Spatial-temporal Fraction Map Fusion with Multi-scale Remotely Sensed Images

Yihang Zhang [a], Giles M. Foody [b], Feng Ling [a*], Xiaodong Li [a], Yong Ge [c], Yun Du [a], Peter M. Atkinson [d]

a. Key Laboratory of Monitoring and Estimate for Environment and Disaster of Hubei Province, Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan 430077, PR China;

b. School of Geography, University of Nottingham, University Park, Nottingham NG7 2RD, UK;

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

d. Lancaster Environment Center, Faculty of Science and Technology, Lancaster University, Lancaster LA1 4YQ, UK;

(Corresponding author: lingf@whigg.ac.cn)

Abstract: Given the common trade-off between the spatial and temporal resolutions of current satellite sensors, spatial-temporal data fusion methods could be applied to produce fused remotely sensed data with synthetic fine spatial resolution (FR) and high repeat frequency. Such fused data are required to provide a comprehensive understanding of Earth's surface land cover dynamics. In this research, a novel Spatial-Temporal Fraction Map Fusion (STFMF) model is proposed to produce a series of fine-spatial-temporal-resolution land cover fraction maps by fusing coarse-spatial-fine-temporal and fine-spatial-coarse-temporal fraction maps, which may be generated from multi-scale remotely sensed images. The STFMF has two main stages. First, FR fraction change maps are generated using kernel ridge regression. Second, a FR fraction map for the date of prediction is predicted using a temporal-weighted fusion model. In comparison to two established spatial-temporal fusion methods of spatial-temporal super-resolution land cover mapping model and spatial-temporal image reflectance fusion model, STFMF holds the following characteristics and advantages: (1) it takes account of the mixed pixel problem in FR remotely sensed images; (2) it directly uses the fraction maps as input, which could be generated from a range of satellite images or other suitable data sources; (3) it focuses on the estimation of fraction changes happened through time and can predict the land cover change more accurately. Experiments using synthetic multi-scale fraction maps simulated from Google Earth images, as well as synthetic and real MODIS-Landsat images were undertaken to test the performance of the proposed STFMF approach against two benchmark spatial-temporal reflectance fusion methods: the Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model (ESTARFM) and the Flexible Spatiotemporal Data Fusion (FSDAF) model. In both visual and quantitative evaluations, STFMF was able to generate more accurate FR fraction maps and provide more spatial detail than ESTARFM and FSDAF, particularly in areas with substantial land cover changes. STFMF has great potential to produce accurate time-series fraction maps with fine-spatial-temporal-resolution that can support studies of land cover dynamics at the sub-pixel scale.

Keywords: Land cover, fraction maps, spatial-temporal fusion, spectral unmixing, super-resolution mapping.

1. Introduction

With the capabilities of broad spatial coverage and temporally repeated imaging from Earth observation sensors, remote sensing has considerable potential to provide time-series satellite images for studying land surface dynamics (Townshend et al. 1991; Yang and Lo 2002). In heterogeneous areas, land surface dynamics, such as urban expansion, flooding and deforestation, often occur at a fine spatial scale and within a short period. It is, therefore, necessary to collect fine-spatial-temporal-resolution remote sensing images to monitor fine scale land cover changes in a timely manner. Due to the common trade-off between the spatial resolution and the temporal repeat frequency of satellite sensing systems, there is so far no single satellite sensor that can provide remote sensing images with both fine spatial and temporal resolutions (Gao et al. 2006; Li et al. 2017; Zhu et al. 2016). Generally, fine spatial resolution (FR) satellite images are acquired infrequently and have a relatively coarse temporal resolution, making it hard to monitor rapid land cover changes. On the contrary, coarse spatial resolution (CR) satellite sensors acquire data with a high repeat frequency. However, their spatial resolutions are often too coarse to allow the detection of land cover changes occurring in small areas. Therefore, to deal with this dilemma, methods for spatial-temporal data fusion are highly desirable for application to both kinds of remotely sensed imagery to provide remote sensing data with fine spatial and temporal resolutions for studying land surface dynamics (Gao et al. 2006; Gong et al. 2013; Hansen and Loveland 2012; Li et al. 2015; Ling et al. 2016a; Ling et al. 2011; Zhu and Woodcock 2014).

Recently, the spatial-temporal super-resolution mapping (STSRM) method proposed by Ling et al. (2011) has become a promising spatial-temporal fusion method to extract fine spatial and temporal resolution land cover change information (Li et al. 2016; Ling et al. 2016a; Wang et al. 2015; Wu et al.

2017; Xu et al. 2017). STSRM aims to predict a FR land cover map from CR fraction maps, assuming that another FR land cover map, acquired at previous time for the same area, is available. STSRM can be considered as an extension of the traditional super-resolution mapping approach applied to a mono-temporal image, by incorporating information about the land cover changes through time. The key of STSRM is the multi-scale land cover change principle that is using coarse-to-fine resolution change detection between current CR fraction maps and previous FR land cover map to predict the potential locations of current land cover labels of FR land cover map (Ling et al. 2011). The multi-scale land cover change principle in STSRM was further analyzed and assessed by using existing land cover maps, and it has been demonstrated consistently that the principle could be suitable for most current satellite sensors (Ling et al. 2016a). Some popular super-resolution mapping algorithms applied on mono-temporal remote sensing images were also extended to the spatial-temporal domain, leading to various STSRM models (He et al. 2016; Li et al. 2015; Li et al. 2017; Wang et al. 2016; Xu and Huang 2014; Zhang et al. 2017). Compared with the traditional super-resolution mapping methods applied to mono-temporal remote sensing imagery, STSRM can provide details about the spatial distribution of different land cover classes and their changes over time. It is a promising means to produce fine spatial and temporal resolution land cover maps from multi-scale remote sensing imagery.

It is noteworthy that in all existing STSRM models the FR pixels are treated as pure units. That is, the fine pixels within the input and the resultant FR land cover maps are all considered as pure pixels, and each of them is labeled as representing an area comprised of one and only one land cover class. This assumption is reasonable in some cases because the proportion of mixed pixels in an image is typically positively related to pixel size. However, the limitation of this assumption is also obvious, as mixing may still exist in FR image pixels, especially if the land cover mosaic is highly fragmented and heterogeneous.

In practice, the satellite sensor's instantaneous field-of-view often includes more than one land cover feature irrespective of the scale of measurement. Indeed, the mixed pixel problem is widely observed in remote sensing images across different spatial scales (Keshava and Mustard 2002). It is well known that CR remote sensing data, such as those obtained from the Advanced Very High Resolution Radiometer (AVHRR), MEdium Resolution Imaging Spectrometer (MERIS) and MODerate resolution Imaging Spectroradiometer (MODIS) images, contain a large number of mixed pixels. However, the mixed pixel problem is also evident in medium and high spatial resolution satellite sensor images, such as Landsat (Lu and Weng 2004; Powell et al. 2007), ASTER (Weng et al. 2009), IKONOS (Lu and Weng 2009) and Quickbird (Lu et al. 2010), and spectral unmixing techniques may still be needed to obtain fraction maps to enhance the representation of land cover. In this situation, the assumption that all FR pixels are pure in STSRM models may be unreasonable in some real applications.

