections. The variables are defined on first use but also a summary of the notation is provided in appendix A.

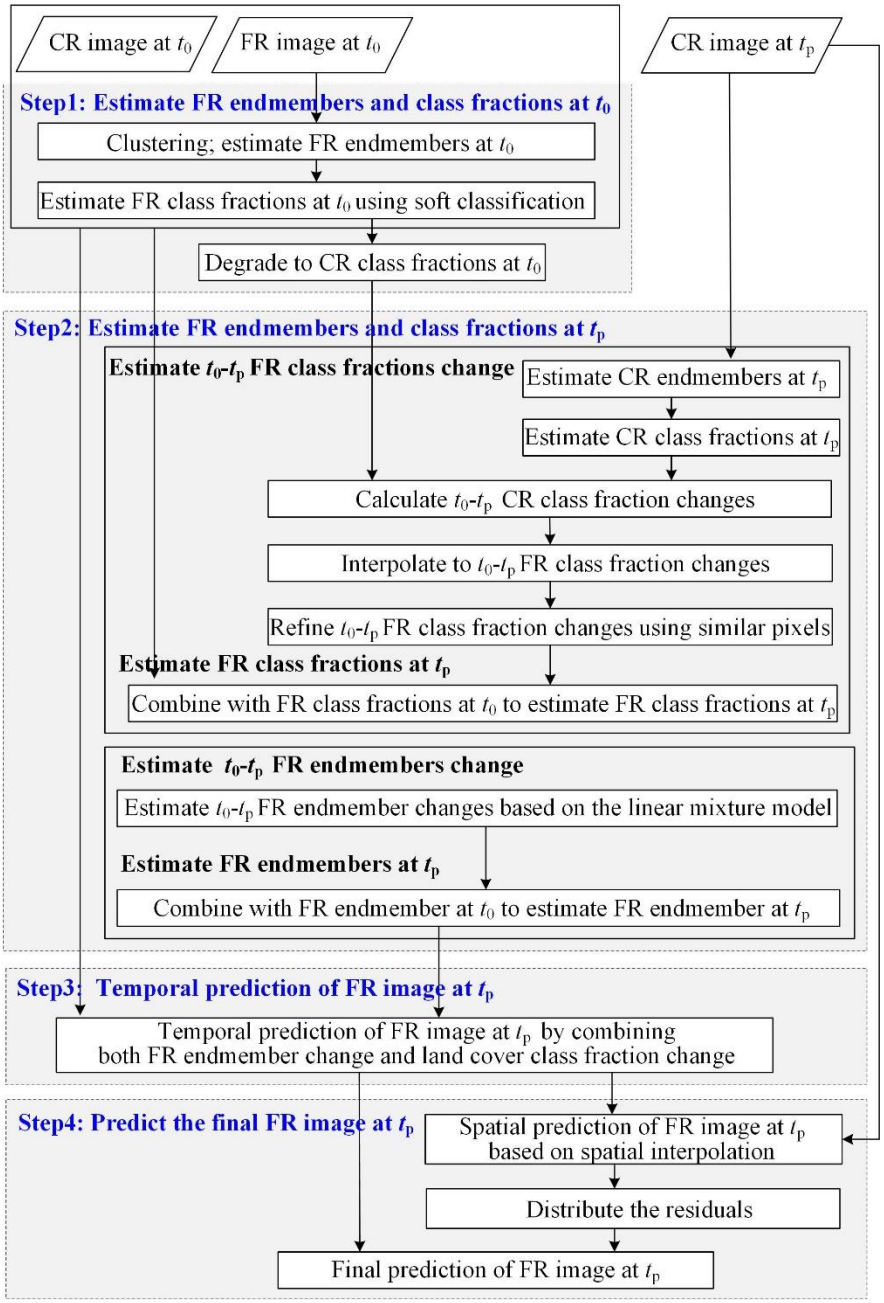

Fig. 1. Flowchart of the proposed SFSDAF. 'CR' stands for coarse spatial resolution, and 'FR' stands for fine spatial resolution.

2.1 Estimation of the FR endmembers and class fractions at t_0

2.1.1 Clustering and the estimation of FR endmembers at t_0

First, the FR image at t_0 is clustered into a FR land cover map using an unsupervised clustering algorithm such as *k*-means or ISODATA. Only the number of classes in the image needs to be defined,

and this may be informed by priori knowledge of the site or the observed heterogeneity. After producing the land cover map, the reflectance of each endmember is the average of the FR pixels from the FR image at t_0 according to the class type in the clustered map.

2.1.2 Estimation of FR class fraction images at t_0

The land cover map produced by the unsupervised classification is a hard classification map in which each FR pixel can belong to only one land cover class. In SFSDAF, with the aforementioned estimated endmembers at t_0, a soft classification is applied to FR image at t_0 to produce the FR land cover fraction images at t_0 to indicate sub-pixel scale land cover information (Alpaydin 1998). Assume the FR image at t_0 contains B spectral bands and is clustered with l land cover classes. $E^{FR}(c,b,t_0)$ is the b^{th} spectrum (b =1,…, B) in the c^{th} endmember (c =1,…, l) in the FR image at t_0. The FR class fraction or abundance is calculated as:

$$A^{FR}(x_{ij},y_{ij},t_0,c)=\frac{\left(\left\|\boldsymbol{v}(x_{ij},y_{ij},t_0)-\boldsymbol{\mu}(c,t_0)\right\|_{\Sigma}\right)^{-1}}{\sum_{c=1}^{l}\left(\left\|\boldsymbol{v}(x_{ij},y_{ij},t_0)-\boldsymbol{\mu}(c,t_0)\right\|_{\Sigma}\right)^{-1}} \quad (1)$$

where $\boldsymbol{v}(x_{ij},y_{ij},t_0)$ is the reflectance vector for FR pixel (x_{ij},y_{ij}) at t_0, and $\boldsymbol{\mu}(c,t_0)$ is the c^{th} cluster centroid at t_0.

With the FR land cover fraction images at t_0, the CR class fraction $A^{CR}(x_i,y_i,t_0,c)$ at t_0 is obtained by spatially degrading the FR class fractions of the c^{th} class within the CR pixel (x_i,y_i) :

$$A^{CR}(x_i,y_i,t_0,c)=\frac{1}{m}\sum_{j=1}^{m}A^{FR}(x_{ij},y_{ij},t_0,c) \quad (2)$$

where m is the number of FR pixels within one CR pixel.

2.2 Estimation of the FR endmembers and class fractions at t_p

The FR class fractions and endmembers at t_p are estimated in this step. Since the FR land cover class fractions at t_0 are estimated in the aforementioned step, if the FR land cover fraction changes from t_0 to

t_p can be estimated, the FR land cover class fractions at t_p can be calculated by combining FR land cover class fractions at t_0 with land cover fraction changes from t_0 to t_p. Similarly, if the FR endmember changes from t_0 to t_p can be estimated, the FR endmembers at t_p can be calculated by combining FR endmembers at t_0 with endmember changes from t_0 to t_p. The main steps in estimating the FR class fraction changes and endmember changes from t_0 to t_p are explained in the below.

2.2.1 Estimation of the FR class fraction change from t_0 to t_p

The CR class fractions at t_p are estimated and then compared with the CR class fractions at t_0 to produce the CR class fraction change images from t_0 to t_p. The CR class fraction change images are then downscaled to the FR class fraction change fractions from t_0 to t_p.

The CR land cover class fractions at t_p are generated first. Before generating the land cover class fractions, the CR endmembers are first estimated based on an inversion method of linear mixture equations as with the previous study (Li et al. 2016). First, n ($n>l$) CR pixels are selected according to the criteria used in (Zhu et al. 2016) to avoid the collinearity problem and reduce the impact of land cover change. Then, assuming the CR pixel value is a linear combination of the reflectance of all endmembers resident within it, the CR endmembers at t_p are estimated using the least square error (LSE) method:

$$\begin{bmatrix} R^{CR}(x_1,y_1,t_p,b) \\ \vdots \\ R^{CR}(x_i,y_i,t_p,b) \\ \vdots \\ R^{CR}(x_n,y_n,t_p,b) \end{bmatrix} = \begin{bmatrix} A^{CR}(x_1,y_1,t_0,1) & A^{CR}(x_1,y_1,t_0,2) & \cdots & A^{CR}(x_1,y_1,t_0,l) \\ \vdots & \vdots & & \vdots \\ A^{CR}(x_i,y_i,t_0,1) & A^{CR}(x_i,y_i,t_0,2) & \cdots & A^{CR}(x_i,y_i,t_0,l) \\ \vdots & \vdots & & \vdots \\ A^{CR}(x_n,y_n,t_0,1) & A^{CR}(x_n,y_n,t_0,2) & \cdots & A^{CR}(x_n,y_n,t_0,l) \end{bmatrix} \begin{bmatrix} E^{CR}(1,b,t_p) \\ \vdots \\ E^{CR}(c,b,t_p) \\ \vdots \\ E^{CR}(l,b,t_p) \end{bmatrix}. \quad (3)$$

where $R^{CR}(x_i,y_i,t_p,b)$ is the b^{th} band reflectance value in the CR pixel (x_i,y_i) at t_p, $E^{CR}(c,b,t_p)$ is b^{th} spectrum in the c^{th} endmember in the CR images at t_p. The c^{th} class fraction in the CR pixel (x_i,y_i) at t_p , i.e. $A^{CR}(x_i,y_i,t_p,c)$, is unknown at present, and can be substituted with $A^{CR}(x_i,y_i,t_0,c)$ in Eq. (3). Once the endmembers in the CR image at t_p are obtained, the land cover class fractions in the CR

image at t_p are obtained based on the fully constrained linear mixing model with 'non-negative' and 'sum-to-one' constraints for the generated class fractions (Li et al. 2016).

The CR class fraction change images from t_0 to t_p are then calculated. Assume $\Delta A^{CR}(x_i, y_i, c)$ is the c^{th} class fraction change at CR pixel (x_i, y_i), which is produced as:

$$\Delta A^{CR}(x_i, y_i, c) = A^{CR}(x_i, y_i, t_p, c) - A^{CR}(x_i, y_i, t_0, c) . \quad (4)$$

The estimated CR class fraction changes from t_0 to t_p are downscaled to FR scale using spatial interpolation methods, such as bicubic interpolation (Keys 1981) and thin plate spline (TPS) interpolation (Dubrule 1984). The bicubic and TPS spatial interpolations are based on spatial correlation assumption, and would predict smooth FR class fraction change images. Therefore, a similar pixels-based approach is used to refine the FR class fraction change images from spatial interpolation. It is assumed that similar pixels would have similar land cover fraction changes from t_0 to t_p in SFSDAF. The similar pixels, which have similar spectral reflectance in the FR image at t_0, are selected through a moving window using the same method in STARFM (Gao et al. 2006). The difference in reflectance between k^{th} FR pixel in the moving window and the target FR pixel is used in the selection of similar neighbor FR pixels. Only a selected number of FR pixels which have the smallest spectral difference to the target FR pixel in the FR pixel at t_0 are involved, and the weight of the k^{th} similar pixel, i.e. w_k, is calculated based on the relative spatial distance between the k^{th} FR similar pixel and the target FR pixel (x_{ij}, y_{ij}) as is defined in Gao et al. (2006).

With the weights of each similar FR pixel, the refined FR class fraction change for a target FR pixel (x_{ij}, y_{ij}) for the c^{th} class, i.e., $\Delta A^{FR}_{SI-Refine}(x_{ij}, y_{ij}, c)$, is calculated as:

$$\Delta A^{FR}_{SI-Refine}(x_{ij}, y_{ij}, c) = \sum_{k=1}^{N} w_k \times \Delta A^{FR}_{SI}(x_k, y_k, c) \quad (5)$$

where $\Delta A^{FR}_{SI}(x_k, y_k, c)$ is the downscaled c^{th} class fraction change at the k^{th} similar pixel from spatial

interpolation, and N is the number of similar pixels..

2.2.2 Estimation of the FR class fractions at t_p

The FR class fraction images at t_p are then combined by adding the FR class fraction images at t_0 and the FR class fraction change images from t_0 to t_p. For a FR pixel (x_{ij}, y_{ij}), the combined FR class fraction for the c^{th} class at t_p is calculated as:

$$A^{FR}_{Combination}(x_{ij}, y_{ij}, t_p, c) = A^{FR}(x_{ij}, y_{ij}, t_0, c) + \Delta A^{FR}_{SI-Refine}(x_{ij}, y_{ij}, c) . \quad (6)$$

The combined FR class fractions are viewed as an initial value because it may fail the 'non-negative' and 'sum-to-one' constraints. Therefore, all the negative fractions are set to 0 and $A^{FR}_{Combination}(x_{ij}, y_{ij}, t_p, c)$ is thus changed to $A^{FR}_{Non\text{-}negative}(x_{ij}, y_{ij}, c)$. Then the fractions are normalized as:

$$A^{FR}(x_{ij}, y_{ij}, t_p, c) = A^{FR}_{Non-negative}(x_{ij}, y_{ij}, t_p, c) \Big/ \sum_{c=1}^{l} A^{FR}_{Non-negative}(x_{ij}, y_{ij}, t_p, c) \quad (7)$$

where $A^{FR}(x_{ij}, y_{ij}, t_p, c)$ is the final FR c^{th} class fraction for FR pixel (x_{ij}, y_{ij}) at t_p. A flowchart for estimating the FR class fraction images at t_p in SFSDAF is in Fig. 2.

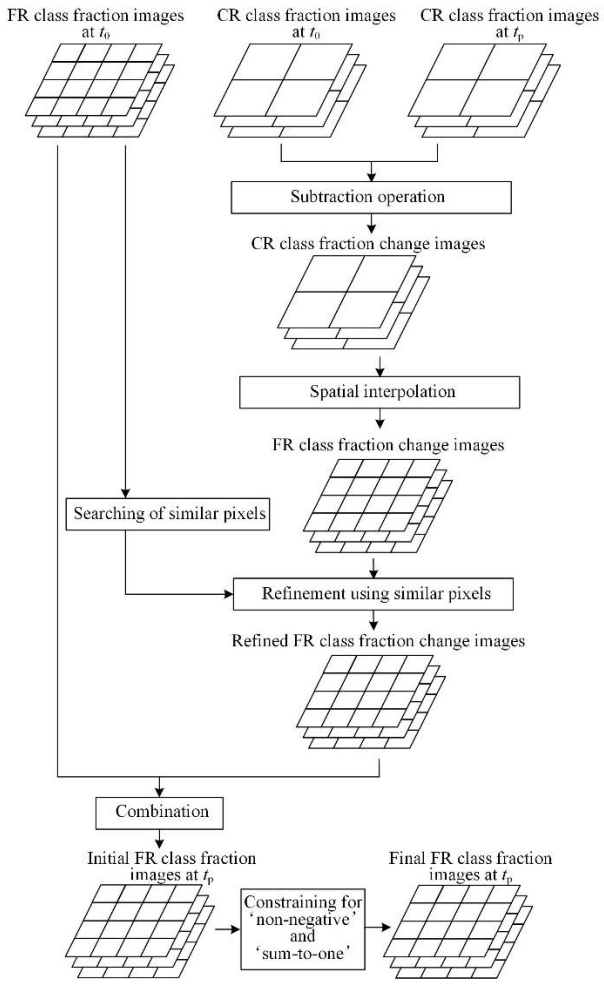

Fig.2. Flowchart of the estimation of FR land cover fraction images at t_p in SFSDAF.