Another limitation of using the pure pixel assumption in STSRM model is that land cover change information used by it may be partial and possibly erroneous. With the assumption, only one land cover class can be associated with a pixel and hence the only change that can be characterized is that it represents a complete alteration in land cover class: a land cover conversion (e.g. a change from forest to grassland). However, many important land cover changes happed at the sub-pixel scale (finer than the spatial resolution of pixel) may not involve a change in class label. For example, a pixel may represent a forested region which may undergo a substantial change such as a major reduction in tree cover and yet still remain classed as a forest. Changes of the latter type, therefore, do not involve a change in label but a change in the character of the land cover: a land cover modification. Land cover modifications cannot be studied using methods that assume pure pixels but they, and the land cover conversions, can be studied if mixed pixels are allowed such as via the application of soft classification techniques (Foody 2001).

Given the two limitations arising from the pure pixel assumption, error and uncertainty could be introduced in the resultant fine spatial and temporal resolution land cover maps produced by STSRM. Since land cover class fraction values produced by unmixing or soft classification analyses can be used to obtain more accurate land cover information at the sub-pixel scale than discrete land cover labels produced by hard classification (Foody 2002; Foody and Doan 2007), they may have a potential role to play in increasing the accuracy of the STSRM approach.

A different approach to the STSRM for fusing fine-spatial-coarse-temporal and coarse-spatial-fine-temporal remotely sensed images is the spatial-temporal reflectance fusion model. Unlike the STSRM approach that aims to predict land cover class labels at a fine resolution, the spatial-temporal reflectance fusion approach is used to blend reflectance values of remotely sensed images. Gao et al. (2006) first proposed the spatial and temporal adaptive reflectance fusion model (STARFM) to blend Landsat and MODIS reflectance images and produce daily 30 m synthetic Landsat-like reflectance images. Hilker et al. (2009) developed a spatial and temporal adaptive fusion model (STAARCH) to explore spatio-temporal pattern details of forest disturbance based on Landsat and MODIS images. Thereafter, STARFM was developed as an enhanced spatial-temporal adaptive reflectance fusion model (ESTARFM) (Zhu et al. 2010) and a flexible spatio-temporal data fusion (FSDAF) model (Zhu et al. 2016). Moreover, other image spatial temporal fusion models, such as the unmixing based fusion model (Gevaert and Garcia-Haro 2015; Zhukov et al. 1999; Zurita-Milla et al. 2008), the sparse representation based fusion model (Huang and Song 2012; Song and Huang 2013) and spatial and temporal reflectance fusion considering the sensor difference (Shen et al. 2013), have also been proposed. Once the fine spatial and temporal resolution remote sensing images have been produced by the spatial-temporal reflectance image fusion method, a spectral unmixing approach can then be used to produce the corresponding fine spatial and temporal fraction maps. The effectiveness of this approach, however, depends greatly on the spatial-temporal reflectance fusion method, which often suffers from two major limitations when the final objective is to produce fraction maps. First, most spatial-temporal reflectance fusion methods do not account for land cover changes that may have occurred within the period represented by the time-series of remotely sensed images (Gevaert and Garcia-Haro 2015; Zhu et al. 2016). Second, spatial-temporal reflectance fusion methods can generally deal with image pairs with similar spectral bands. Given that many satellite sensors produce images with unique spectral bands, the range of application of these spatial-temporal reflectance fusion methods is thus limited. In comparison, STSRM-based approaches are free from the assumption of sensor-based coherence and can accommodate information on class label change, but not the land cover fraction changes.

In this paper, a novel Spatial-Temporal Fraction Map Fusion (STFMF) model is proposed to generate fraction maps that have a fine resolution in both the spatial and temporal domains by fusing coarse-spatial-fine-temporal and fine-spatial-coarse-temporal remotely sensed images. Critically, the STFMF approach addresses limitations of other methods and hence forms an important contribution to the realization of the potential of remote sensing as a source of information on land cover fraction change. STFMF is based on the fraction maps generated from multi-scale remotely sensed images and uses kernel ridge regression (KRR) to predict FR fraction change maps through time, which are finally used to generate the time-series FR fraction maps with a temporal-weighted model. Compared with the STSRM method, the input and output FR data of STFMF are fraction maps, not the hard land cover class maps used in STSRM model, such that the mixed pixel problem can be dealt with at the fine spatial scale to some extent. Fraction maps with fine resolution have greater superiority than the hard land cover class maps in real applications, such as dynamic monitoring of impervious surfaces (Michishita et al. 2012;

Wu and Murray 2003), tree canopy estimation (Goodwin et al. 2005; Pu et al. 2003) and sub-pixel snow cover mapping (Rosenthal and Dozier 1996; Vikhamar and Solberg 2003), as they have more information at the sub-pixel scale. Compared with the spatial-temporal reflectance fusion approach, such as STARFM and ESTARFM, the proposed STFMF approach is applied directly to land cover fraction maps and could focus more on the fraction land cover changes through time. Meanwhile, there is no need for STFMF to ensure that the collected coarse and fine spatial resolution remote sensing images have similar bands and, thus, a greater number of available pairs of coarse and fine spatial resolution images can be used.

The objectives of this research are three-fold. First, we proposed a new spatial-temporal fraction maps fusion method to produce fraction maps that have a fine resolution in both the spatial and temporal domains, and support more accurate studies of land cover dynamics at the sub-pixel scale. Second, we analyzed the performance and uncertainty of traditional spatial-temporal reflectance fusion approaches for predicting fraction maps. Although the spatial and temporal reflectance fusion approach has been applied widely to produce land cover maps at the per-pixel scale, few studies applied it to produce fraction maps at the sub-pixel scale. This study aims simultaneously to provide a benchmark comparison of their performances in predicting fraction maps. Third, we quantify the proposed approach in revealing spatial-temporal changes at the sub-pixel scale, based on the resultant time-series FR fraction maps within a short period of time (e.g. one month).

2. Methods

The central feature of concern is the prediction of a FR fraction map for a date that lies between dates at which other appropriate remotely sensed are available. Thus, information from imagery that pre- and post-date the date of prediction are critical to spatial-temporal fusion.

2.1 Problem formulation

Let $\mathbf{F}_{T_i}^c$, $\mathbf{F}_{T_p}^c$ and $\mathbf{F}_{T_j}^c$ be the time-series CR fraction maps at previous date T_i, predicted date T_p and posterior date T_j with the same K land cover classes and $M_1 \times M_2$ coarse pixels. In addition, let $\mathbf{F}_{T_i}^f$ and $\mathbf{F}_{T_j}^f$ be the corresponding FR fraction maps at times T_i and T_j with $(M_1 \times z) \times (M_2 \times z)$ fine pixels, where z is the spatial resolution ratio (zoom factor) between the coarse and fine spatial resolution fraction maps. Note that the superscripts *f* and *c* indicate the fine and coarse spatial resolution fraction maps respectively. The objective of the proposed STFMF approach is to predict the FR fraction maps $\mathbf{F}_{T_p}^f$ from the available CR fraction maps $\mathbf{F}_{T_p}^c$, with the aid of pre- and post-date coarse and fine spatial resolution fraction maps, that is, $\mathbf{F}_{T_i}^c$, $\mathbf{F}_{T_j}^c$, $\mathbf{F}_{T_i}^f$ and $\mathbf{F}_{T_j}^f$. Note that the data and methods used to generate the time-series coarse and fine spatial resolution fraction maps $\mathbf{F}_{T_i}^c$, $\mathbf{F}_{T_p}^c$, $\mathbf{F}_{T_j}^c$, $\mathbf{F}_{T_i}^f$ and $\mathbf{F}_{T_j}^f$ are not specific. They can, for example, be produced from existing datasets or produced from corresponding remote sensing images (e.g., CR MODIS and FR Landsat images) through the use of a soft classification (Foody et al. 1997), a spectral unmixing model such as linear spectral mixture model (LSM) (Adams et al. 1986) or a multiple endmember spectral mixture analysis model (Powell et al. 2007).

One possible way to obtain the FR fraction maps $\mathbf{F}_{T_p}^f$ is to downscale the CR fraction maps $\mathbf{F}_{T_p}^c$ to the target fine resolution through the use of an appropriate spatial interpolation approach. With this approach, however, the spatial and temporal prior information within pre- (e.g. T_i) and post-date of prediction (e.g. T_j) coarse and fine spatial resolution fraction maps cannot be utilized. Moreover, the outcome of spatial interpolation is to some extent a smoothed representation, which would lead to edge blur and ringing effects around the boundaries of different land cover features. In general, during the period between T_i and T_j , fraction values of different land cover classes at the FR may have changed from those in $\mathbf{F}_{T_i}^f$ to those in $\mathbf{F}_{T_k}^f$, and may also have changed to those in $\mathbf{F}_{T_j}^f$. As the fraction values in $\mathbf{F}_{T_i}^f$ and $\mathbf{F}_{T_j}^f$ are inputs, if we can predict the changes of FR fraction values of different land cover classes between $\mathbf{F}_{T_p}^f$ and $\mathbf{F}_{T_i}^f$ or $\mathbf{F}_{T_p}^f$ and $\mathbf{F}_{T_i}^f$, the FR fraction maps $\mathbf{F}_{T_p}^f$ can thus be predicted.

Let $F_{T_i}^f(k)$ be the FR fraction map of kth land cover class in $\mathbf{F}_{T_i}^f$ and $F_{T_i}^c(k)$ be the CR fraction map of kth land cover class in $\mathbf{F}_{T_i}^c$. Assuming that the CR fraction map $F_{T_i}^c(k)$ has been geo-referenced to the coordinate system of the FR fraction map $F_{T_i}^f(k)$, and $F_{T_i}^c(\uparrow_z, k)$ is the FR fraction maps which has been downscaled to the spatial resolution of $F_{T_i}^f(k)$ with a downscaling method. The relationship between $F_{T_i}^c(\uparrow_z, k)$ and $F_{T_i}^f(k)$ could be formulated as

$$F_{T_i}^f(k) = F_{T_i}^c(\uparrow_z, k) + \varepsilon_{T_i}(k), \quad \forall k = 1, 2, \ldots, K, \tag{1}$$

in which $\uparrow_z$ indicates a downscaling operation used to increase the spatial resolution (i.e. make pixel size smaller) of $F_{T_i}^c(k)$ to that of $F_{T_i}^f(k)$, and $\varepsilon_{T_i}(k)$ is denoted as the fraction difference between $F_{T_i}^f(k)$ and $F_{T_i}^c(k)$. It is noteworthy that fraction map is not a physical variable directly observed by satellite sensors and generally produced from satellite images at different spatial resolutions. Therefore, the fraction difference $\varepsilon_{T_i}(k)$ between $F_{T_i}^f(k)$ and $F_{T_i}^c(k)$ is associated with differences between the data sources, the means of endmember selection and the spectral unmixing methods used in the generation of the fine and coarse spatial resolution fraction maps. Likewise, the relationship shown in equation (1) applies equally at T_p , and is expressed as

$$F_{T_p}^{f}(k)=F_{T_p}^{c}(\uparrow_z, k)+\varepsilon_{T_p}(k), \quad \forall k=1,2,\ldots,K . \tag{2}$$

In this section, it is assumed that the data source, principles of endmember selection and spectral unmixing method for the generation of fine and coarse spatial resolution fraction maps in equations (1) and (2) at time T_i are the same at T_p. $\varepsilon_{T_p}(k)$ at time T_p is thus considered unchanged by comparing with $\varepsilon_{T_i}(k)$ at time T_i. Therefore, combining equations (1) and (2), the estimation of FR fraction maps $F_{T_p}^{f}(k)$ can be expressed as

$$F_{T_p}^{f}(k)=F_{T_i}^{f}(k)+(F_{T_p}^{c}(\uparrow_z, k)-F_{T_i}^{c}(\uparrow_z, k)), \quad \forall k=1,2,\ldots,K . \tag{3}$$

Denote $\Delta_{T_iT_p}^{f}(k)=F_{T_p}^{c}(\uparrow_z, k)-F_{T_i}^{c}(\uparrow_z, k)$ as the kth land cover fraction change map with spatial resolution equal to that of $F_{T_i}^{f}(k)$ and $F_{T_p}^{f}(k)$, and $\Delta_{T_iT_p}^{c}(k)=F_{T_p}^{c}(k)-F_{T_i}^{c}(k)$ as the CR fraction change map of the kth land cover class. Assume that $F_{T_i}^{f}(k)$ at time T_i is known, the estimation of FR fraction map $F_{T_p}^{f}(k)$ becomes a key process of predicting the FR fraction change map $\Delta_{T_iT_p}^{f}(k)$ from the CR fraction change map $\Delta_{T_iT_p}^{c}(k)$.