2.2.3 Estimation of the FR endmember change from t_0 to t_p

The FR endmember change is estimated based on an inversion method of linear mixture equations which is used in FSDAF. Assume $\Delta E^{FR}(c,b)$ is the change in the b^{th} spectrum in the c^{th} endmember in the FR images from t_0 to t_p. Assume the change in the spectrum in the endmember are invariant in the CR and FR images (Gao et al. 2006; Zhu et al. 2016). The CR pixel reflectance change can be calculated as a linear combination of endmember reflectance change weighted by the fraction of each class within the CR pixel:

$$\Delta R^{CR}(x_i, y_i, b) = \sum_{c=1}^{l} A^{CR}(x_i, y_i, t_0, c) \times \Delta E^{FR}(c,b) \quad (8)$$

where $\Delta R^{CR}(x_i, y_i, b)$ is the reflectance change in the b^{th} band in the CR pixel (x_i, y_i) from t_0 to t_p:

$$\Delta R^{CR}(x_i, y_i, b) = R^{CR}(x_i, y_i, t_p, b) - R^{CR}(x_i, y_i, t_0, b) . \quad (9)$$

$R^{CR}(x_i, y_i, t_0, b)$ and $R^{CR}(x_i, y_i, t_p, b)$ are the b^{th} band reflectance values in the CR pixel (x_i, y_i) at t_0 and t_p, respectively. Eq. (8) can be formulated as Eq. (10), which is valid when the land cover composition is unchanged in the CR pixels from t_0 to t_p. n ($n>l$) CR pixels are selected to constitute the equation set as with the previous study of Zhu et al. (2016), and $\Delta E^{FR}(c, b)$ is estimated using the LSE method to solve the inversion of Eq. (10).

$$\begin{bmatrix} \Delta R^{CR}(x_1, y_1, b) \\ \vdots \\ \Delta R^{CR}(x_i, y_i, b) \\ \vdots \\ \Delta R^{CR}(x_n, y_n, b) \end{bmatrix} = \begin{bmatrix} A^{CR}(x_1, y_1, t_0, 1) & A^{CR}(x_1, y_1, t_0, 2) & \cdots & A^{CR}(x_1, y_1, t_0, l) \\ \vdots & \vdots & & \vdots \\ A^{CR}(x_i, y_i, t_0, 1) & A^{CR}(x_i, y_i, t_0, 2) & \cdots & A^{CR}(x_i, y_i, t_0, l) \\ \vdots & \vdots & & \vdots \\ A^{CR}(x_n, y_n, t_0, 1) & A^{CR}(x_n, y_n, t_0, 2) & \cdots & A^{CR}(x_n, y_n, t_0, l) \end{bmatrix} \begin{bmatrix} \Delta E^{FR}(1, b) \\ \vdots \\ \Delta E^{FR}(c, b) \\ \vdots \\ \Delta E^{FR}(l, b) \end{bmatrix} . \quad (10)$$

2.2.4 Estimation of the FR endmembers at t_p

The endmembers in the FR image at t_p are calculated according to the endmembers in the FR image at t_0 and the endmember change in the FR images from t_0 to t_p:

$$E^{FR}(c, b, t_p) = E^{FR}(c, b, t_0) + \Delta E^{FR}(c, b) \quad (11)$$

where $E^{FR}(c, b, t_p)$ is the b^{th} spectrum in the c^{th} endmember in the FR images at t_p.

2.3 Temporal prediction of the FR image at t_p

With the estimated temporal land cover class fraction changes and temporal endmember changes from t_0 to t_p, the temporal prediction for FR pixel (x_{ij}, y_{ij}) in the b^{th} band at t_p (i.e., $R_{TP}^{FR}(x_{ij}, y_{ij}, t_p, b)$) is calculated as:

$$R_{TP}^{FR}(x_{ij}, y_{ij}, t_p, b) = R^{FR}(x_{ij}, y_{ij}, t_0, b) + \left[\sum_{c=1}^{l} A^{FR}(x_{ij}, y_{ij}, t_p, c) \times E^{FR}(c, b, t_p) - \sum_{c=1}^{l} A^{FR}(x_{ij}, y_{ij}, t_0, c) \times E^{FR}(c, b, t_0) \right] . \quad (12)$$

where $R^{FR}(x_{ij}, y_{ij}, t_0, b)$ is the FR pixel reflectance in the b^{th} band at t_0. $\sum_{c=1}^{l} A^{FR}(x_{ij}, y_{ij}, t_{\text{p}}, c) \times E^{FR}(c, b, t_{\text{p}})$ represents the reflectance of a FR pixel (x_{ij}, y_{ij}) given the FR land cover fractions and endmembers in the b^{th} band of FR image at t_{p}, and $\sum_{c=1}^{l} A^{FR}(x_{ij}, y_{ij}, t_0, c) \times E^{FR}(c, b, t_0)$ represents the reflectance of FR pixel (x_{ij}, y_{ij}) given the FR land cover fractions and endmembers in the b^{th} band of FR image at t_0. The term $\sum_{c=1}^{l} A^{FR}(x_{ij}, y_{ij}, t_{\text{p}}, c) \times E^{FR}(c, b, t_{\text{p}}) - \sum_{c=1}^{l} A^{FR}(x_{ij}, y_{ij}, t_0, c) \times E^{FR}(c, b, t_0)$ in Eq. (12) accounts for combined effect from land cover fraction change and endmember change in the land surface reflectance change.

Note that although $\sum_{c=1}^{l} \hat{A}^{FR}(x_{ij}, y_{ij}, t_{\text{p}}, c) \times E^{FR}(c, b, t_{\text{p}})$ can represent the FR reflectance information at t_{p}, it uses endmember information which represents the average spectral reflectance for a class instead of the reflectance of one FR pixel, and is insufficient to represent the spatial heterogeneity and intra-class variability in reflectance values at the FR scale. Thus, the FR reflectance image at t_0, which contains FR spatial heterogeneity in reflectance, is used in temporal predicting the FR image at t_{p} in Eq. (12).

2.4 Prediction of the final FR image at t_{p}

In addition to the land cover class fraction information contained in the CR image at t_{p}, the reflectance values in this CR image can also provide land cover change information if the change is apparent at the CR scale (Liu et al. 2019a; Zhu et al. 2016). To address this issue, like FSDAF, the CR image at t_{p} is spatially interpolated to the FR scale using TPS or bicubic interpolation to produce a spatial prediction image at t_{p} (Zhu et al. 2016). In SFSDAF, the temporal prediction and spatial prediction refer to the steps of producing the FR reflectance images at t_{p}. Although the spatial interpolation is also used on the CR class fraction change images, the purpose is the generation of land cover class fraction change images from t_0 to t_{p} instead of FR reference images at t_{p}, and we do not call this process a spatial

prediction. Like FSDAF, the spatial prediction in SFSDAF only downscales the CR reflectance image at t_p, and the land cover class information contained is not used.

The spatial prediction image and temporal prediction image are combined to produce the final FR image at t_p. The combination of spatial prediction and temporal prediction as with the previous study in FSDAF (Zhu et al. 2016) is directly used in SFSDAF. In particular, the combination method considers the spatial distribution of FR pixel reflectance residual between the predicted and true reflectance values and assumes that the error distribution is related to the landscape heterogeneity. The details of calculating a FR pixel residual term $r^{FR}\left(x_{ij}, y_{ij}, b\right)$, which measures the difference between the prediction and true reflectance values of FR pixel $\left(x_{ij}, y_{ij}\right)$ in the b^{th} band, can be referred in Eqs (14)-(19) in Zhu et al. (2016).

With the estimated FR pixel residual $r\left(x_{ij}, y_{ij}, b\right)$, the total reflectance change of a FR pixel (i.e., $\Delta R^{FR}\left(x_{ij}, y_{ij}, b\right)$), in the b^{th} band for FR pixel (x_{ij}, y_{ij}), can be calculated by adding the temporal change of this FR pixel as:

$$\Delta R^{FR}\left(x_{ij}, y_{ij}, b\right) = \left[\sum_{c=1}^{l} A^{FR}(x_{ij}, y_{ij}, t_p, c) \times E^{FR}(c, b, t_p) - \sum_{c=1}^{l} A^{FR}(x_{ij}, y_{ij}, t_0, c) \times E^{FR}(c, b, t_0)\right] + r^{FR}\left(x_{ij}, y_{ij}, b\right) \quad (13)$$

Like FSDAF, similar pixels are used in the combine the total FR reflectance change and the FR reflectance image at t_0 to avoid predicting 'blocky' output (Zhu et al. 2016). The final prediction for FR pixel (x_{ij}, y_{ij}) in the b^{th} band at t_p (i.e., $R^{FR}\left(x_{ij}, y_{ij}, t_p, b\right)$), is the sum of the observation at t_0 and the final estimate of total reflectance change as:

$$R^{FR}\left(x_{ij}, y_{ij}, t_p, b\right) = R^{FR}\left(x_{ij}, y_{ij}, t_0, b\right) + \sum_{k=1}^{N} w_k \times \Delta R^{FR}\left(x_k, y_k, b\right) \quad (14)$$

3. Experiments

The potential of SFSDAF and a set of comparison methods were evaluated using degraded and real

remotely sensed images. The first experiment was based on Landsat and real MODIS imagery for a heterogeneous landscape. The second experiment was based on Landsat and real MODIS imagery for a region that undergoes some land cover change. The FR images are the Landsat surface reflectance images downloaded from Google Earth Engine. The CR images are real MODIS images of MCD43A4 surface reflectance products, which are a daily product adjusted using a bidirectional reflectance distribution function (BRDF) downloaded from the USGS. Details of the experiments based on degraded images can be found in Sections S1 and S2 in the Supplementary data. In all experiments, six spectral bands were used: Red (R), Green (G), Blue (B), Near Infrared (NIR) and two Short Wavelength Infrared bands (SWIR-1 and SWIR-2).

3.1 Experiment 1—Landsat and real MODIS imagery for a heterogeneous landscape

Two Landsat 5 TM images acquired on August 31, 2007 (Fig. 3(c)) and March 10, 2008 (Fig. 3(d)) near Canberra, Australia (147°52'22"E, 34°4'21"S) were used. Two MODIS surface reflectance product MCD43A4 images acquired at the same date to the Landsat images were used, and were re-projected from sinusoidal projection to UTM project with a spatial resolution of 480 m. The scale factor between the MODIS and Landsat images was 16. The data covers an area of 29 km × 29 km (960 × 960 Landsat image pixels). This area is mainly composed of farmland and has a high degree of spatial heterogeneity. The image pair on August 31, 2007 (Fig. 3(a) and (c)) and the MODIS image on March 10, 2008 (Fig. 3(b)) were used to produce a prediction of the Landsat image on March 10, 2008. The actual Landsat image for March 10, 2008 (Fig. 3(d)) was used for validation.

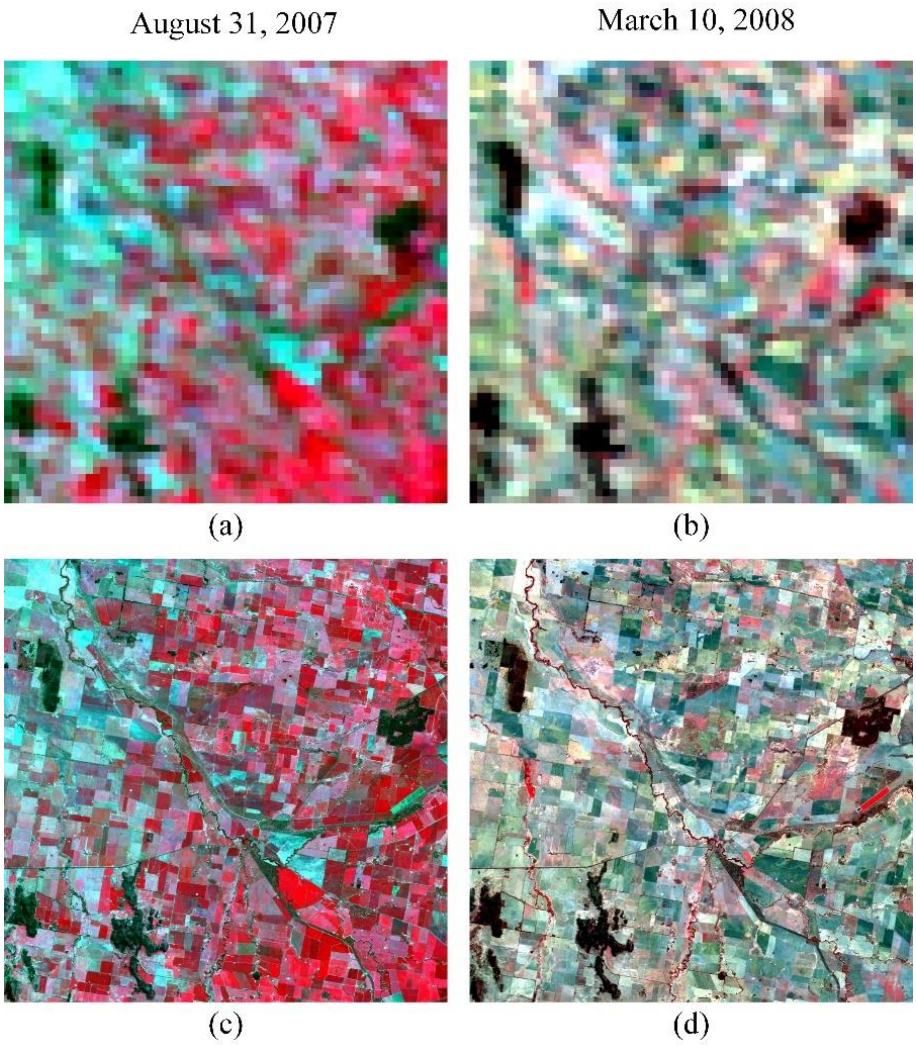

Fig. 3. Test data for Landsat and real MODIS imagery for a heterogeneous landscape. (a) MODIS MCD43A4 image acquired on August 31, 2007, (b) MODIS MCD43A4 image acquired on March 10, 2008, (c) Landsat image acquired on August 31, 2007, and (b) Landsat image acquired on March 10, 2008.

3.2 Experiment 2—Landsat and real MODIS imagery for a region that undergoes some land cover change

Two Landsat 5 TM images acquired on November 26, 2004 (Fig. 4(b)) and December 12, 2004 (Fig. 4(d)) in Gwydir, Australia (149°16'45"E, 29°5'12"S) were used (Emelyanova et al. 2013). Two MODIS surface reflectance product MCD43A4 images acquired on the same date as the Landsat images were

used, and were re-projected from sinusoidal projection to UTM project with a spatial resolution of 480 m. The scale factor between the MODIS and Landsat images was 16. The data covers an area of 48 km × 48 km (1600 × 1600 Landsat image pixels). A large flood occurred in the study area and is evident in the image on December 12, 2004 (Fig. 4(d)). The flood resulted in a (temporary) land cover conversion to water for certain pixels. The images pair on November 26, 2004 (Fig. 4(a) and (c)) and the MODIS image on December 12, 2004 (Fig. 4(b)) were used to generate a Landsat image on December 12, 2004. The actual Landsat image for December 12, 2004 (Fig. 4(d)) was used for validation.

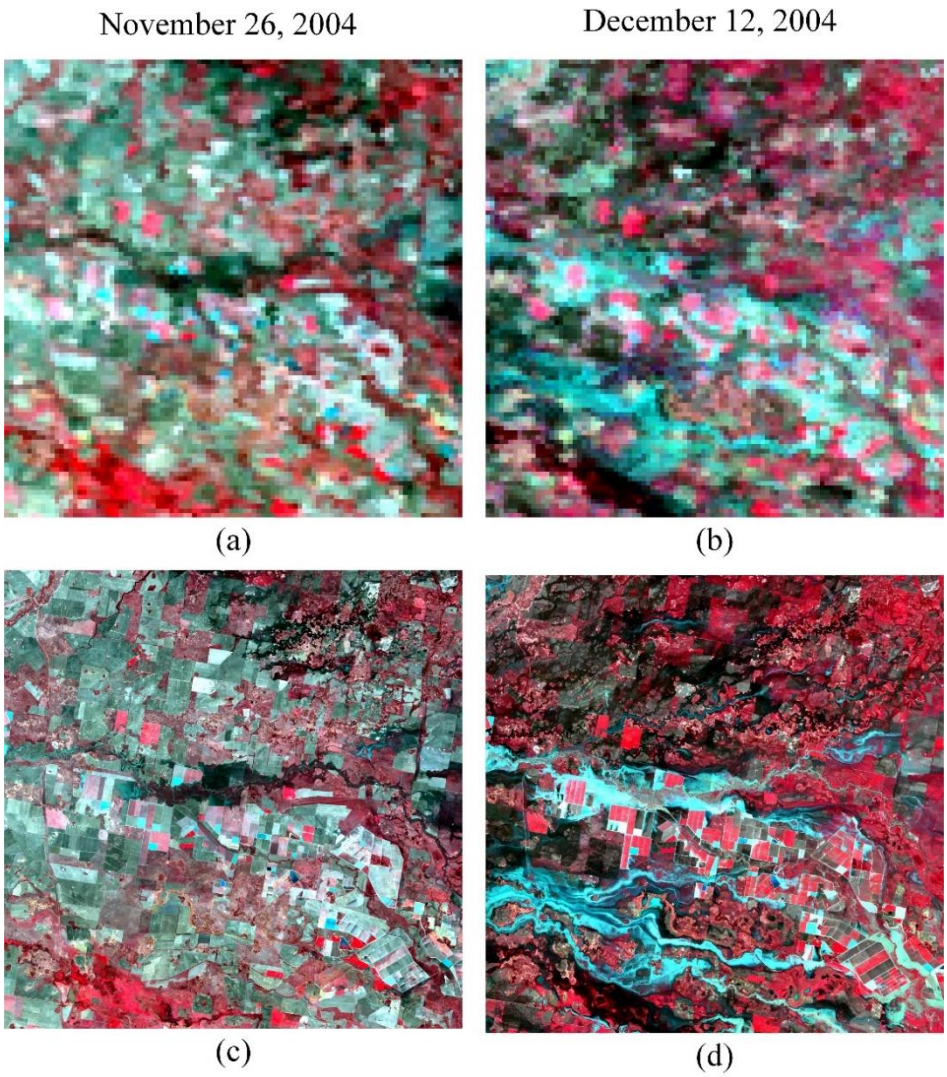

Fig. 4. Test data for Landsat and real MODIS imagery for a region that undergoes some land cover change. (a) MODIS MCD43A4

image acquired on November 26, 2004, (b) MODIS MCD43A4 image acquired on December 12, 2004, (c) Landsat image acquired on November 26, 2004, and (b) Landsat image acquired on December 12, 2004.