Likewise, for equation (3), fine and coarse spatial resolution fraction maps $F_{T_i}^{f}(k)$ and $F_{T_i}^{c}(k)$ at pre-time T_i could be replaced as fraction maps $F_{T_j}^{f}(k)$ and $F_{T_j}^{c}(k)$ at post-time T_j. The estimation of $F_{T_p}^{f}(k)$ is, therefore, to predict the FR fraction change map $\Delta_{T_iT_p}^{f}(k)$ or $\Delta_{T_pT_j}^{f}(k)$ from the observed CR fraction change map $\Delta_{T_iT_p}^{c}(k)$ or $\Delta_{T_pT_j}^{c}(k)$ according to equation (3). Note that the corresponding CR fraction change maps $\Delta_{T_iT_p}^{c}(k)$, $\Delta_{T_pT_j}^{c}(k)$ and $\Delta_{T_iT_j}^{c}(k)$ can be calculated from the known CR fraction maps $F_{T_i}^{c}(k)$, $F_{T_p}^{c}(k)$ and $F_{T_j}^{c}(k)$. $\Delta_{T_iT_p}^{c}(k)$, $\Delta_{T_pT_j}^{c}(k)$ and $\Delta_{T_iT_j}^{c}(k)$ are, therefore, expressed as

$$\Delta_{T_iT_p}^{c}(k)=F_{T_p}^{c}(k)-F_{T_i}^{c}(k), \tag{4}$$

$$\Delta_{T_pT_j}^{c}(k)=F_{T_j}^{c}(k)-F_{T_p}^{c}(k), \tag{5}$$

$$\Delta_{T_iT_j}^{c}(k)=F_{T_j}^{c}(k)-F_{T_i}^{c}(k). \tag{6}$$

Moreover, the FR fraction change maps $\Delta^{f}_{T_iT_j}(f)$ can be calculated from the known FR fraction maps $F^{f}_{T_i}(k)$ and $F^{f}_{T_j}(k)$, expressed as

$$\Delta^{f}_{T_iT_j}(k) = F^{f}_{T_j}(k) - F^{f}_{T_i}(k)\,. \tag{7}$$

Therefore, according to equations (6) and (7), a coarse and fine spatial resolution fraction change maps pair $\left[\Delta^{c}_{T_iT_j}(k), \Delta^{f}_{T_iT_j}(k)\right]$ can be obtained, where $k = 1, \ldots, K$. Assuming that the relationships between the coarse and fine spatial resolution fraction maps pairs $\left[\Delta^{c}_{T_iT_p}(k), \Delta^{f}_{T_iT_p}(k)\right]$ and $\left[\Delta^{c}_{T_pT_j}(k), \Delta^{f}_{T_pT_j}(k)\right]$ are similar to those of $\left[\Delta^{c}_{T_iT_j}(k), \Delta^{f}_{T_iT_j}(k)\right]$, the FR fraction change maps $\Delta^{f}_{T_iT_p}(k)$ and $\Delta^{f}_{T_pT_j}(k)$ can then be predicted from $\Delta^{c}_{T_iT_p}(k)$ and $\Delta^{c}_{T_iT_j}(k)$, respectively.

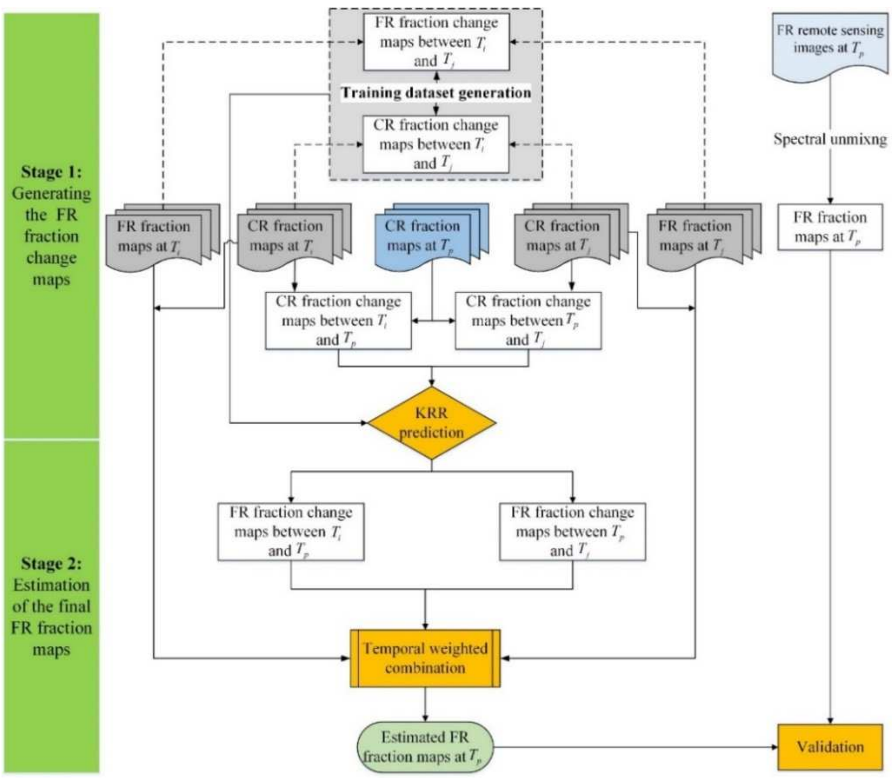


Figure 1. Flowchart of the proposed the proposed STFMF approach.

Fig. 1 shows the whole flowchart of the proposed STFMF approach. Fig 1 highlights especially that the model inputs are the coarse and fine spatial resolution fraction map pairs at dates that pre- and post- the date of prediction together with the CR fraction maps for the date of prediction. STFMF is composed of two main stages: generating FR fraction change maps and estimation of the final FR fraction maps.

2.2 Generating FR fraction change maps

The estimation of fine resolution fraction change maps $\Delta^f_{T_iT_p}(k)$ and $\Delta^f_{T_pT_j}(k)$ from $\Delta^c_{T_iT_p}(k)$ and $\Delta^c_{T_pT_j}(k)$ can be considered as an image reconstruction process, and can generally be achieved via spatial interpolation or image super-resolution approaches (Kim and Kwon 2010; Ni and Nguyen 2007). In this research, a super-resolution reconstructing approach based on kernel ridge regression (KRR) was applied (Kim and Kwon 2010). The first step of this approach is to learn the relationship between the coarse and fine spatial resolution fraction change maps pair $\left[\Delta^c_{T_iT_j}(k), \Delta^f_{T_iT_j}(k)\right]$. Then, the learned relationship is applied to estimate the FR fraction change maps $\Delta^f_{T_iT_p}(k)$ and $\Delta^f_{T_pT_j}(k)$ from $\Delta^c_{T_iT_p}(k)$ and $\Delta^c_{T_pT_j}(k)$ respectively. In the super-resolution reconstruction process, the FR fraction change maps are estimated class by class, and it has three main steps: training dataset generation, candidate neighbors search and fine image patch reconstruction.

2.2.1 Training dataset generation

The training dataset is used to obtain the relationship between the coarse and fine spatial resolution images. Instead of directly using the whole coarse and fine spatial resolution fraction change maps pair $\left[\Delta^c_{T_iT_j}(k), \Delta^f_{T_iT_j}(k)\right]$, image patch pairs generated from them are used as the training dataset. As shown in Fig. 2, an example is used here to illustrate the generation process of image patch pairs in training dataset, where the spatial ratio z is set to be 4 and the window size P is set to be 3. The image patch pairs are composed of a large number of small sized coarse and fine spatial resolution image patch pairs extracted from corresponding fraction change maps of $\Delta^c_{T_iT_j}(k)$ and $\Delta^f_{T_iT_j}(k)$. Let $X_{T_{ij},k} = \{x^m_{T_{ij},k}\}_{m=1}^{M_1\times M_2}$ be the

CR image patch sets generated from the kth CR fraction change map of $\Delta^c_{T_iT_j}(k)$, and $x^m_{T_{ij},k}$ be the mth CR image patch that is expressed as

$$x^m_{T_{ij},k} = [f^m_{T_{ij},k}(1), f^m_{T_{ij},k}(2), \cdots, f^m_{T_{ij},k}(P \times P)], \tag{8}$$

where P is the square window size of the CR image patch and $f^m_{T_{ij},k}(V)$ is the kth fraction change value of coarse pixel V in the mth CR image patch. Let $Y_{T_{ij},k} = \{y^m_{T_{ij},k}\}^{M_1 \times M_2}_{m=1}$ be the FR image patch sets generated from the kth fraction change map of $\Delta^f_{T_iT_j}$, and $y^m_{T_{ij},k}$ be the mth FR patch that is

$$y^m_{T_{ij},k} = [I^m_{T_{ij},k}(1), I^m_{T_{ij},k}(2), \cdots, I^m_{T_{ij},k}(z \times z)], \tag{9}$$

where $I^m_{T_{ij},k}(v)$ is the kth fraction change value of the fine pixel v in the mth FR image patch.

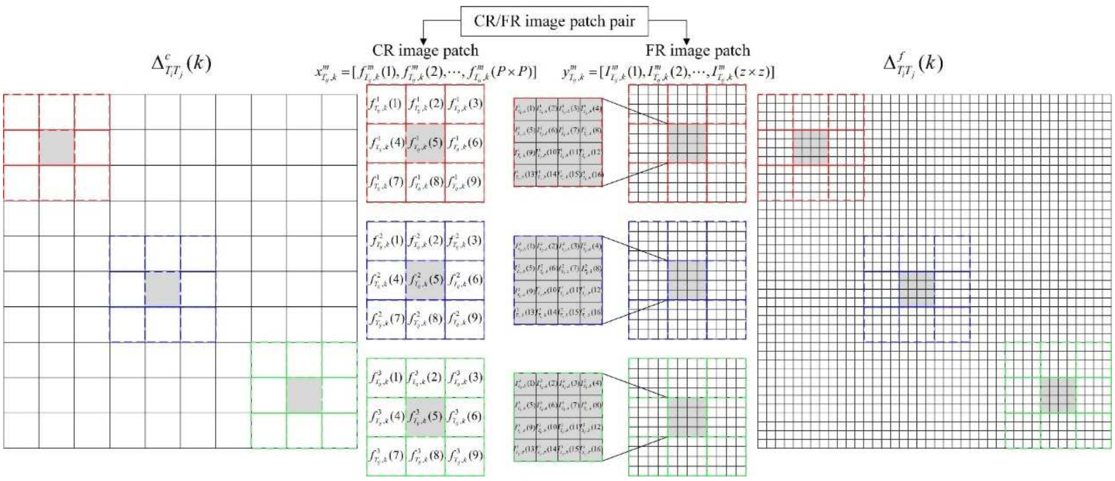


Figure 2. An example of a coarse and fine spatial resolution image patch pair in the training dataset.

As shown in Fig. 2, $y^m_{T_{ij},k}$ contains $z \times z$ fine pixels within the mth central coarse pixel, and $x^m_{T_{ij},k}$ contains $P \times P$ coarse pixels which is composed of the mth central pixel and neighboring $P \times P - 1$ coarse pixels. Training dataset is denoted as $\left[X_{T_{ij},k}, Y_{T_{ij},k}\right]$ which is composed of the image pairs of CR image patches $X_{T_{ij},k}$ and FR image patches $Y_{T_{ij},k}$ for land cover class k, where $X_{T_{ij},k} \in \Delta^c_{T_iT_j}$ and $Y_{T_{ij},k} \in \Delta^f_{T_iT_j}$. Therefore, there is a total of $M_1 \times M_2$ image patch pairs in the training dataset $\left[X_{T_{ij},k}, Y_{T_{ij},k}\right]$. More information about the training dataset generating process could be found in Zhang et al. (2014) and Ling et al. (2016b).

2.2.2 Searching for candidate neighboring patch pairs

To reconstruct the FR fraction change maps $\Delta^{f}_{T_iT_p}(k)$ and $\Delta^{f}_{T_pT_j}(k)$ by using the training dataset $\left[X_{T_{ij},k}, Y_{T_{ij},k}\right]$, similar CR and FR patch pairs need to be searched from the training dataset for each CR patch in the CR fraction change maps $\Delta^{c}_{T_iT_p}(k)$ and $\Delta^{c}_{T_pT_j}(k)$. Let $X_{T,k} = \{x^m_{T,k}\}_{m=1}^{M_1 \times M_2}$ be the CR patches dataset generated from the input kth CR fraction change maps of $\Delta^{c}_{T_iT_p}(k)$ or $\Delta^{c}_{T_pT_j}(k)$. For a certain CR patch $x^m_{T,k}$, similar CR patches in the training dataset $\left[X_{T_{ij},k}, Y_{T_{ij},k}\right]$ can be searched according to the following criterion

$$\Delta f(x^m_{T,k}, x^m_{T_{ij},k}) = \sqrt{\frac{1}{P \times P}\sum_{V=1}^{P \times P}(f^m_{T,k}(V) - f^m_{T_{ij},k}(V))^2} < \theta, \tag{10}$$

where $\Delta f(x^m_{T,k}, x^m_{T_{ij},k})$ is the difference of fraction change values between the CR patch $x^m_{T,k}$ in $\Delta^{c}_{T_iT_p}(k)$ or $\Delta^{c}_{T_pT_j}(k)$ and $x^m_{T_{ij},k}$ in the training dataset $\left[X_{T_{ij},k}, Y_{T_{ij},k}\right]$. $f^m_{T,k}(V)$ is the fraction change value of pixel V in CR patch $x^m_{T,k}$, and $f^m_{T_{ij},k}(V)$ is the fraction change value of the corresponding pixel V in CR patch $x^m_{T_{ij},k}$. The more similar the patches $x^m_{T,k}$ and $x^m_{T_{ij},k}$, the less the value of Δf. The threshold θ is a pre-defined parameter that is the tolerable fraction difference between two patches.