3.3 Comparison and accuracy assessment

The outputs obtained from SFSDAF were compared visually and quantitatively with those from three popular methods: STARFM (Gao et al. 2006), UBDF (Zurita-Milla et al. 2008) and FSDAF (Zhu et al. 2016). These comparator methods are all one-pair case spatio-temporal image fusion methods. For STARFM, FSDAF, and SFSDAF, the number of classes was set to 4, the number of similar pixels was set to 20, and the size of the moving window was set to 16 (Gao et al. 2006; Zhu et al. 2016). The accuracy of image prediction was assessed by comparison to the relevant reference image and expressed using the root mean square error (RMSE), average absolute difference (AAD), correlation coefficient (CC), and structure similarity (SSIM). The closer the value of RMSE or AAD to 0 and the closer value of CC or SSIM to 1 the more similar the predicted image is to the true image.

4. Results

4.1 Experiment 1—Landsat and real MODIS imagery for a heterogeneous landscape

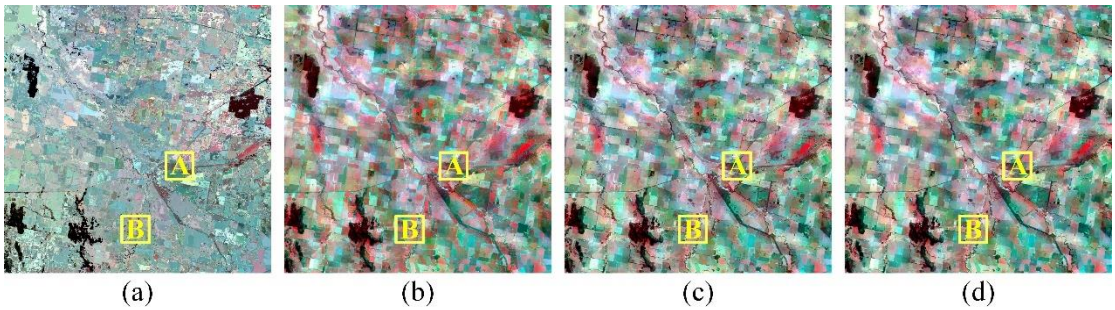

Fig. 5. Images predicted by the different methods. (a) UBDF, (b) STARFM, (c) FSDAF, and (d) SFSDAF.

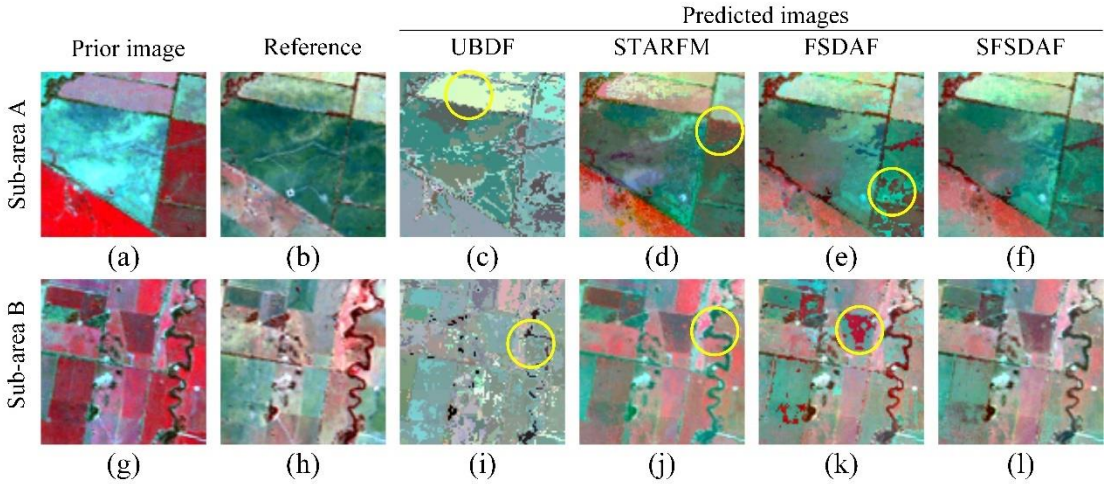

Fig. 6. Zoomed images (96 × 96 Landsat image pixels) of sub-areas A and B shown in Fig. 5. (a) Prior image acquired on November 24, 2001 in sub-area A, (b) reference image acquired on February 12, 2002 in sub-area A, (c) UBDF predicted image in sub-area A, (d) STARFM predicted image in sub-area A, (e) FSDAF predicted image in sub-area A, (f) SFSDAF predicted image in sub-area A, (g) prior image acquired on November 24, 2001 in sub-area B, (h) reference image acquired on February 12, 2002 in sub-area B, (i) UBDF predicted image in sub-area B, (j) STARFM predicted image in sub-area B, (k) FSDAF predicted image in sub-area B, and (l) SFSDAF predicted image in sub-area B.

Table 1

Accuracies of the different methods for Landsat and real MODIS imagery for a heterogeneous landscape in experiment 3. Bold data indicate the most accurate method.

Spectral	RMSE				AAD				CC				SSIM			
band	UBDF	STARFM	FSDAF	SFSDAF	UBDF	STARFM	FSDAF	SFSDAF	UBDF	STARFM	FSDAF	SFSDAF	UBDF	STARFM	FSDAF	SFSDAF
B	0.0182	0.0111	0.0116	**0.0106**	0.0147	0.0086	0.0090	**0.0082**	0.4829	0.6831	0.6864	**0.7311**	0.4652	0.6753	0.6860	**0.7277**
G	0.0239	0.0180	0.0171	**0.0165**	0.0191	0.0143	0.0135	**0.0131**	0.5346	0.7317	0.7521	**0.7702**	0.5162	0.7134	0.7425	**0.7586**
R	0.0288	0.0238	0.0231	**0.0216**	0.0222	0.0181	0.0177	**0.0166**	0.5721	0.7096	0.7375	**0.7675**	0.5634	0.7003	0.7346	**0.7622**
NIR	0.0415	0.0399	0.0382	**0.0336**	0.0326	0.0302	0.0292	**0.0254**	0.5103	0.5999	0.6391	**0.6888**	0.5048	0.5819	0.6202	**0.6792**
SWIR-1	0.0523	0.0486	0.0447	**0.0430**	0.0401	0.0377	0.0346	**0.0333**	0.5956	0.6723	0.7267	**0.7459**	0.5671	0.6661	0.7196	**0.7373**
SWIR-2	0.0706	0.0615	0.0581	**0.0562**	0.0562	0.0474	0.0460	**0.0444**	0.5390	0.6010	0.6711	**0.6992**	0.5243	0.5961	0.6674	**0.6943**

The predicted images obtained from the different methods are shown in Fig. 5 and for the sub-areas of A and B in Fig. 6. The predicted image from SFSDAF was visually the most accurate. The predicted

image obtained from UBDF contained patches that had homogeneous spectral values as highlighted for the area in the yellow circle in Fig. 6(c). The river appeared blurred in the predicted image obtained from UBDF in Fig. 6(i). This is because UBDF predicted a FR pixel reflectance image using the spectra of several neighbor CR pixels around it. When UBDF predicted the FR pixel reflectance of the river, only few pixels of the associated class were involved in the unmixing approach, and this resulted in an inaccurate estimation of that class. STARFM predicted reflectance values that were dissimilar to the reference when the regions were spatially heterogeneous, an example is highlighted in the yellow circles in Fig. 6(d) and (j). The FSDAF predicted image contained small patches of dissimilar spectral values for regions that were relatively homogeneous in the reference image, see for example the areas highlighted in Fig. 6(e) and (k). In contrast, the prediction image from SFSDAF was most similar to the reference image, and did not include the patches produced by FSDAF (Fig. 6(f) and (l)). Quantitative measures demonstrate that the proposed SFSDAF predicted the lowest RMSE and AAD and the highest CC and SSIM among all of the assessed methods (Table 1).

4.2 Experiment 2—Landsat and real MODIS imagery for a region that undergoes some land cover

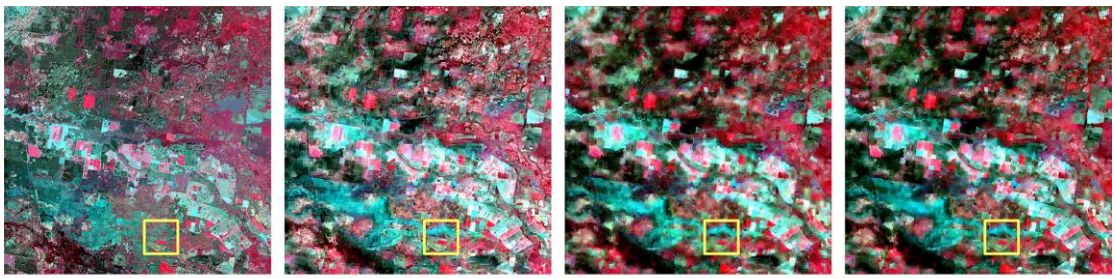

Fig. 7. Images predicted by the different methods. (a) UBDF, (b) STARFM, (c) FSDAF, and (d) SFSDAF.

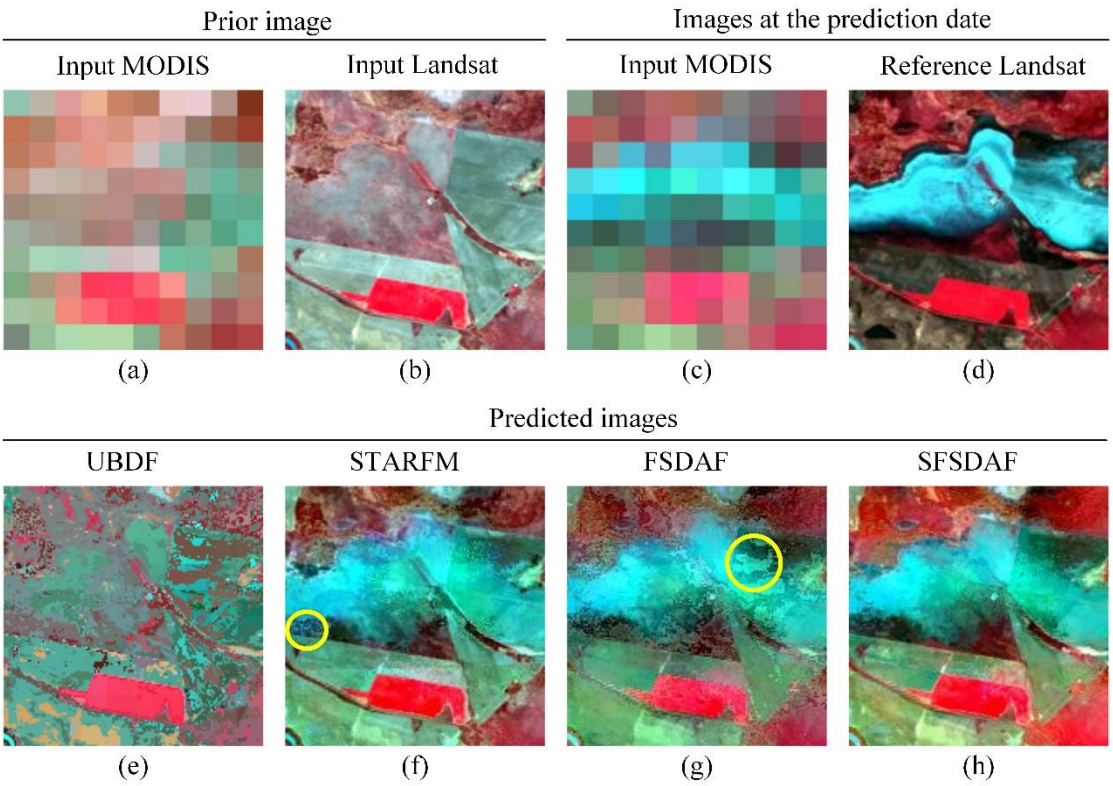

Fig. 8. Zoomed images (160 × 160 Landsat image pixels) of the sub-areas shown in Fig. 7. (a) Prior MODIS image acquired on November 26, 2004, (b) prior Landsat image acquired on November 26, 2004, (c) MODIS image acquired on December 12, 2004, (d) Landsat image acquired on December 12, 2004 as the reference image, (e) UBDF predicted image, (f) STARFM predicted image, (g) FSDAF predicted image, (h) SFSDAF predicted image.

Table 2

Accuracies of the different methods for Landsat and real MODIS imagery for a region that undergoes some land cover change in experiment 4. Bold data indicate the most accurate method.

Spectral	RMSE				AAD				CC				SSIM			
band	UBDF	STARFM	FSDAF	SFSDAF	UBDF	STARFM	FSDAF	SFSDAF	UBDF	STARFM	FSDAF	SFSDAF	UBDF	STARFM	FSDAF	SFSDAF
B	0.0202	0.0177	0.0173	**0.0170**	0.0144	0.0126	0.0124	**0.0120**	0.4553	0.6193	0.6253	**0.6393**	0.4305	0.6090	0.6052	**0.6215**
G	0.0301	0.0261	0.0255	**0.0248**	0.0231	0.0191	0.0186	**0.0177**	0.4572	0.6110	0.6147	**0.6390**	0.4268	0.5977	0.5903	**0.6143**
R	0.0367	0.0318	0.0318	**0.0306**	0.0280	0.0222	0.0223	**0.0207**	0.4729	0.6233	0.6135	**0.6463**	0.4472	0.6130	0.5901	**0.6231**
NIR	0.0637	0.0387	0.0412	**0.0376**	0.0496	0.0295	0.0320	**0.0290**	0.6204	0.8094	0.7808	**0.8170**	0.6085	0.8090	0.7805	**0.8168**
SWIR-1	0.0947	0.0635	0.0618	**0.0584**	0.0769	0.0493	0.0480	**0.0450**	0.4626	0.6851	0.6631	**0.7036**	0.4189	0.6830	0.6467	**0.6881**
SWIR-2	0.0659	0.0466	0.0436	**0.0410**	0.0531	0.0363	0.0329	**0.0308**	0.4523	0.6862	0.6676	**0.7132**	0.4121	0.6853	0.6520	**0.7002**

Figs. 7 and 8 present results for the entire study area and selected sub-area respectively. A flood

resulted in a temporary land cover change for part of the region including the sub-area from November 26, 2004 to December 12, 2004. This flood is, for example, evident by comparing the Landsat images in Fig. 8(b) and (d). The MODIS image acquired on December 12, 2004 recorded the flood, and the spatio-temporal image fusion methods were used to predict the FR image for the flooded landscape (Fig. 8(e)-(h)). The predicted image from UBDF that was dissimilar to the reference image, with the flood not represented well. This is because UBDF assumed land cover was unchanged and was not suitable for the prediction of reflectance change caused by such an abrupt land cover change. The flood was apparent in the image predicted by STARFM, FSDAF and SFSDAF images (Fig. 8(f)-(h)). The sub-area is not in a complex heterogeneous landscape, and STARFM, which is suitable in imagery fusion in homogeneous regions, captured well the reflectance change caused by the flood. FSDAF predicted the flood with sharp reflectance changes between the neighboring FR pixels that were dissimilar to the reference image, see for examples that are highlighted in the yellow circle in Fig. 8(g). STARFM contained large errors that are highlighted in the yellow circle in Fig. 8(f), while the SFSDAF prediction was the most similar to the reference image.

The quantitative measures in Table 2 show that UBDF predicted the highest RMSE and AAD and lowest CC and SSIM. This is because UBDF is based on the assumption that land cover is unchanged, and hence is not suitable when land cover change occurs. STARFM, FSDAF, and SFSDAF could capture some land cover change information and generated high accuracy. FSDAF usually generated lower RMSE and AAD values than STARFM, but STARFM generated higher CC and SSIM than FSDAF in the NIR and SWIR bands. This is explained by the fact that the study area was not spatially heterogeneous and STARFM could predict reflectance values more satisfactorily. SFSDAF generated the lowest RMSE and AAD values and the highest CC and SSIM values among all methods, indicating the advantage of

the proposed method.

5. Discussion

The comparison of the different methods to account for land cover class change information was stressed in this section. In particular, since the difference between FSDAF and SFSDAF is in the temporal prediction step, a theoretical comparison between the temporal prediction steps used in FSDAF and SFSDAF was conducted, and the experimental results between the predicted images from FSDAF and SFSDAF temporal prediction steps were revisited. The influence of the number of classes on SFSDAF and further improvements for the proposed method were discussed in this section.