If the Δf between patches $x^m_{T_{ij},k}$ and $x^m_{T,k}$ is less than the threshold value θ, $x^m_{T_{ij},k}$ in training dataset is thus considered as the neighboring patch of $x^m_{T,k}$. It is assumed that if the CR patches $x^m_{T_{ij},k}$ and $x^m_{T,k}$ have a similar spatial pattern, their corresponding FR patches $y^m_{T_{ij},k}$ and predicted $y^m_{T,k}$ should also be similar to each other (Freeman et al. 2002). $\left[x^m_{T_{ij},k}, y^m_{T_{ij},k}\right]$ is thus regarded as the candidate neighboring patch pair for CR patch $x^m_{T,k}$ in $\Delta^{c}_{T_iT_p}(k)$ or $\Delta^{c}_{T_pT_j}(k)$. It is noteworthy that only one candidate neighboring patch pair is always insufficient for the predicting of FR patch $y^m_{T,k}$. We assume that N candidate neighboring image patch pairs, which are represented as $\left\{x^m_{T_{ij},k}(l), y^m_{T_{ij},k}(l)\right\}_{l=1}^{N}$, have been searched from the training dataset $\left[X_{T_{ij},k}, Y_{T_{ij},k}\right]$.

However, different CR patch should have different threshold value θ, and it is almost infeasible to define a fixed θ to search *N* candidate neighboring image patch pairs for various CR patches. An alternative solution for this is to directly find the *N* nearest neighboring patches from the training dataset for each CR patch. It is assumed that if there were enough elements in the training dataset, the searched nearest neighboring patches would be regarded as the *N* candidate neighboring image patch pairs. K-D tree search algorithm (Bentley 1975; Freeman et al. 2002) is used here to find the *N* nearest neighboring patches from the training dataset, as it holds the advantages of simple and efficient. K-D tree search algorithm first builds a K-D tree struct (with $M_1 \times M_2$ elements) from the training dataset $\left[X_{T_{ij},k}, Y_{T_{ij},k}\right]$. Δf, which are values between all of the $M_1 \times M_2$ CR patch $x^m_{T_{ij},k}$ and each input CR patch $x^m_{T,k}$, are then calculated. Finally, all of the Δf values are arranged in an ascending order, and the first *N* elements are, therefore, regarded as the candidate neighboring image patch pairs. The searched *N* candidate neighboring image patch pairs $\left\{x^m_{T_{ij},k}(l), y^m_{T_{ij},k}(l)\right\}_{l=1}^{N}$ are used to reconstruct the latent FR fraction change maps $\Delta^f_{T_iT_p}(k)$ and $\Delta^f_{T_pT_j}(k)$.

2.2.3 FR fraction change map estimation with KRR

Let $y^m_{T,k}$ be the HR patch of the input LR patch $x^m_{T,k}$ extracted from the CR fraction change map $\Delta^c_{T_iT_p}(k)$ or $\Delta^c_{T_pT_j}(k)$, $y^m_{T_{ij},k}$ and $x^m_{T_{ij},k}$ be the HR and LR patch pair extracted from the FR and CR fraction change maps $\Delta^c_{T_iT_j}(k)$ and $\Delta^f_{T_iT_j}(k)$. If the root mean square error between $x^m_{T,k}$ and $x^m_{T_{ij},k}$ is lower than a value (e.g. 0.10), it is considered that $y^m_{T,k}$ is equal to $y^m_{T_{ij},k}$. Otherwise, the estimation of $y^m_{T,k}$ is based on the similar neighbors searched from candidate image patch pairs $\left\{x^m_{T_{ij},k}(l), y^m_{T_{ij},k}(l)\right\}_{l=1}^{N}$. Since $x^m_{T,k}$ and $x^m_{T_{ij},k}(l)$ are similar, we also consider that the spatial distribution information of the predicted $y^m_{T,k}$ within $x^m_{T,k}$ should be similar to that of $y^m_{T_{ij},k}(l)$ within $x^m_{T_{ij},k}(l)$. Given the searched similar image patch pairs $\left\{x_{T_{ij},k}^{m}(l), y_{T_{ij},k}^{m}(l)\right\}_{l=1}^{N}$, the machine learning approach of KRR (Kim and Kwon 2010) is applied here to estimate the FR fraction change image patch $y_{T,k}^{m}$.

Assume a function model $y = f(x) + w$, where w is the estimation noise, x is the input variable and y is the corresponding regression value, KRR aims to estimate the regression function f. Given a set of training data $\left\{(x_{T_{ij},k}^{m}(1), y_{T_{ij},k}^{m}(1)), \cdots, (x_{T_{ij},k}^{m}(N), y_{T_{ij},k}^{m}(N))\right\}$, we can estimate $\hat{f}$ by solving an optimization problem:

$$\hat{f} = \arg\min_{f \in H} \frac{1}{2}\sum_{m=1}^{N}(y_{T_{ij},k}^{m} - f(x_{T_{ij},k}^{m}))^2 + \frac{\lambda}{2}\|f\|_H^2, \tag{11}$$

where H is a kernel Hilbert space with kernel K, and λ is a regularization constant parameter. The first term of equation (11) is the data fidelity term, while the second is the regularization term. Then the optimal solution for $\hat{f}$ from equation (11) has the following form:

$$\hat{f}(\cdot) = \sum_{m=1}^{N}\alpha_m \cdot K(\cdot, x_{T_{ij},k}^{m}), \tag{12}$$

$$\|f\|_H^2 = \sum_{n,m=1}^{N}\alpha_n\alpha_m K(x_{T_{ij},k}^{n}, x_{T_{ij},k}^{m}). \tag{13}$$

Let $\mathbf{y} = [y_{T_{ij},k}^{1}, \cdots, y_{T_{ij},k}^{N}]$ and $\mathbf{K} = K_{nm} = K(x_{T_{ij},k}^{n}, x_{T_{ij},k}^{m})$, and then the original optimization problem shown in equation (11) is formulated as:

$$\hat{\alpha} = \arg\min \frac{1}{2}\|\mathbf{y} - \mathbf{K}\alpha\|_2^2 + \frac{\lambda}{2}\alpha^T\mathbf{K}\alpha, \tag{14}$$

by calculating the gradient of equation (14), we can obtain the following equation:

$$\nabla C(\alpha) = -\mathbf{K}\mathbf{y} + \mathbf{K}^2\mathbf{y}\alpha + \lambda\mathbf{K}\alpha = 0. \tag{15}$$

One solution for equation (15) is $\hat{\alpha} = (\mathbf{K} + \lambda\mathbf{I})^{-1}\mathbf{y}$, and this is the only solution due to the form of $\hat{f}(\cdot)$. Therefore, the estimate of $\hat{f}(\cdot)$ is:

$$\hat{f}(\cdot) = \sum_{n=1}^{N}\hat{\alpha}_n K(\cdot, x_{T_{ij},k}^{n}), \tag{16}$$

In this research, the kernel function K is based on a Gaussian kernel and is presented as:

$$K(s,t)=\exp\left(-\frac{\|s-t\|^2}{\delta}\right). \tag{17}$$

Therefore, for any input LR image patch $x^m_{T,k}$, the corresponding FR image patch $y^m_{T,k}$ can be predicted by equation (16). Once the FR image patches dataset $\{y^m_{T,k}, m=1,\cdots,M_1\times M_2\}$ has been produced, the FR fraction change maps $\Delta^f_{T_iT_j}(k)$ and $\Delta^f_{T_pT_j}(k)$ can then be produced by merging the FR image patches with a spatial averaging filter. More information about the merging process is presented in Zhang et al. (2015).