5.1 Comparison of the different methods to account for land cover class change information

The established spatio-temporal fusion methods could accurately predict gradual reflectance changes caused by endmember change. However, in order to accurately predict the reflectance change of FR pixels caused by land cover class change (particularly that which is abrupt), FR land cover class change information is necessary. In the established spatio-temporal image fusion methods, the reason that the two-pairs case methods can often produce a more accurate result than the one-pair case methods is that they can better capture the land cover class change information by comparing the FR images that were acquired before and after the prediction date. For existing one-pair case methods, the FR land cover class change information is usually unavailable. As the impact of land cover class change on the FR pixel reflectance change is generally different from that caused by endmember change, the confusion of reflectance change caused by the different sources may lead to blurry outputs for the pixels associated with changed land cover predicted by the existing one-pair case methods. Land cover class information is used in the unmixing-based methods, but these methods, such as UBDF, assume the land cover class

is unchanged. FSDAF spatially interpolated the CR image at t_p to FR scale to include land cover class change information, but the information about how land covers are changed from t_0 to t_p is still not directly derived. In contrast to existing one-pair case methods, the proposed SFSDAF method accommodates for the fact that the FR reflectance change caused by endmember and land cover class change have different characteristics, and models land cover class and endmember changes separately. As a result, the SFSDAF increased the prediction accuracy, especially for pixels representing sites that experienced both land cover class and endmember changes.

5.2 Theoretical comparison between FSDAF and SFSDAF in temporal prediction

The FSDAF derives endmember change in the temporal prediction step, whereas SFSDAF directly derives endmember change and land cover class fraction change simultaneously in its temporal prediction step in Eq. (12). In SFSDAF, if the FR pixels are simply assumed to be pure pixels, Eq. (12) can be rewritten as:

$$R_{TP}^{FR}(x_{ij}, y_{ij}, t_p, b) = R^{FR}(x_{ij}, y_{ij}, t_0, b) + \left(E^{FR}(\beta, b, t_p) - E^{FR}(\alpha, b, t_0)\right) \quad (15)$$

where α and β (α=1, …, l, β=1, …, l) are the single class for the fine pixel (x_{ij}, y_{ij}) at t_0 and t_p, respectively. Furthermore, if it is assumed that the land cover class is unchanged from t_0 to t_p (α=β=c) in Eq. (15), then

$$E^{FR}(c, b, t_p) - E^{FR}(c, b, t_0) = \Delta E^{FR}(c, b). \quad (16)$$

Eq. (16) can be rewritten as:

$$R_{TP}^{FR}(x_{ij}, y_{ij}, t_p, b) = R^{FR}(x_{ij}, y_{ij}, t_0, b) + \Delta E^{FR}(c, b) \quad (17)$$

Eq. (17) is the same algorithm used in the temporal prediction of the FR image at t_p in FSDAF. Therefore, the FSDAF temporal prediction can thus be viewed as a special case of the SFSDAF temporal prediction if the fine pixels are pure and have an unchanged land cover class.

5.3 Experimental comparison between FSDAF and SFSDAF in temporal prediction

In order to explicitly compare the FSDAF and SFSDAF to show if including land cover fraction change could improve the temporal prediction result, the second experiment reported, which is in a flooded region for land cover change prediction, was revisited.

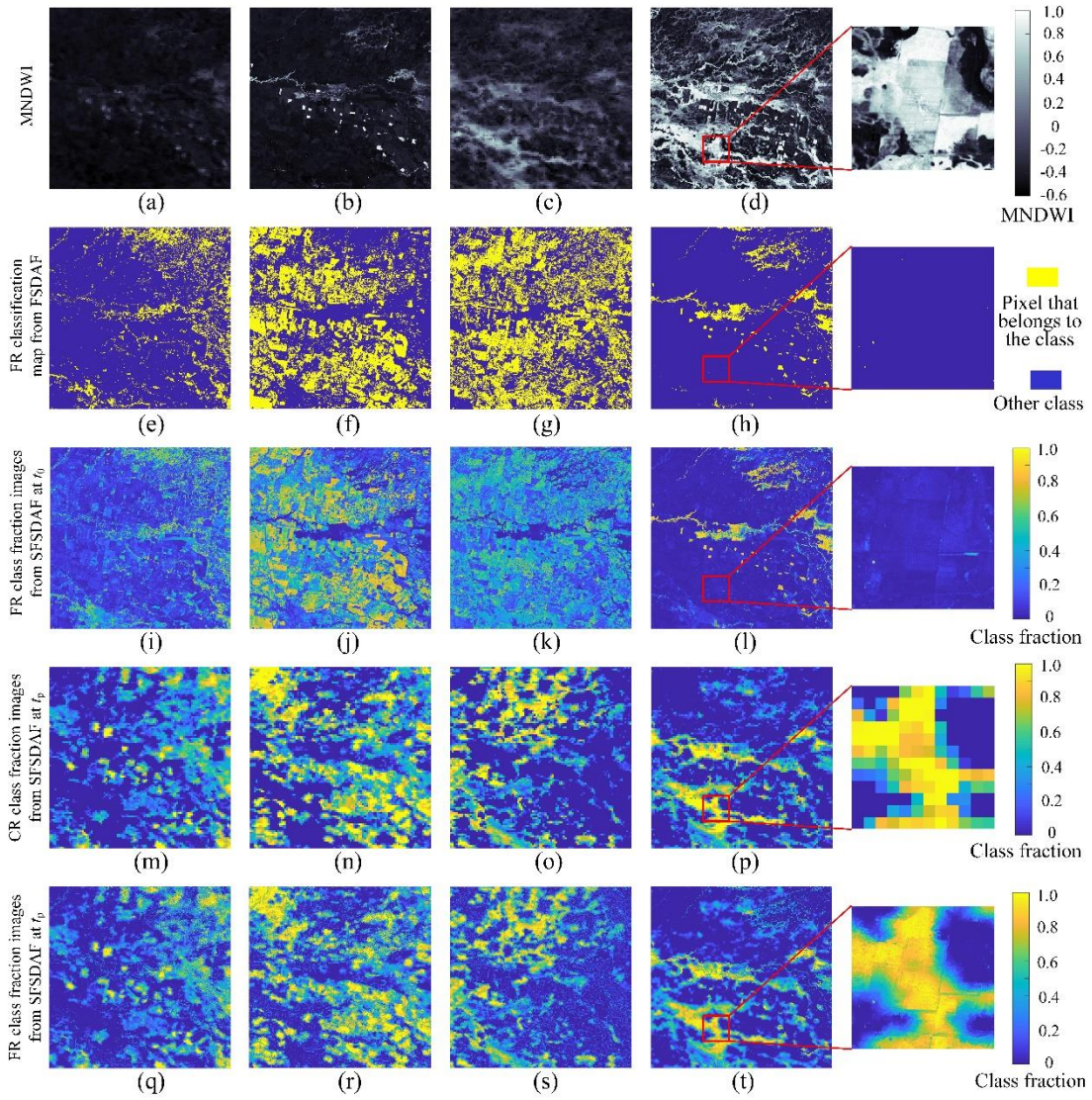

Fig. 9. Comparison of temporal prediction image from FSDAF and SFSDAF in the Landsat and real MODIS image experiment in the flood area in Gwydir, Australia. (a) and (b) are MNDWI derived from the MODIS and Landsat images on November 26, 2004, and (c) and (d) are MNDWI derived from the MODIS and Landsat images on December 12, 2004. (e)-(h) are the binary FR land cover maps for the four land cover classes on November 26 in FSDAF, 2004; yellow means pixels that belong to that class, and

blue means the pixel belongs to one of the other classes. (i)-(l) are the FR class fraction images for the four land cover classes on November 26, 2004 in SFSDAF. (m)-(p) are the CR class fraction images for the four land cover classes on December 12, 2004 in SFSDAF. (q)-(t) are the FR class fraction images for the four land cover classes on December 12, 2004 in SFSDAF.

The comparison of FSDAF and SFSDAF temporal prediction in a landscape that includes some land cover change was assessed using the Landsat and real MODIS imagery used in the experiment focused on the flood in Gwydir, Australia. For this study area, the flood made an obvious (temporary) land cover change. The modified normalized difference water index (MNDWI) was used to provide the water information at different dates (Xu 2006). Only a small proportion of the study area was covered with water in the image on November 26, 2004 (Fig. 9(b)). On December 12, 2004 the study area had experienced a flood that influenced a large proportion of pixels (Fig. 9(d)). In FSDAF, the unsupervised *k*-means algorithm was applied on the FR Landsat image on November 26, 2004 and used to cluster the FR image into four classes (Fig. 9(e)-(h)). By comparison with the MNDWI image on November 26, 2004 (Fig. 9(b)) the 4th cluster in Fig. 9(h) could indicate the water class. FSDAF assumed land cover was unchanged in the temporal prediction and used a classification map, which is separated into four binary land cover maps in Fig. 9(e)-(h), in temporal predicting the FR image at t_p. Obviously, these binary categorical maps only represent information on November 26, 2004, but could not reflect the flood that is evident on December 12, 2004. In contrast, the unmixed CR class fraction image for the 4th class (Fig. 9(p)) and the downscaled FR fraction image for the 4th class (Fig. 9(p)) predicted by SFSDAF could represent the flood on December 12, 2004. With the predicted sub-pixel scale land cover fraction change information, SFSDAF could accommodate the combined effect from both endmember change and land cover class fraction change in relation to the total land surface change. The comparison of FSDAF and SFSDAF temporal prediction images in a heterogeneous region is referred to Section S3 in the

Supplementary data, and visual comparison shows that the SFSDAF temporal prediction image as more similar to the reference image.

Quantitative measures of the FR temporal prediction images at t_p from FSDAF and SFSDAF in the experiments using real MODIS imagery are shown in Figs. S8-S9 in the Supplementary data. In all the experiments, the SFSDAF temporal prediction decreased the RSME and AAD values, and increased the CC and SSIM values compared with the FSDAF temporal prediction, and the scatter plots shown in Fig. S10 in the Supplementary data show that the SFSDAF temporal prediction points were closer to the 1:1 line than those in the FSDAF temporal prediction image. Figs. S8-S9 in the Supplementary data also show that the SFSDAF final predictions, which were combinations of both spatial and temporal prediction images, were more accurate than SFSDAF temporal predictions. The SFSDAF spatial prediction can directly capture abrupt reflectance change from the CR reflectance image at t_p when the change is detectable in the CR pixel scale, but the spatial prediction usually generates smooth result in which the spatial details are lost (Liu et al. 2019a; Zhu et al. 2016). The SFSDAF temporal prediction captures abrupt reflectance change based on the information of endmember change and land cover class fraction change, but the CR reflectance changes from t_0 to t_p were not directly used. Combining the spatial and temporal predictions could take advantage of both steps to make the final prediction more accurate for FSDAF and SFSDAF.

5.4 Computation times comparison

Since SFSDAF is based on FSDAF, the computation times of the two methods were compared. FSDAF and SFSDAF were performed based on a computer having Intel(R) Xeon(R) Silver 4116 processor (2.10GHz) and 32 GB RAM. FSDAF was programmed based on IDL, while SFSDAF was programmed based on Matlab. The number of clusters was set to four in FSDAF and SFSDAF in all

experiments. FSDAF divided the entire image into different blocks or sub-regions in predicting the reflectance image. The FSDAF default block size was set to 30, and each block contained 30 × 30 CR pixels. The first experiment contained 60 × 60 pixels in the CR image, which was divided into 4 blocks. The second experiment contained 100 × 100 pixels in the CR image, which was divided into 16 blocks. The FSDAF spatial interpolation method applied to the CR image at t_p was TPS. The SFSDAF for comparison did not divide the image into blocks, and the spatial interpolation method applied to the CR class fraction images and CR reflectance image at t_p was the bicubic interpolation.

Table 3

The computations time of FSDAF and SFSDAF in the two experiments using real MODIS images.

	Experiment 1	Experiment 2
FSDAF	320s	989s
SFSDAF	378s	1007s

The computation times of SFSDAF are longer than those of FSDAF. In particular, SFSDAF used about 50s more than FSDAF in the first experiment, and about 20s more than FSDAF in the second experiment. The FSDAF computation time is nearly proportional to the number of blocks divided, while the SFSDAF computation time is nearly proportional to the number of CR pixels contained in the image. The SFSDAF running times are longer because SFSDAF introduced additional processes compared with FSDAF. The three main processes introduced by SFSDAF are: step (1) generating FR fraction images at t_0 based on a clustering map and endmembers, step (2) unmixing the CR image at t_p to class fraction images, and step (3) downscaling the CR class fraction images to FR scale. Take the first experiment for example, the running time is about 20s in step (1), about 33s in step (2) and about 52s in step (3). These three steps take account for about 28% in the SFSDAF total computation time.

5.5 Influence of the number of classes on SFSDAF

The inputs to SFSDAF include the number of classes, the similar neighbor FR pixel number and the moving window size. The selections of the optimal value of similar pixel number and the moving window size are the same as FSDAF. The optimal number of classes was between four and six in this study. SFSDAF involves a linear spectral unmixing approach in estimating the land cover fractions at the dates of the prior image pair and prediction, which usually requires that the number of classes be less than the number of spectral bands. When fusion MODIS and Landsat imagery, which have six similar spectral bands (R, G, B, NIR, SWIR-1, and SWIR-2), the optimal endmember number should be no more than the number of spectral bands in the imagery to get reliable results in SFSDAF. In addition, SFSDAF can be employed to predict not only reflectance images but also products (e.g. NDVI and NDWI) which are linearly additive in space. Different to STARFM and FSDAF which can be directly applied to predict a vegetation index based on CR and FR indices, SFSDAF cannot be directly used to predict a vegetation index. This is because the vegetation index image has only one band, while the linear mixture model used in SFSDAF requires the number of bands should be no less than the number of classes. In this case, SFSDAF can first be used to predict reflectance images and then the obtained vegetation index products can be generated from the resulting reflectance images.

Increasing the number of classes in SFSDAF does not always increase the accuracy but increases the computation time. Setting the optimal number of classes from four to six is competent even applying SFSDAF to a complex area with a spatially heterogeneous landscape. While the number of classes is set for the entire image, it is possible for the endmembers to be derived globally or locally. In the global approach, the endmembers are estimated based on typical pixels selected from the entire image. In the local approach, the study area can be divided into sub-regions, and the endmember for each class is derived in each sub-region to account for intra-class variability in endmembers.

5.6 Further improvement of SFSDAF

The proposed SFSDAF introduces sub-pixel class fraction information into spatio-temporal image fusion, and showed superior predictions relative to existing state-of-art methods. The main improvement of SFSDAF is in the temporal prediction step. SFSDFA extended FSDAF by directly deriving land cover class fraction change in this step. In SFSDAF, the CR image at t_p is first unmixed to CR class fraction images, which are compared with the CR class fraction images at t_0 in calculating the CR class fraction change image. The CR class fraction change image is then downscaled to represent the FR scale land cover class fraction change. It needs to be emphasized that the unmixing and downscaling processes are open problems which could be explored further. First, the unmixing of the CR image at t_p is based on the linear mixture model, and other methods such as the nonlinear mixture model could also be used. Second, the SFSDAF considers the one-pair case, and could be extended to the two-pairs case, and several fraction change downscaling methods could be used. For instance, Zhang et al. (2018) proposed a class fraction imagery fusion method using two-pairs of FR class fraction images that pre- and post-date the CR class fraction images to estimate the FR class fraction images. FR class fraction change images are derived based on kernel ridge regression, and then a temporal-weighted fusion model is applied to predict the FR class fraction images at the prediction date. This method could also be applied in SFSDAF when two-pair images are available.

Although SFSDAF could predict abrupt reflectance change to a certain extent, the predicted image may have blurring effects where the reflectance change is abrupt. An example is shown in Fig. 8(h) in the flooded area in the second experiment. The prediction of FR abrupt reflectance change in SFSDAF is mainly attributed to the downscaling of the CR class fraction change images in the temporal prediction step and the downscaling of the CR reflectance image at t_p in the spatial prediction step. Both steps use

bicubic or TPS interpolation, which is based on spatial dependence, for high computation efficiency. The downscaled class fraction change images and reflectance image are smoothed, which may result in blurring effects in the abrupt change area in the final prediction image. Powerful downscaling methods, such as single image super-resolution and learning-based methods especially the convolutional neural network-based deep learning method, could be a promising solution to maintain the spatial details for abrupt change (Belgiu and Stein 2019; Liu et al. 2019b; Song et al. 2018).

The SFSDAF used the same spatial prediction process as FSDAF, and the combination of temporal prediction and spatial prediction could be explored further. The FSDAF combination is based on the assumption that errors between real and predicted imagery depend mainly on the landscape homogeneity. This is modified by IFSDAF which considered the spatial autocorrelation in NDVI and used constrained linear squares in the combination of temporal prediction and spatial prediction of NDVI imagery (Liu et al. 2019a). Other temporal prediction and spatial prediction combination methods could be developed in the future.

The proposed SFSDAF is not only suitable to fuse MODIS and Landsat imagery, but also appropriate in fusing other CR and FR imagery which have similar spectral bands. For instance, SFSDAF can generate nearly daily Sentinel-2 imagery by fusing the blue, green, red and NIR bands from Sentinel-2 and Sentinel-3 imagery (Mileva et al. 2018; Wang and Atkinson 2018). A comprehensive study of using SFSDAF to fuse imagery from various sensors will be further developed.

6. Conclusions

Spatio-temporal fusion of remotely sensed imagery that accommodates reflectance change caused by both endmember and land cover class changes is a challenge for present methods. Most existing fusion methods accommodate endmember change, but are unable to directly derive and use land cover class

change information. In this paper, a novel SFSDAF that accommodates both endmember and land cover class fraction change was proposed. SFSDAF is built on FSDAF which has been widely used for its simplicity and flexibility. FSDAF combines a temporal prediction which considers endmember changes at different dates and a spatial prediction which interpolates the CR image at the prediction time to FR scale in comparison with the temporal prediction image to represent land cover change. FSDAF assumes the land cover class is unchanged and only uses the endmember change information in the temporal prediction, while SFSDAF improved FSDAF by addressing the combined effect from both endmember change and class fraction change. Using only a single prior CR and FR image pair and the CR image at the prediction date, SFSDAF predicts sub-pixel scale land cover class fractions for the FR pixels at the prediction date and uses the FR land cover class fraction change between the dates of prior image pair and prediction. With the derived FR land cover class fraction change information, SFSDAF allows accurate prediction even for regions that undergo an abrupt land cover change. SFSDAF is perhaps the first one-pair case spatio-temporal image fusion method that predicts image reflectance change by directly exploring FR land cover class change information, and therefore opens a new view for spatio-temporal remotely sensed image fusion. This greatly helps the realization of the full potential of satellite remote sensing in land cover change studies, especially those focused on contemporary change and/or requiring near-real-time analysis.

SFSDAF was compared with UBDF, STARFM, and FSDAF for heterogeneous landscapes and for landscapes that experienced land cover change. Results show that the SFSDAF predicted images were the most similar to the reference image in all of the experiments reported. UBDF predicted images with patches of homogeneous reflectance values, and STARFM predictions were poor in heterogeneous regions. In the fusion of imagery with the phenological change, FSDAF predicted the overly sharp change

in reflectance values for the neighboring FR pixels that were dissimilar to the reference in shape and spectral values, while SFSDAF successfully captured the reflectance changes and maintained the shape of land cover patches. In the fusion of imagery for sites that experienced land cover change, SFSDAF predicted FR scale class fractions that represented the change (e.g. flood) well unlike FSDAF. In all experiments, the SFSDAF predicted images were the most accurate, having the lowest RMSE and AAD values and the highest CC and SSIM values. SFSDAF improved the popular FSDAF framework and could be used in various applications because of its flexibility. The SFSDAF Matlab package is available from https://www.researchgate.net/profile/Xiao_Li52.

Acknowledgment

This work was supported in part by the Hubei Province Natural Science Fund for Distinguished Young Scholars (Grant No. 2018CFA062), in part by the Youth Innovation Promotion Association CAS (Grant No. 2017384), in part by the Natural Science Foundation of China (Grant No. 61671425), in part from the Hubei Province Natural Science Fund for Innovation Groups (Grant No. 2019CFA019), in part by the Strategic Priority Research Program of Chinese Academy of Sciences (Grant No XDA 2003030201), and in part by the National Science Fund for Distinguished Young Scholars (Grant No. 41725006). The authors would like to thank Xiaolin Zhu and Jin Chen for providing the FSDAF programs. The authors would also like to thank Dr Leo Lymburner and team (Geoscience Australia) for making the Landsat-5 TM Gwydir database available.

Appendix A

Notations

i: the index of CR pixel

j: the index of FR pixel

l: the number of classes

(x_i, y_i): i^{th} CR pixel

(x_{ij}, y_{ij}): j^{th} FR pixel in the i^{th} CR pixel

$R^{CR}(x_i, y_i, t_0, b)$ and $R^{CR}(x_i, y_i, t_p, b)$: the b^{th} band reflectance values at CR pixel (x_i, y_i) at t_0 and t_p

$R^{FR}(x_{ij}, y_{ij}, t_0, b)$ and $R^{FR}(x_{ij}, y_{ij}, t_p, b)$: the b^{th} band reflectance values at FR pixel (x_{ij}, y_{ij}) at t_0 and t_p

$\Delta R^{CR}(x_i, y_i, b)$: the change between $R^{CR}(x_i, y_i, t_0, b)$ and $R^{CR}(x_i, y_i, t_p, b)$

$A^{CR}(x_i, y_i, t_0, c)$ and $A^{CR}(x_i, y_i, t_p, c)$: the c^{th} class fractions at CR pixel (x_i, y_i) at t_0 and t_p

$A^{FR}(x_{ij}, y_{ij}, t_p, c)$ and $A^{FR}(x_{ij}, y_{ij}, t_p, c)$: the c^{th} class fractions a at FR pixel (x_{ij}, y_{ij}) at t_0 and t_p

$\Delta A^{CR}(x_i, y_i, c)$: the c^{th} class fraction change at CR pixel (x_i, y_i) from t_0 to t_p

$\Delta A_{SI}^{FR}(x_k, y_k, c)$: the c^{th} class fraction change at the k^{th} FR pixel from spatial interpolation

$\Delta A_{SI-Refine}^{FR}(x_{ij}, y_{ij}, c)$: the refined FR class fraction change for the c^{th} class at FR pixel (x_{ij}, y_{ij})

$E^{CR}(c,b,t_p)$: the b^{th} spectrum in the c^{th} endmember in the CR image at t_p

$E^{FR}(c,b,t_0)$ and $E^{FR}(c,b,t_p)$: the b^{th} spectrums in the c^{th} endmember in the FR image at t_0 and t_p

$\Delta E^{FR}(c,b)$: the change between $E^{FR}(c,b,t_0)$ and $E^{FR}(c,b,t_p)$

References

Alpaydin, E. (1998). Soft vector quantization and the EM algorithm. *Neural Networks, 11*, 467-477

Alves, D.B., Lloveria, R.M., Perez-Cabello, F., & Vlassova, L. (2018). Fusing Landsat and MODIS data to retrieve multispectral information from fire-affected areas over tropical savannah environments in the Brazilian Amazon. *International Journal of Remote Sensing, 39*, 7919-7941

Amoros-Lopez, J., Gomez-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, 23*, 132-141

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

Chapin, F.S., Zavaleta, E.S., Eviner, V.T., Naylor, R.L., Vitousek, P.M., Reynolds, H.L., Hooper, D.U., Lavorel, S., Sala, O.E., Hobbie, S.E., Mack, M.C., & Diaz, S. (2000). Consequences of changing biodiversity. *Nature, 405*, 234-242

Chen, B., Chen, L., Huang, B., Michishita, R., & Xu, B. (2018). Dynamic monitoring of the Poyang Lake wetland by integrating Landsat and MODIS observations. *ISPRS Journal of Photogrammetry and Remote Sensing, 139*, 75-87

Chen, B., Huang, B., & Xu, B. (2017). A hierarchical spatiotemporal adaptive fusion model using one image pair. *International Journal of Digital Earth, 10*, 639-655

Dubrule, O. (1984). Comparing splines and kriging. *Computers & Geosciences, 10*, 327-338

Emelyanova, I.V., McVicar, T.R., Van Niel, T.G., Li, L.T., & van Dijk, A.I.J.M. (2013). Assessing the accuracy of blending Landsat-

MODIS surface reflectances in two landscapes with contrasting spatial and temporal dynamics: A framework for algorithm selection. *Remote Sensing of Environment, 133*, 193-209

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., Helkowski, J.H., Holloway, T., Howard, E.A., Kucharik, C.J., Monfreda, C., Patz, J.A., Prentice, I.C., Ramankutty, N., & Snyder, P.K. (2005). Global consequences of land use. *Science, 309*, 570-574

Fu, D., Chen, B., Wang, J., Zhu, X., & Hilker, T. (2013). An improved image fusion approach based on enhanced spatial and temporal the adaptive reflectance fusion model. *Remote Sensing, 5*, 6346-6360

Gaertner, P., Foerster, M., & Kleinschmit, B. (2016). The benefit of synthetically generated RapidEye and Landsat 8 data fusion time series for riparian forest disturbance monitoring. *Remote Sensing of Environment, 177*, 237-247

Gao, F., Anderson, M.C., Zhang, X., Yang, Z., Alfieri, J.G., Kustas, W.P., Mueller, R., Johnson, D.M., & Prueger, J.H. (2017). Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. *Remote Sensing of Environment, 188*, 9-25

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*, 2207-2218

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

Hilker, T., Wulder, M.A., Coops, N.C., Linke, J., McDermid, G., Masek, J.G., Gao, F., & White, J.C. (2009). A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on Landsat and MODIS. *Remote Sensing of Environment, 113*, 1613-1627

Huang, B., & Song, H.H. (2012). Spatiotemporal reflectance fusion via sparse representation. *IEEE Transactions on Geoscience and Remote Sensing, 50*, 3707-3716

Huang, B., & Zhang, H. (2014). Spatio-temporal reflectance fusion via unmixing: accounting for both phenological and land-cover

changes. *International Journal of Remote Sensing, 35*, 6213-6233

Ju, J., & Roy, D.P. (2008). The availability of cloud-free Landsat ETM plus data over the conterminous United States and globally. *Remote Sensing of Environment, 112*, 1196-1211

Keshava, N., & Mustard, J.F. (2002). Spectral unmixing. *IEEE Signal Processing Magazine, 19*, 44-57

Keys, R.G. (1981). Cubic convolution interpolation for digital image processing. *IEEE Transactions on Acoustics Speech and Signal Processing, 29*, 1153-1160

Li, X., Du, Y., & Ling, F. (2015). Sub-pixel-scale land cover map updating by integrating change detection and sub-pixel mapping. *Photogrammetric Engineering and Remote Sensing, 81*, 59-67

Li, X., Ling, F., Foody, G.M., & Du, Y. (2016). A superresolution land-cover change detection method using remotely sensed images with different spatial resolutions. *IEEE Transactions on Geoscience and Remote Sensing, 54*, 3822-3841

Li, X., Ling, F., Foody, G.M., Ge, Y., Zhang, Y., & Du, Y. (2017). Generating a series of fine spatial and temporal resolution land cover maps by fusing coarse spatial resolution remotely sensed images and fine spatial resolution land cover maps. *Remote Sensing of Environment, 196*, 293-311

Liao, C., Wang, J., Pritchard, I., Liu, J., & Shang, J. (2017). A spatio-temporal data fusion model for generating NDVI time series in heterogeneous regions. *Remote Sensing, 9*

Ling, F., Li, W., Du, Y., & Li, X. (2011). Land cover change mapping at the subpixel scale with different spatial-resolution remotely sensed imagery. *IEEE Geoscience and Remote Sensing Letters, 8*, 182-186

Liu, M., Yang, W., Zhu, X., Chen, J., Chen, X., Yang, L., & Helmer, E.H. (2019a). An Improved Flexible Spatiotemporal DAta Fusion (IFSDAF) method for producing high spatiotemporal resolution normalized difference vegetation index time series. *Remote Sensing of Environment, 227*, 74-89

Liu, X., Deng, C., Chanussot, J., Hong, D., & Zhao, B. (2019b). StfNet: A two-stream convolutional neural network for spatiotemporal image fusion. *IEEE Transactions on Geoscience and Remote Sensing, 57*, 6552-6564

Maselli, F., Chiesi, M., & Pieri, M. (2019). A new method to enhance the spatial features of multitemporal NDVI image series. *IEEE Transactions on Geoscience and Remote Sensing, 57*, 4967-4979

Mileva, N., Mecklenburg, S., & Gascon, F. (2018). New tool for spatiotemporal image fusion in remote sensing - a case study approach using Sentinel-2 and Sentinel-3 data. In L. Bruzzone, & F. Bovolo (Eds.), *Image and Signal Processing for Remote Sensing Xxiv*

Schmidt, M., Lucas, R., Bunting, P., Verbesselt, J., & Armston, J. (2015). Multi-resolution time series imagery for forest disturbance and regrowth monitoring in Queensland, Australia. *Remote Sensing of Environment, 158*, 156-168

Song, H., & Huang, B. (2013). Spatiotemporal satellite image fusion through one-pair image learning. *IEEE Transactions on Geoscience and Remote Sensing, 51*, 1883-1896

Song, H., Liu, Q., Wang, G., Hang, R., & Huang, B. (2018). Spatiotemporal satellite image fusion using deep convolutional neural networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11*, 821-829

Sun, Y., Zhang, H., & Shi, W. (2019). A spatio-temporal fusion method for remote sensing data using a linear injection model and local neighbourhood information. *International Journal of Remote Sensing, 40*, 2965-2985

Vitousek, P.M., Mooney, H.A., Lubchenco, J., & Melillo, J.M. (1997). Human domination of Earth's ecosystems. *Science, 277*, 494-499

Walker, J.J., de Beurs, K.M., Wynne, R.H., & Gao, F. (2012). Evaluation of Landsat and MODIS data fusion products for analysis of dryland forest phenology. *Remote Sensing of Environment, 117*, 381-393

Wang, J., & Huang, B. (2017). A rigorously-weighted spatiotemporal fusion model with uncertainty analysis. *Remote Sensing, 9*

Wang, Q., & Atkinson, P.M. (2018). Spatio-temporal fusion for daily Sentinel-2 images. *Remote Sensing of Environment, 204*, 31-42

Wang, Q., Shi, W., & Atkinson, P.M. (2016). Spatiotemporal subpixel mapping of time-series images. *IEEE Transactions on Geoscience and Remote Sensing, 54*, 5397-5411

Wang, Q., Zhang, Y., Onojeghuo, A.O., Zhu, X., & Atkinson, P.M. (2017). Enhancing spatio-temporal fusion of MODIS and Landsat data by incorporating 250 m MODIS data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10*, 4116-4123

Wu, B., Huang, B., & Zhang, L. (2015). An Error-Bound-Regularized Sparse Coding for Spatiotemporal Reflectance Fusion. *IEEE Transactions on Geoscience and Remote Sensing, 53*, 6791-6803

Wu, M., Niu, Z., Wang, C., Wu, C., & Wang, L. (2012). Use of MODIS and Landsat time series data to generate high-resolution temporal synthetic Landsat data using a spatial and temporal reflectance fusion model. *Journal of Applied Remote Sensing, 6*

Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing, 27*, 3025-3033

Xu, Y., & Huang, B. (2014). A spatio-temporal pixel-swapping algorithm for subpixel land cover mapping. *IEEE Geoscience and Remote Sensing Letters, 11*, 474-478

Zhang, L., Weng, Q., & Shao, Z. (2017). An evaluation of monthly impervious surface dynamics by fusing Landsat and MODIS time series in the Pearl River Delta, China, from 2000 to 2015. *Remote Sensing of Environment, 201*, 99-114

Zhang, Y., Foody, G.M., Ling, F., Li, X., Ge, Y., Du, Y., & Atkinson, P.M. (2018). Spatial-temporal fraction map fusion with multi-scale remotely sensed images. *Remote Sensing of Environment, 213*, 162-181

Zhao, Y., Huang, B., & Song, H. (2018). A robust adaptive spatial and temporal image fusion model for complex land surface changes. *Remote Sensing of Environment, 208*, 42-62

Zhong, D., & Zhou, F. (2019). Improvement of clustering methods for modelling abrupt land surface changes in satellite image fusions. *Remote Sensing, 11*

Zhu, X., Cai, F., Tian, J., & Williams, T.K.-A. (2018). Spatiotemporal fusion of multisource remote sensing data: literature survey, taxonomy, principles, applications, and future directions. *Remote Sensing, 10*

Zhu, X., Chen, J., Gao, F., Chen, X., & Masek, J.G. (2010). An enhanced spatial and temporal adaptive reflectance fusion model

for complex heterogeneous regions. *Remote Sensing of Environment, 114*, 2610-2623

Zhu, X., Helmer, E.H., Gao, F., Liu, D., Chen, J., & Lefsky, M.A. (2016). A flexible spatiotemporal method for fusing satellite images with different resolutions. *Remote Sensing of Environment, 172*, 165-177

Zhukov, B., Oertel, D., Lanzl, F., & Reinhackel, G. (1999). Unmixing-based multisensor multiresolution image fusion. *IEEE Transactions on Geoscience and Remote Sensing, 37*, 1212-1226

Zurita-Milla, R., Clevers, J.G.P.W., & Schdepman, M.E. (2008). Unmixing-based Landsat TM and MERIS FR data fusion. *IEEE Geoscience and Remote Sensing Letters, 5*, 453-457

Zurita-Milla, R., Kaiser, G., Clevers, J.G.P.W., Schneider, W., & Schaepman, M.E. (2009). Downscaling time series of MERIS full resolution data to monitor vegetation seasonal dynamics. *Remote Sensing of Environment, 113*, 1874-1885

Supplementary Data:

SFSDAF: an enhanced FSDAF that incorporates sub-pixel class fraction change information for spatio-temporal image fusion

Section S1. Experiment based on Landsat and MODIS-like imagery for a heterogeneous

Data description

Two Landsat 7 TM images acquired on November 24, 2001 (Fig. S1(c)) and February 12, 2002 (Fig. S1(d)) for Coleambally, southern New South Wales, Australia (145°56'40"E, 34°52'26"S) were used. The data covers an area of 29 km × 29 km (960 × 960 Landsat image pixels) for a heterogeneous agricultural region. The Landsat images were spatially degraded to the MODIS-like images (Fig. S1(a) and (b)). The scale factor between the MODIS-like and Landsat images was 16. The image pair on November 24, 2001 (Fig. S1(a) and (c)) and the MODIS-like image on February 12, 2002 (Fig. S1(b)) were used to produce a prediction of the Landsat image on February 12, 2002. The actual Landsat image for February 12, 2002 (Fig. S1(d)) was used for validation.