2.3 Final FR fraction map estimation

With the estimated FR fraction change maps $\mathbf{\Delta}^f_{T_iT_p}$ and $\mathbf{\Delta}^f_{T_pT_j}$, the final FR fraction maps $\mathbf{F}^f_{T_p}$ can, thus, be predicted using equation (3). To take advantage of the predicted results being based on the FR fraction maps $\mathbf{F}^f_{T_i}$ and $\mathbf{F}^f_{T_j}$ that respectively pre- and post-date it, a temporal weighted model is used here to predict $\mathbf{F}^f_{T_p}$. In the absence of knowledge on the land cover changes, the model is based on the assumption that the FR fraction maps at time T_p are a linearly weighted combination of the FR fraction maps and corresponding FR fraction change maps at both pre- and post-time T_i and T_j. Consequently, $\mathbf{F}^f_{T_p}$ is predicted as:

$$\mathbf{F}^f_{T_p}=\frac{\mathbf{c}_{T_pT_j}}{\mathbf{c}_{T_iT_p}+\mathbf{c}_{T_pT_j}}\cdot(\mathbf{F}^f_{T_i}+\mathbf{\Delta}^f_{T_iT_p})+\frac{\mathbf{c}_{T_iT_p}}{\mathbf{c}_{T_iT_p}+\mathbf{c}_{T_pT_j}}\cdot(\mathbf{F}^f_{T_j}+\mathbf{\Delta}^f_{T_pT_j}), \tag{18}$$

where $\mathbf{c}_{T_1T_2}=[c^1_{T_iT_p},\cdots,c^n_{T_iT_p}]$ is the change ratio vector between fraction maps $\mathbf{F}^c_{T_i}$ and $\mathbf{F}^c_{T_p}$, and $\mathbf{c}_{T_pT_j}=[c^1_{T_pT_j},\cdots,c^n_{T_pT_j}]$ is the change ratio vector between fraction maps $\mathbf{F}^c_{T_p}$ and $\mathbf{F}^c_{T_j}$. $c^k_{T_iT_p}$ and $c^k_{T_pT_j}$ are the change ratio between fraction maps of the kth land cover class ($k\in 1,\cdots,K$), and they are presented as

$$c^k_{T_iT_p}=\sqrt{\frac{1}{M_1\times M_2}\sum_{V=1}^{M_1\times M_2}(f_{T_i,k}(V)-f_{T_p,k}(V))^2}, \tag{19}$$

$$c_{T_pT_j}^k = \sqrt{\frac{1}{M_1 \times M_2} \sum_{V=1}^{M_1 \times M_2} (f_{T_p,k}(V) - f_{T_j,k}(V))^2}\ , \tag{20}$$

where $f_{T_i,k}(V)$, $f_{T_p,k}(V)$ and $f_{T_j,k}(V)$ are the fraction values for coarse pixel V in the kth fraction maps $\mathbf{F}_{T_i}^c$, $\mathbf{F}_{T_p}^c$ and $\mathbf{F}_{T_j}^c$. Since $\mathbf{c}_{T_iT_p}$ and $\mathbf{c}_{T_pT_j}$ can be calculated from $\mathbf{F}_{T_i}^c$, $\mathbf{F}_{T_p}^c$ and $\mathbf{F}_{T_j}^c$, and $\mathbf{F}_{T_i}^f$ and $\mathbf{F}_{T_j}^f$ are already known, the final FR fraction map $\mathbf{F}_{T_p}^f$ can be predicted once the FR fraction change maps $\mathbf{\Delta}_{T_iT_p}^f$ and $\mathbf{\Delta}_{T_pT_j}^f$ have been estimated.

Theoretically, fraction values of the different land cover classes in the predicted FR fraction maps $\mathbf{F}_{T_p}^f$ should be in the range of 0 and 1, and the sum of fraction values of different land cover class for each fine pixel in $\mathbf{F}_{T_p}^f$ should be exactly 1. To make the resultant FR fraction maps $\mathbf{F}_{T_p}^f$ satisfy both restrictions, a normalization operation is further applied. Let $I_{T_p,k}(v)$ be the fraction value of fine pixel v in the kth fraction map of original predicted $\mathbf{F}_{T_p}^f$ and $I_{T_p,k}^*(v)$ be the corrected fraction values in the normalized $\mathbf{F}_{T_p}^f$, and $I_{T_p,k}^*(v)$ be expressed as

$$I_{T_p,k}^*(v) = \frac{I_{T_p,k}(v)}{\sum_{k=1}^{K} I_{T_p,k}(v)}. \tag{21}$$

2.4 Accuracy Assessment

Four indices are used for the quantitative evaluation of the resultant FR fraction maps obtained from the various approaches: the correlation coefficient (CC), root mean square error (RMSE), absolute average difference (AAD), and universal image quality index (UIQI) (Wang and Bovik 2002). The CC index indicates the degree of correlation (or similarity) between the predicted and reference fraction maps, and its value lies in the range of 0 and 1, where a larger value means a better match. By contrast, RMSE reflects the difference between the predicted and reference fraction maps with small RMSE values indicating a closer match, the ideal value of RMSE is 0. AAD is used to assess the average bias of the individual predicted fraction maps, with small values indicating high quality. UIQI accounts for an estimation of CC and differences in the mean luminance and contrast, and it was designed to overcome some limitations of RMSE (Vivone et al. 2015). UIQI varies in the range of -1 to 1, and larger values denote better fidelity to the reference fraction maps.

3. Experiments and results

To assess the performance of the proposed STFMF approach, two experiments based on the synthetic fraction maps simulated from Google Earth images (GEI), as well as synthetic and real MODIS-Landsat images for study areas with different land cover mosaics were undertaken. In the first experiment, the input fraction maps were simulated by downscaling the FR GEI land cover maps. In the second experiment, the input fraction maps were generated from the MODIS and Landsat images using the linear spectral mixture (LSM) model (Keshava and Mustard 2002). To implement the LSM model in the MODIS-Landsat experiment, spectral endmembers were obtained using the Pixel Purity Index algorithm (Chang and Plaza 2006) and manual selection, and the fully constrained least squares spectral unmixing analysis (Heylen et al. 2011) was applied to generate fraction maps from the MODIS and Landsat images.

Two popular spatial-temporal reflectance fusion algorithms, that is, ESTARFM (Zhu et al. 2010) and FSDAF (Zhu et al. 2016), are used as the comparative methods against which the performance of STFMF was evaluated. ESTARFM needs two pairs of CR and FR remotely sensed reflectance images, and both coarse and fine spatial resolution remotely sensed reflectance images at T_i and T_j were used as the input. FSDAF needs only one reflectance image pair. To have a comprehensive comparison, FSDAF based on the reflectance image pair at T_i and FSDAF based on the image pair at T_j were applied as the comparative methods.