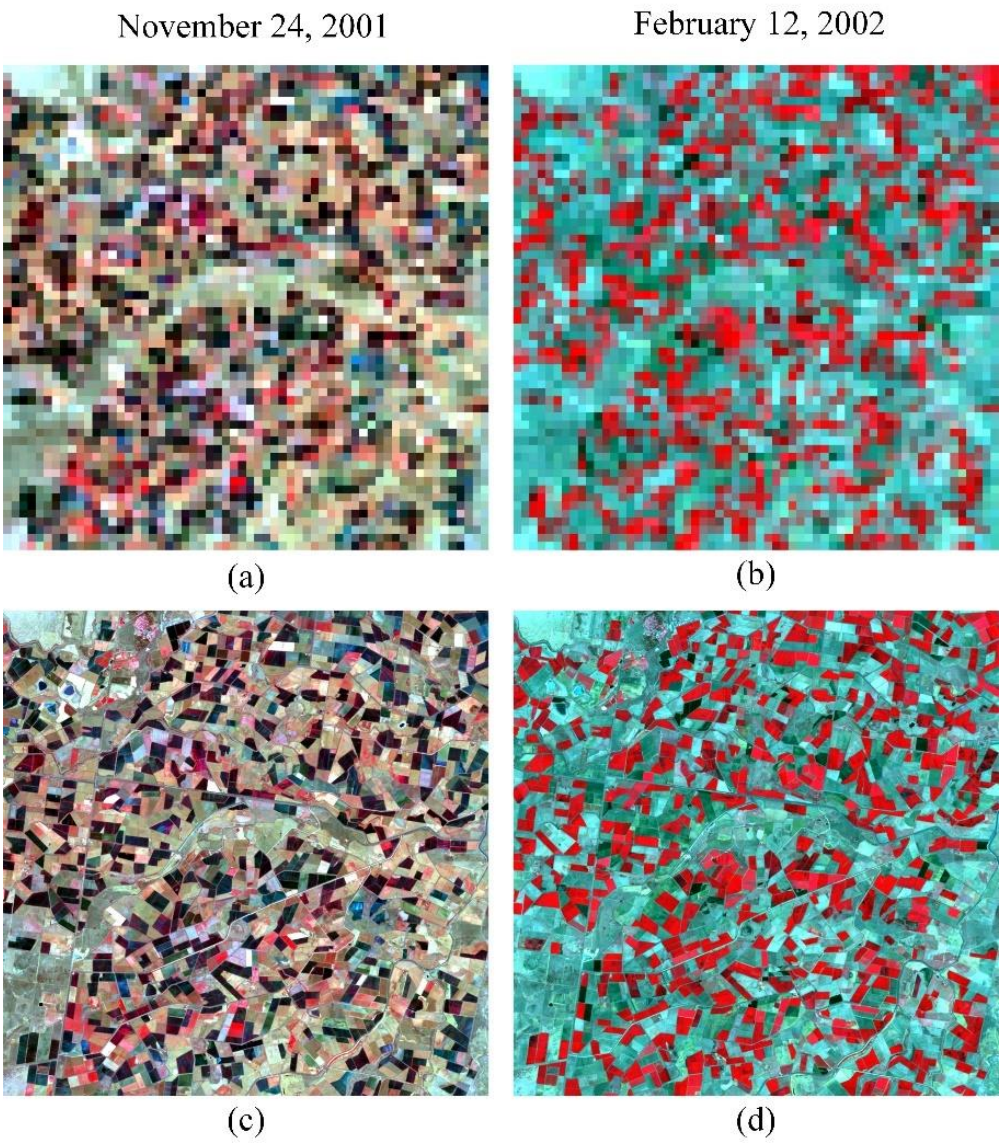

Fig. S1. Test data for Landsat and MODIS-like imagery for a heterogeneous landscape. This region is an irrigation area for rice. The region undergoes an apparent phenological change during the summer growing season for rice, and the false color composite images (RGB: bands 432) are different in reflectance values at different dates in Fig. S1. (a) MODIS-like image acquired on November 24, 2001, (b) MODIS-like image acquired on February 12, 2002, (c) Landsat image acquired on November 24, 2001, and (b) Landsat image acquired on February 12, 2002.

Results

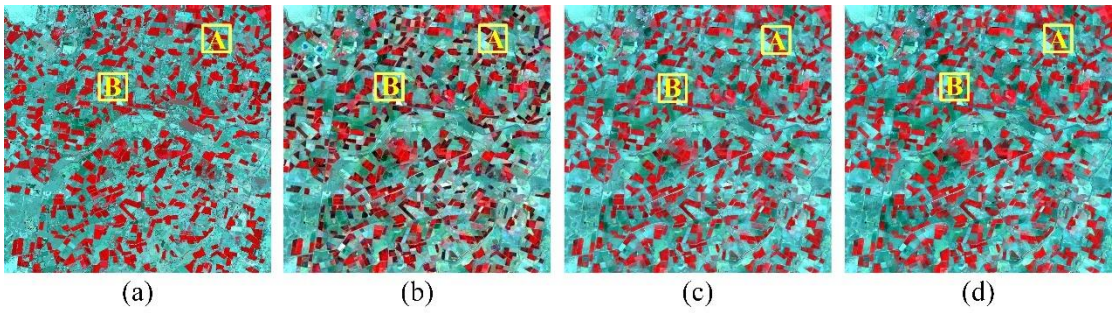

Fig. S2. Images predicted by the different methods. (a) UBDF, (b) STARFM, (c) FSDAF, and (d) SFSDAF.

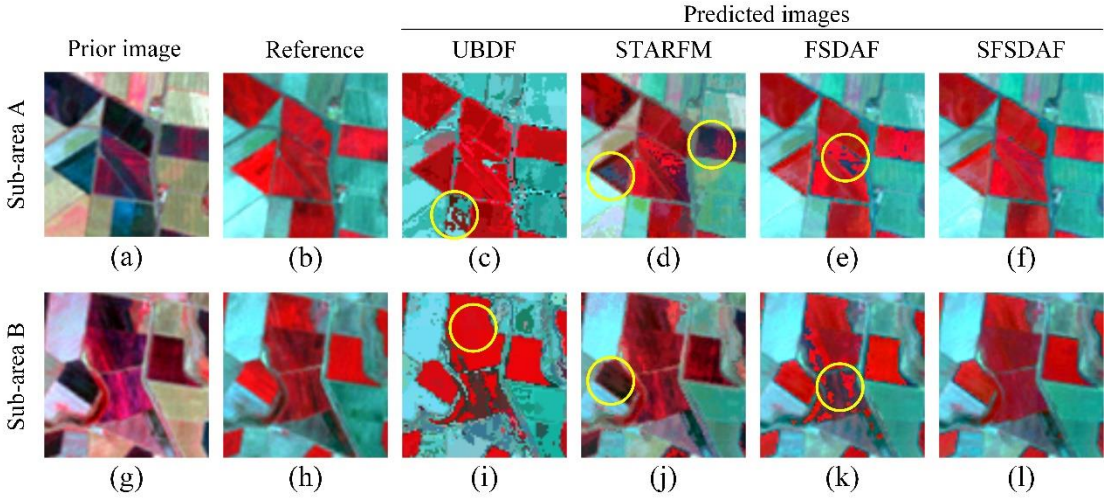

Fig. S3. Zoomed images (80 × 80 Landsat image pixels) of sub-areas A and B shown in Fig. S2 (a) Prior image acquired on November 24, 2001 in sub-area A, (b) reference image acquired on February 12, 2002 in sub-area A, (c) UBDF predicted image in sub-area A, (d) STARFM predicted image in sub-area A, (e) FSDAF predicted image in sub-area A, (f) SFSDAF predicted image in sub-area A, (g) prior image acquired on November 24, 2001 in sub-area B, (h) reference image acquired on February 12, 2002 in sub-area B, (i) UBDF predicted image in sub-area B, (j) STARFM predicted image in sub-area B, (k) FSDAF predicted image in sub-area B, and (l) SFSDAF predicted image in sub-area B.

The predicted image from SFSDAF was more similar to the reference image than those generated by the other methods (Fig. S2). The differences between the outputs generated by the four methods are particularly evident in the sub-area areas shown in Fig. S3. The image predicted by UBDF contained some clear errors. For example, the homogeneous region in yellow in Fig. S3(c) was shown as a mosaic of discrete patches. This is because UBDF assigned the reflectance of endmembers to the corresponding FR pixels; the FR pixels that clustered to different classes may have a large difference in spectral values

if the assigned endmembers have a large spectral difference. Additionally, the predicted reflectance was inappropriately homogeneous for large regions, see for examples that are highlighted with a yellow circle in Fig. S3(i) because the FR pixels of the same class were assigned the same reflectance values within a moving window. The predicted image generated by STARFM contained more spatial detail than that from UBDF, but dissimilar reflectance was predicted in boundary areas, an example is highlighted in a yellow circle in Fig. S3(d) and (j). This situation arose because STARFM is more suitable for homogeneous areas (Gao et al. 2006). FSDAF generated a predicted image that was more similar to the reference image at the boundaries than that from STARFM, but it predicted overly sharp changes in reflectance values for neighboring FR pixels that lead to discontinuities in homogeneous region, example highlighted in yellow in Fig. S3(e) and (k). In contrast, SFSDAF in Fig. S3(f) and (l) contained more spatial detail than UBDF, and successfully captured the reflectance changes and maintained the shape of patches better than STARFM and FSDAF.

Table S1

Accuracies of the different methods for Landsat and MODIS-like imagery for a heterogeneous landscape in section S1. Bold data indicate the most accurate method.

Spectral	RMSE				AAD				CC				SSIM			
band	UBDF	STARFM	FSDAF	SFSDAF	UBDF	STARFM	FSDAF	SFSDAF	UBDF	STARFM	FSDAF	SFSDAF	UBDF	STARFM	FSDAF	SFSDAF
B	0.0153	0.0112	0.0101	**0.0095**	0.0111	0.0082	0.0075	**0.0070**	0.7653	0.8709	0.8962	**0.9072**	0.7651	0.8527	0.8954	**0.9043**
G	0.0212	0.0149	0.0135	**0.0129**	0.0154	0.0107	0.0098	**0.0094**	0.7492	0.8723	0.8974	**0.9050**	0.7490	0.8563	0.8964	**0.9015**
R	0.0322	0.0214	0.0196	**0.0190**	0.0230	0.0152	0.0142	**0.0138**	0.7860	0.9028	0.9196	**0.9239**	0.7860	0.8967	0.9189	**0.9222**
NIR	0.0655	0.0701	0.0525	**0.0422**	0.0462	0.0479	0.0344	**0.0291**	0.6512	0.5020	0.7830	**0.8458**	0.6512	0.4689	0.7827	**0.8424**
SWIR-1	0.0550	0.0413	0.0361	**0.0336**	0.0393	0.0298	0.0262	**0.0241**	0.8060	0.8978	0.9143	**0.9252**	0.8060	0.8967	0.9137	**0.9230**
SWIR-2	0.0510	0.0363	0.0343	**0.0314**	0.0359	0.0251	0.0247	**0.0224**	0.8055	0.8991	0.9097	**0.9235**	0.8055	0.8984	0.9093	**0.9213**

Quantitative indices to indicate the quality of the predictions arising from the four methods are provided in Table S1. For all 6 bands, the image predicted by SFSDAF had the lowest RMSE and AAD and highest CC and SSIM values, showing that SFSDAF was the most accurate method assessed. The difference between SFSDAF and FSDAF was most apparent in the NIR and the two SWIR bands. For instance, compared with FSDAF in the NIR band, SFSDAF decreased RMSE by 0.0103 and AAD by 0.0053, and increased CC by 0.0628 and SSIM by 0.0597.

Section S2. Experiment based on Landsat and MODIS-like imagery for a heterogeneous region that undergoes some land cover change

Data description

Two Landsat 5 TM images acquired on October 18, 2009 (Fig. S4(c)) and April 28, 2010 (Fig. S4(d)) for Louisiana, USA (92°45'7"W, 30°23'29"N) were used. The data covers an area of 29 km × 29 km (960 × 960 Landsat image pixels). This area contains a heterogeneous landscape of farmland, and experienced land cover changes in transitions between crop and bare-land due to farming. The Landsat images were spatially degraded to the MODIS-like images (Fig. S4(a), (b)). The scale factor between the MODIS-like and Landsat images was 16. The image pair on October 18, 2009 (Fig. S4 (a) and (c)) and the MODIS-like image on April 28, 2010 (Fig. S4(b)) were used to produce a prediction of the Landsat image on April 28, 2010. The actual Landsat image for April 28, 2010 (Fig. S4(d)) was used for validation.

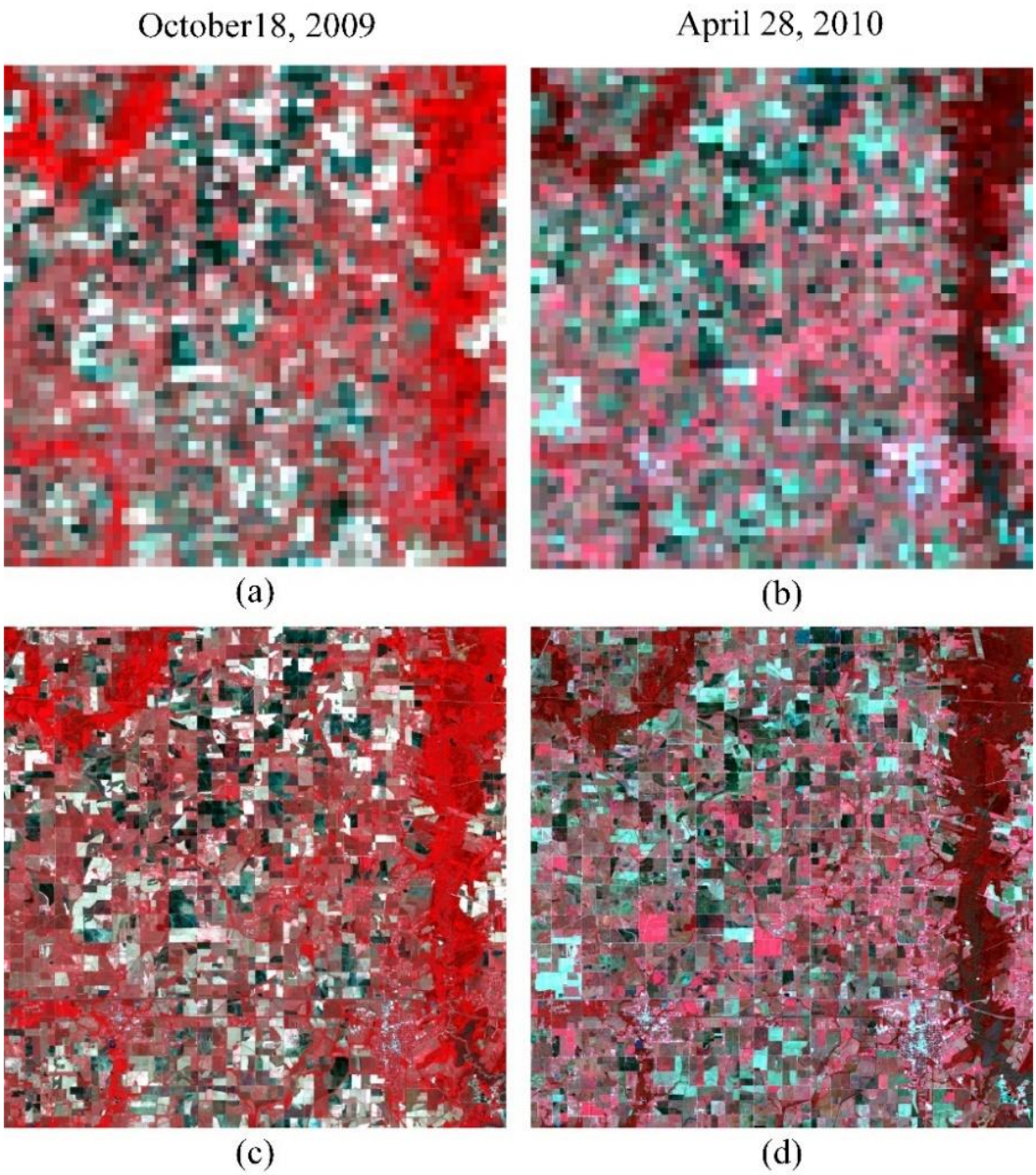

Fig. S4. Test data for Landsat and MODIS-like imagery for a heterogeneous region that undergoes some land cover change. (a) MODIS-like image acquired on October 18, 2009, (b) MODIS-like image acquired on April 28, 2010, (c) Landsat image acquired on October 18, 2009, and (b) Landsat image acquired on April 28, 2010.

Results

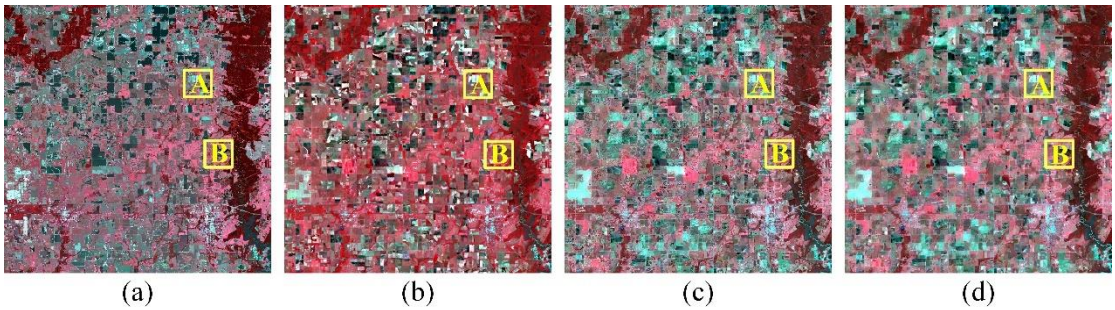

Fig. S5. Images predicted by the different methods. (a) UBDF, (b) STARFM, (c) FSDAF, and (d) SFSDAF.

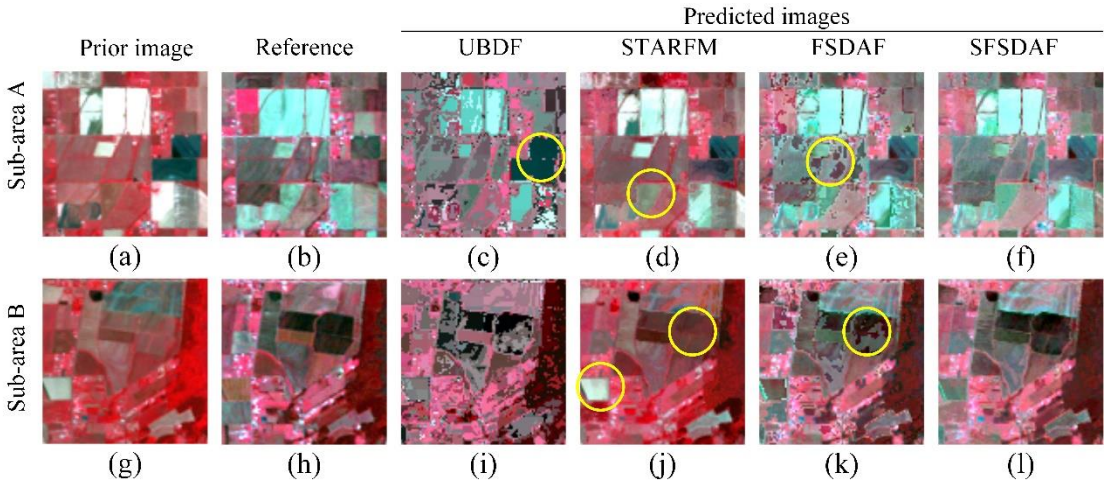

Fig. S6. Zoomed images (96 × 96 Landsat image pixels) of sub-areas A and B shown in Fig. S5. (a) Prior image acquired on November 24, 2001 in sub-area A, (b) reference image acquired on February 12, 2002 in sub-area A, (c) UBDF predicted image in sub-area A, (d) STARFM predicted image in sub-area A, (e) FSDAF predicted image in sub-area A, (f) SFSDAF predicted image in sub-area A, (g) prior image acquired on November 24, 2001 in sub-area B, (h) reference image acquired on February 12, 2002 in sub-area B, (i) UBDF predicted image in sub-area B, (j) STARFM predicted image in sub-area B, (k) FSDAF predicted image in sub-area B, and (l) SFSDAF predicted image in sub-area B.

Table S2

Accuracies of the different methods for Landsat and MODIS-like imagery for a heterogeneous region that undergoes some land cover change in section S2. Bold data indicate the most accurate method.

Spectral	RMSE				AAD				CC				SSIM			
band	UBDF	STARFM	FSDAF	SFSDAF	UBDF	STARFM	FSDAF	SFSDAF	UBDF	STARFM	FSDAF	SFSDAF	UBDF	STARFM	FSDAF	SFSDAF
B	0.0120	0.0132	0.0110	**0.0095**	0.0079	0.0081	0.0072	**0.0060**	0.5200	0.6166	0.6682	**0.7276**	0.5136	0.6031	0.6675	**0.7267**
G	0.0168	0.0180	0.0140	**0.0120**	0.0115	0.0115	0.0099	**0.0084**	0.5732	0.6650	0.7305	**0.7886**	0.5692	0.6513	0.7304	**0.7844**
R	0.0228	0.0245	0.0193	**0.0161**	0.0156	0.0157	0.0135	**0.0112**	0.5544	0.6634	0.7081	**0.7806**	0.5506	0.6455	0.7080	**0.7750**
NIR	0.0538	0.0446	0.0385	**0.0350**	0.0407	0.0327	0.0290	**0.0260**	0.4416	0.6408	0.7125	**0.7573**	0.4394	0.6405	0.7072	**0.7450**
SWIR-1	0.0540	0.0656	0.0437	**0.0382**	0.0395	0.0418	0.0298	**0.0256**	0.5156	0.5827	0.6854	**0.7523**	0.5132	0.5535	0.6824	**0.7419**
SWIR-2	0.0377	0.0539	0.0336	**0.0255**	0.0267	0.0343	0.0240	**0.0181**	0.4983	0.5264	0.6241	**0.7659**	0.4947	0.4759	0.6233	**0.7440**

The predicted images from different methods are in Fig. S5 and the sub-areas of A and B are in Fig. S6. Again, the images predicted by SFSDAF was the most accurate obtained. The image predicted by UBDF contained patches with homogeneous reflectance, highlighted in the yellow circle in Fig. S6(c). STARFM failed to predict the reflectance change for small patches such as that highlighted in Fig. S6(d) and (j) because STARFM is most suitable for homogeneous regions and the sub-areas shown have relatively high spatial heterogeneity. The predicted image from FSDAF better represented reflectance change for these small patches than STARFM, but it contained patches with spectral values and shapes that were different to the reference, such as the example highlighted in the yellow circles in Fig. S6(e) and (k). In contrast, the prediction from SFSDAF improved on that from FSDAF by excluding patches of abnormal spectral values and generating an image that was more similar to the reference image in Fig. S6(f) and (l). The quantitative measures in Table S2 show that UBDF usually generated lower RMSE and AAD than STARFM, and STARFM usually generated higher CC and SSIM than UBDF. In contrast, FSDAF decreased RMSE and AAD and increased CC and SSIM compared with UBDF and STARFM. Among all methods, however, SFSDAF generated the most accurate prediction with the lowest RMSE and AAD and highest CC and SSIM values.

Section S3. Temporal prediction comparison in a heterogeneous landscape

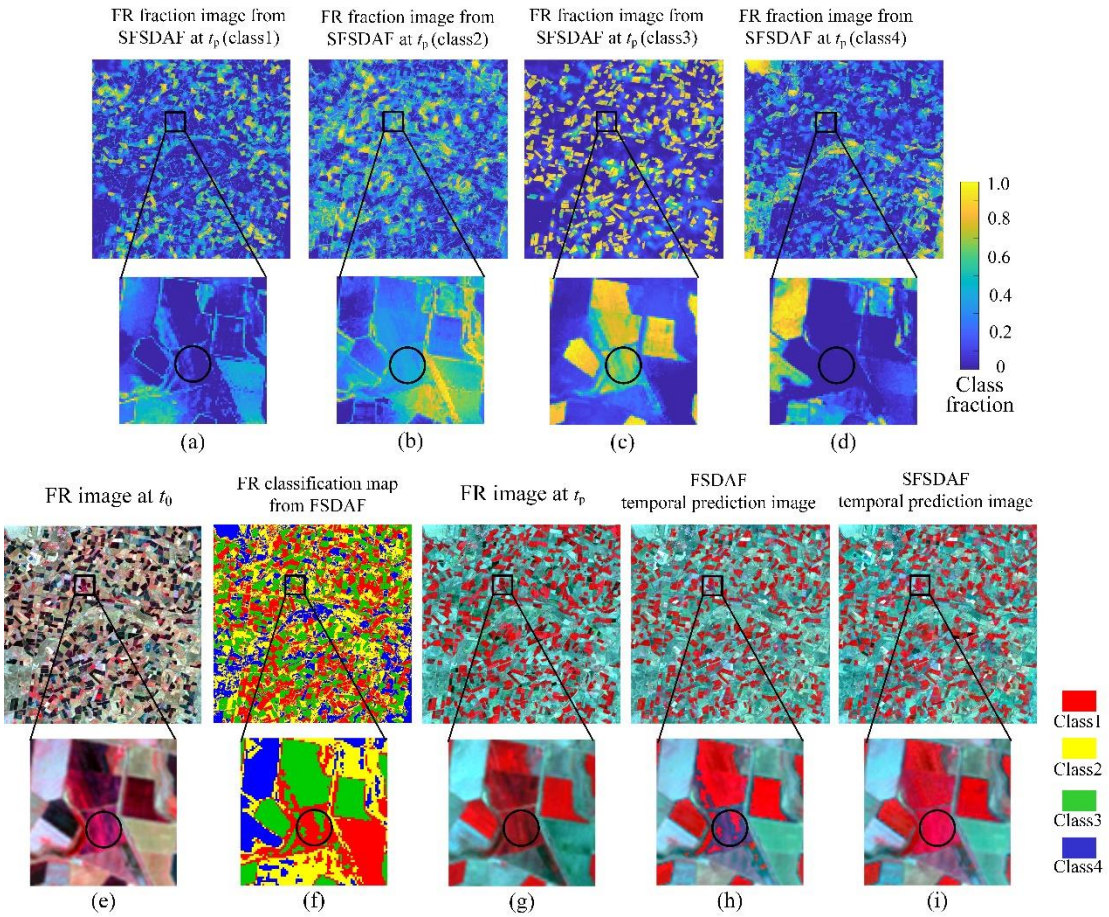

Fig. S7. Comparison of temporal prediction images from FSDAF and SFSDAF in the Landsat and MODIS image experiment in Coleambally, Australia in section S1. The number of classes in FSDAF and SFSDAF was set to 4. (a)-(d) FR class fraction images from SFSDAF at t_p for the 4 classes, (e) FR reflectance image at t_0, (f) FR classification map from FSDAF based on FR image at t_0, (g) FR reflectance image at t_p, (h) FSDAF temporal prediction image, (i) SFSDAF temporal prediction image.

The comparison of FSDAF and SFSDAF temporal prediction for a heterogeneous region in section S1 in Supplementary data was assessed using the Landsat and MODIS-like experiment in Coleambally, Australia. The number of classes was set to four in FSDAF and SFSDAF by the *k*-means method. In the temporal prediction in FSDAF, the change of endmembers was directly added to the prior FR image at t_0 based on the class type in Fig. S7(f). As a result, FR pixels with the same class were added with the same change of endmembers, and FR pixels with the different classes would be added with a different change of endmembers. In Fig. S7(f), the neighboring FR pixels highlighted in a black circle were clustered to “class1” in red and “class3” in green. Since they have different class types, FSDAF would

assign different changes of endmembers. As a result, the FSDAF predicted reflectance was obviously different for pixels of different classes highlighted in a black circle in Fig. S7(h). In contrast to FSDAF, SFSDAF did not use the FR classification, but adopted the FR class fraction images which are spatially continuous in temporal prediction in Fig. S7(a)-(d). SFSDAF multiplied the endmembers with class fractions at t_0 and t_p to account for the total surface reflectance change in temporal prediction in Eq. (18). As a result, the SFSDAF temporal prediction image in Fig. S7(i) captured the reflectance change and were more similar to the reference image than FSDAF highlighted in the black circle in Fig. S7(h). Since this highlighted region had undergone an obvious phenological change, the proposed SFSDAF could improve FSDAF in phenological change prediction in heterogeneous regions. In particular, FSDAF temporal prediction would predict sharp change in reflectance values for neighboring FR pixels if the FR pixels are classified to different classes, whereas the SFSDAF temporal prediction is based on sub-pixel class fraction images and hence does not produce inappropriate land cover patches.

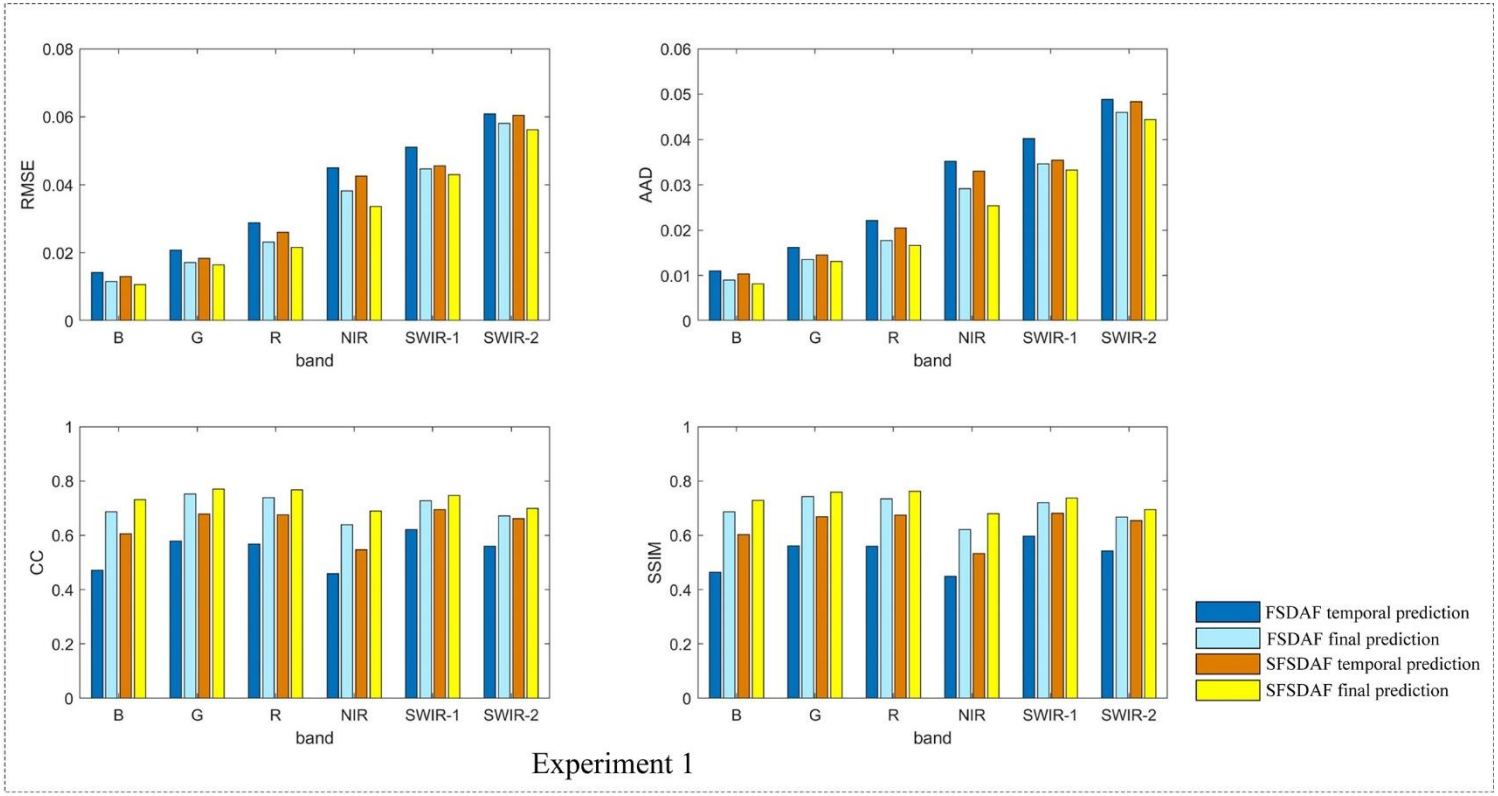

Fig. S8 The accuracies of FSDAF temporal prediction, FSDAF final prediction, SFSDAF temporal prediction and SFSDAF final prediction in Experiment 1.

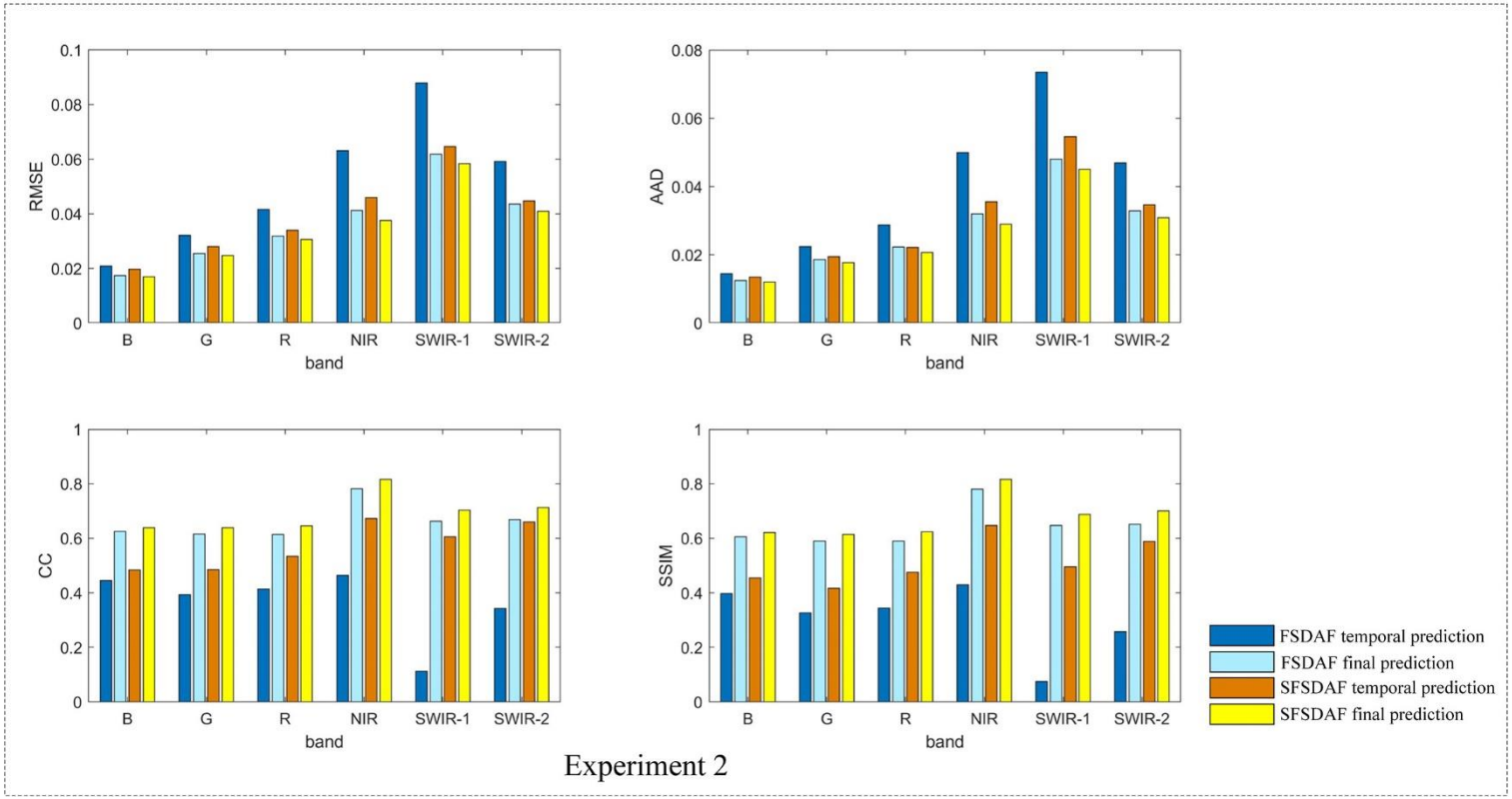

Fig. S9 The accuracies of FSDAF temporal prediction, FSDAF final prediction, SFSDAF temporal prediction and SFSDAF final prediction in Experiment 2.

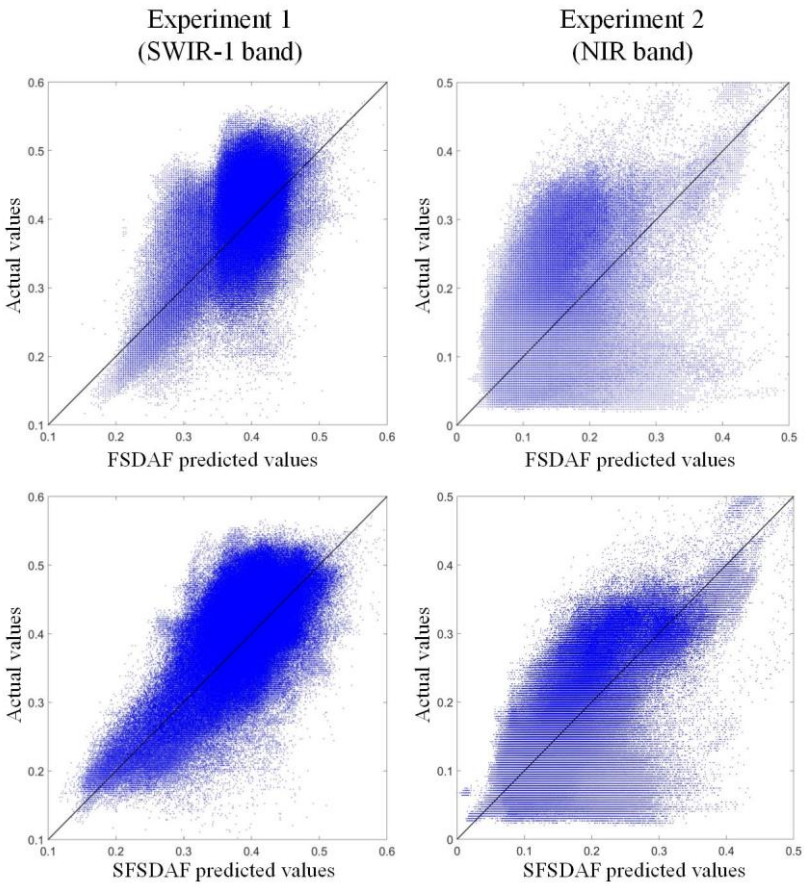

Fig. S10. The scatter plots of temporal prediction images from FSDAF (in the first row) and SFSDAF (in the second row) in the two experiments based on real MODIS images. The Landsat NIR and SWIR-1 bands are selected because they are sensitive to canopy cover and the moisture of soil and vegetation.

Reference：

Alpaydin, E. (1998). Soft vector quantization and the EM algorithm. *Neural Networks, 11*, 467-477

Alves, D.B., Lloveria, R.M., Perez-Cabello, F., & Vlassova, L. (2018). Fusing Landsat and MODIS data to retrieve multispectral information from fire-affected areas over tropical savannah environments in the Brazilian Amazon. *International Journal of Remote Sensing, 39*, 7919-7941

Amoros-Lopez, J., Gomez-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, 23*, 132-141

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

Chapin, F.S., Zavaleta, E.S., Eviner, V.T., Naylor, R.L., Vitousek, P.M., Reynolds, H.L., Hooper, D.U., Lavorel,

S., Sala, O.E., Hobbie, S.E., Mack, M.C., & Diaz, S. (2000). Consequences of changing biodiversity. *Nature, 405*, 234-242

Chen, B., Chen, L., Huang, B., Michishita, R., & Xu, B. (2018). Dynamic monitoring of the Poyang Lake wetland by integrating Landsat and MODIS observations. *Isprs Journal of Photogrammetry and Remote Sensing, 139*, 75-87

Chen, B., Huang, B., & Xu, B. (2017). A hierarchical spatiotemporal adaptive fusion model using one image pair. *International Journal of Digital Earth, 10*, 639-655

Dubrule, O. (1984). Comparing splines and kriging. *Computers & Geosciences, 10*, 327-338

Emelyanova, I.V., McVicar, T.R., Van Niel, T.G., Li, L.T., & van Dijk, A.I.J.M. (2013). Assessing the accuracy of blending Landsat-MODIS surface reflectances in two landscapes with contrasting spatial and temporal dynamics: A framework for algorithm selection. *Remote Sensing of Environment, 133*, 193-209

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., Helkowski, J.H., Holloway, T., Howard, E.A., Kucharik, C.J., Monfreda, C., Patz, J.A., Prentice, I.C., Ramankutty, N., & Snyder, P.K. (2005). Global consequences of land use. *Science, 309*, 570-574

Fu, D., Chen, B., Wang, J., Zhu, X., & Hilker, T. (2013). An improved image fusion approach based on enhanced spatial and temporal the adaptive reflectance fusion model. *Remote Sensing, 5*, 6346-6360

Gaertner, P., Foerster, M., & Kleinschmit, B. (2016). The benefit of synthetically generated RapidEye and Landsat 8 data fusion time series for riparian forest disturbance monitoring. *Remote Sensing of Environment, 177*, 237-247

Gao, F., Anderson, M.C., Zhang, X., Yang, Z., Alfieri, J.G., Kustas, W.P., Mueller, R., Johnson, D.M., & Prueger, J.H. (2017). Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. *Remote Sensing of Environment, 188*, 9-25

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*, 2207-2218

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

Hilker, T., Wulder, M.A., Coops, N.C., Linke, J., McDermid, G., Masek, J.G., Gao, F., & White, J.C. (2009). A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on Landsat and MODIS. *Remote Sensing of Environment, 113*, 1613-1627

Huang, B., & Song, H.H. (2012). Spatiotemporal reflectance fusion via sparse representation. *Ieee Transactions on Geoscience and Remote Sensing, 50*, 3707-3716

Huang, B., & Zhang, H. (2014). Spatio-temporal reflectance fusion via unmixing: accounting for both phenological and land-cover changes. *International Journal of Remote Sensing, 35*, 6213-6233

Ju, J., & Roy, D.P. (2008). The availability of cloud-free Landsat ETM plus data over the conterminous United States and globally. *Remote Sensing of Environment, 112*, 1196-1211

Keshava, N., & Mustard, J.F. (2002). Spectral unmixing. *IEEE Signal Processing Magazine, 19*, 44-57

Keys, R.G. (1981). Cubic convolution interpolation for digital image processing. *Ieee Transactions on Acoustics Speech and Signal Processing, 29*, 1153-1160

Li, X., Du, Y., & Ling, F. (2015). Sub-pixel-scale land cover map updating by integrating change detection and sub-pixel mapping. *Photogrammetric Engineering and Remote Sensing, 81*, 59-67

Li, X., Ling, F., Foody, G.M., & Du, Y. (2016). A superresolution land-cover change detection method using

remotely sensed images with different spatial resolutions. *Ieee Transactions on Geoscience and Remote Sensing, 54*, 3822-3841

Li, X., Ling, F., Foody, G.M., Ge, Y., Zhang, Y., & Du, Y. (2017). Generating a series of fine spatial and temporal resolution land cover maps by fusing coarse spatial resolution remotely sensed images and fine spatial resolution land cover maps. *Remote Sensing of Environment, 196*, 293-311

Liao, C., Wang, J., Pritchard, I., Liu, J., & Shang, J. (2017). A spatio-temporal data fusion model for generating NDVI time series in heterogeneous regions. *Remote Sensing, 9*

Ling, F., Li, W., Du, Y., & Li, X. (2011). Land cover change mapping at the subpixel scale with different spatial-resolution remotely sensed imagery. *Ieee Geoscience and Remote Sensing Letters, 8*, 182-186

Liu, M., Yang, W., Zhu, X., Chen, J., Chen, X., Yang, L., & Helmer, E.H. (2019a). An Improved Flexible Spatiotemporal DAta Fusion (IFSDAF) method for producing high spatiotemporal resolution normalized difference vegetation index time series. *Remote Sensing of Environment, 227*, 74-89

Liu, X., Deng, C., Chanussot, J., Hong, D., & Zhao, B. (2019b). StfNet: A two-stream convolutional neural network for spatiotemporal image fusion. *Ieee Transactions on Geoscience and Remote Sensing, 57*, 6552-6564

Maselli, F., Chiesi, M., & Pieri, M. (2019). A new method to enhance the spatial features of multitemporal NDVI image series. *Ieee Transactions on Geoscience and Remote Sensing, 57*, 4967-4979

Mileva, N., Mecklenburg, S., & Gascon, F. (2018). New tool for spatiotemporal image fusion in remote sensing - a case study approach using Sentinel-2 and Sentinel-3 data. In L. Bruzzone, & F. Bovolo (Eds.), *Image and Signal Processing for Remote Sensing Xxiv*

Schmidt, M., Lucas, R., Bunting, P., Verbesselt, J., & Armston, J. (2015). Multi-resolution time series imagery for forest disturbance and regrowth monitoring in Queensland, Australia. *Remote Sensing of Environment, 158*, 156-168

Song, H., & Huang, B. (2013). Spatiotemporal satellite image fusion through one-pair image learning. *IEEE Transactions on Geoscience and Remote Sensing, 51*, 1883-1896

Song, H., Liu, Q., Wang, G., Hang, R., & Huang, B. (2018). Spatiotemporal satellite image fusion using deep convolutional neural networks. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11*, 821-829

Sun, Y., Zhang, H., & Shi, W. (2019). A spatio-temporal fusion method for remote sensing data using a linear injection model and local neighbourhood information. *International Journal of Remote Sensing, 40*, 2965-2985

Vitousek, P.M., Mooney, H.A., Lubchenco, J., & Melillo, J.M. (1997). Human domination of Earth's ecosystems. *Science, 277*, 494-499

Walker, J.J., de Beurs, K.M., Wynne, R.H., & Gao, F. (2012). Evaluation of Landsat and MODIS data fusion products for analysis of dryland forest phenology. *Remote Sensing of Environment, 117*, 381-393

Wang, J., & Huang, B. (2017). A rigorously-weighted spatiotemporal fusion model with uncertainty analysis. *Remote Sensing, 9*

Wang, Q., & Atkinson, P.M. (2018). Spatio-temporal fusion for daily Sentinel-2 images. *Remote Sensing of Environment, 204*, 31-42

Wang, Q., Shi, W., & Atkinson, P.M. (2016). Spatiotemporal subpixel mapping of time-series images. *Ieee Transactions on Geoscience and Remote Sensing, 54*, 5397-5411

Wang, Q., Zhang, Y., Onojeghuo, A.O., Zhu, X., & Atkinson, P.M. (2017). Enhancing spatio-temporal fusion of MODIS and Landsat data by incorporating 250 m MODIS data. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10*, 4116-4123

Wu, B., Huang, B., & Zhang, L. (2015). An Error-Bound-Regularized Sparse Coding for Spatiotemporal Reflectance Fusion. *Ieee Transactions on Geoscience and Remote Sensing, 53*, 6791-6803

Wu, M., Niu, Z., Wang, C., Wu, C., & Wang, L. (2012). Use of MODIS and Landsat time series data to generate high-resolution temporal synthetic Landsat data using a spatial and temporal reflectance fusion model. *Journal of Applied Remote Sensing, 6*

Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing, 27*, 3025-3033

Xu, Y., & Huang, B. (2014). A spatio-temporal pixel-swapping algorithm for subpixel land cover mapping. *Ieee Geoscience and Remote Sensing Letters, 11*, 474-478

Zhang, L., Weng, Q., & Shao, Z. (2017). An evaluation of monthly impervious surface dynamics by fusing Landsat and MODIS time series in the Pearl River Delta, China, from 2000 to 2015. *Remote Sensing of Environment, 201*, 99-114

Zhang, Y., Foody, G.M., Ling, F., Li, X., Ge, Y., Du, Y., & Atkinson, P.M. (2018). Spatial-temporal fraction map fusion with multi-scale remotely sensed images. *Remote Sensing of Environment, 213*, 162-181

Zhao, Y., Huang, B., & Song, H. (2018). A robust adaptive spatial and temporal image fusion model for complex land surface changes. *Remote Sensing of Environment, 208*, 42-62

Zhong, D., & Zhou, F. (2019). Improvement of clustering methods for modelling abrupt land surface changes in satellite image fusions. *Remote Sensing, 11*

Zhu, X., Cai, F., Tian, J., & Williams, T.K.-A. (2018). Spatiotemporal fusion of multisource remote sensing data: literature survey, taxonomy, principles, applications, and future directions. *Remote Sensing, 10*

Zhu, X., Chen, J., Gao, F., Chen, X., & Masek, J.G. (2010). An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sensing of Environment, 114*, 2610-2623

Zhu, X., Helmer, E.H., Gao, F., Liu, D., Chen, J., & Lefsky, M.A. (2016). A flexible spatiotemporal method for fusing satellite images with different resolutions. *Remote Sensing of Environment, 172*, 165-177

Zhukov, B., Oertel, D., Lanzl, F., & Reinhackel, G. (1999). Unmixing-based multisensor multiresolution image fusion. *Ieee Transactions on Geoscience and Remote Sensing, 37*, 1212-1226

Zurita-Milla, R., Clevers, J.G.P.W., & Schdepman, M.E. (2008). Unmixing-based Landsat TM and MERIS FR data fusion. *Ieee Geoscience and Remote Sensing Letters, 5*, 453-457

Zurita-Milla, R., Kaiser, G., Clevers, J.G.P.W., Schneider, W., & Schaepman, M.E. (2009). Downscaling time series of MERIS full resolution data to monitor vegetation seasonal dynamics. *Remote Sensing of Environment, 113*, 1874-1885