3.1 The GEI experiment

The study area of this experiment is Wuhan city, China. With the FR (5 m) GEIs [see Figs. 3(a)-(c)] acquired on April 24, 2012, December 20, 2014 and February 20, 2016, the corresponding FR land cover maps, as shown in Figs. 3 (d)-(f), were generated by manually digitizing. Each of the land cover maps includes four land cover classes of water, vegetation, bareland and impervious surface. Then, the 30 m Landsat-like fraction maps and the 480 m MODIS-like fraction maps were simulated from the FR land cover maps by spatial degrading. The original GEI contains 1920 × 1920 pixels, and thus the Landsat-like fraction map contains 320 × 320 pixels and the MODIS-like fraction map contains 20 × 20 pixels. The MODIS-like fraction maps at 2014 were used as the input CR images at the predicted time (e.g. T_p). ESTARFM, FSDAF and STFMF were then applied to produce the Landsat-like FR fraction maps at 2014.

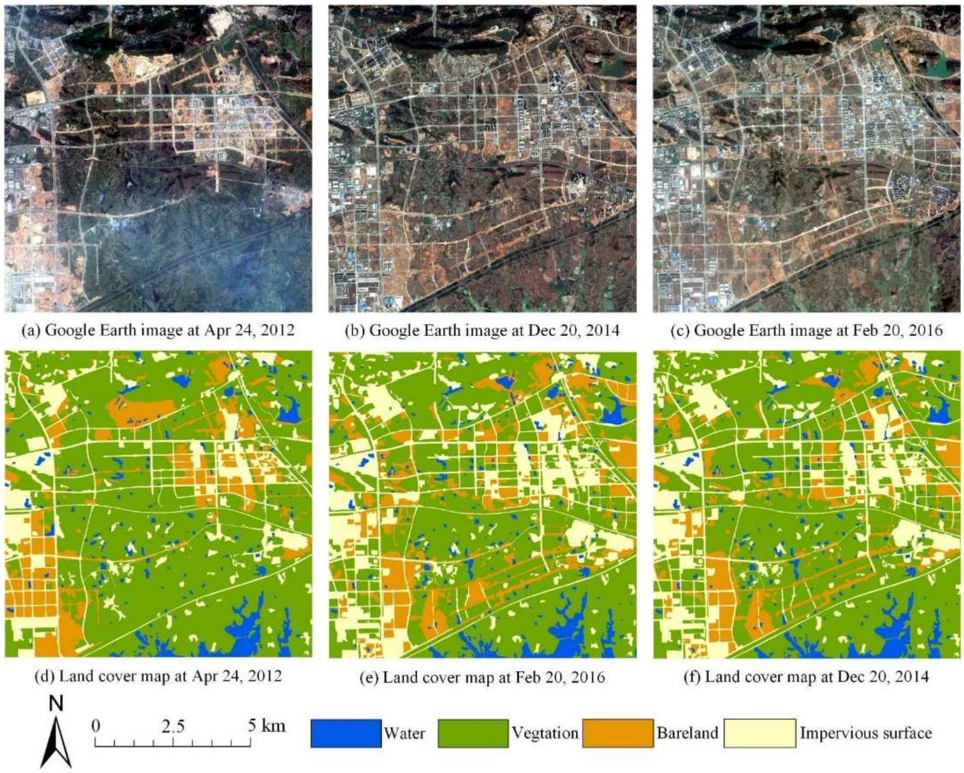


Figure 3. Time-series 5 m Google Earth images and corresponding land cover maps in the GEI experiment.

For ESTARFM and FSDAF, they were designed originally to predict FR reflectance images. As there are no satellite reflectance images in the GEI experiment, the simulated fraction maps were then used as the input of ESTARFM and FSDAF to directly predict the FR fraction maps at 2014. The Landsat-like and MODIS-like fraction maps at 2012 and 2016 were used as the input FR and CR data that pre- (e.g. T_i) and post- (e.g. T_j) the date of prediction in ESTARFM and STFMF. For FSDAF, as only one image pair pre- (2012) or post- (2016) the date of prediction is needed. The FSDAF based on the pair of fraction maps that include the data for 2012 is regarded as FSDAF^{2012}, while that based on the pair of fraction maps that include the data for 2016 is regarded as FSDAF^{2016}. The advantages of using simulated fraction maps is that it could represent greater control on the errors arising from factors such as the spectral unmixing analysis, geographical mis-registration and differences in satellite sensor properties. Moreover, the reference data (e.g. Landsat-like fraction maps at 2014) are known at the date of prediction and could thus be used objectively to assess the quality of results produced by different methods.

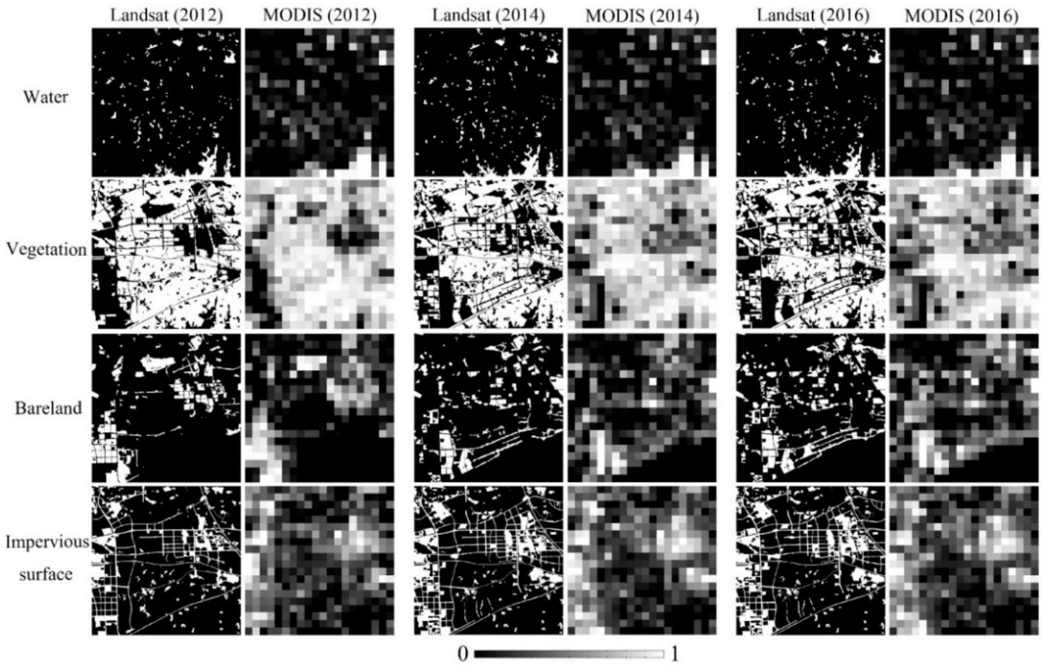


Figure 4. Time-series Landsat-like and MODIS-like fraction maps of four land covers in the GEI experiment.

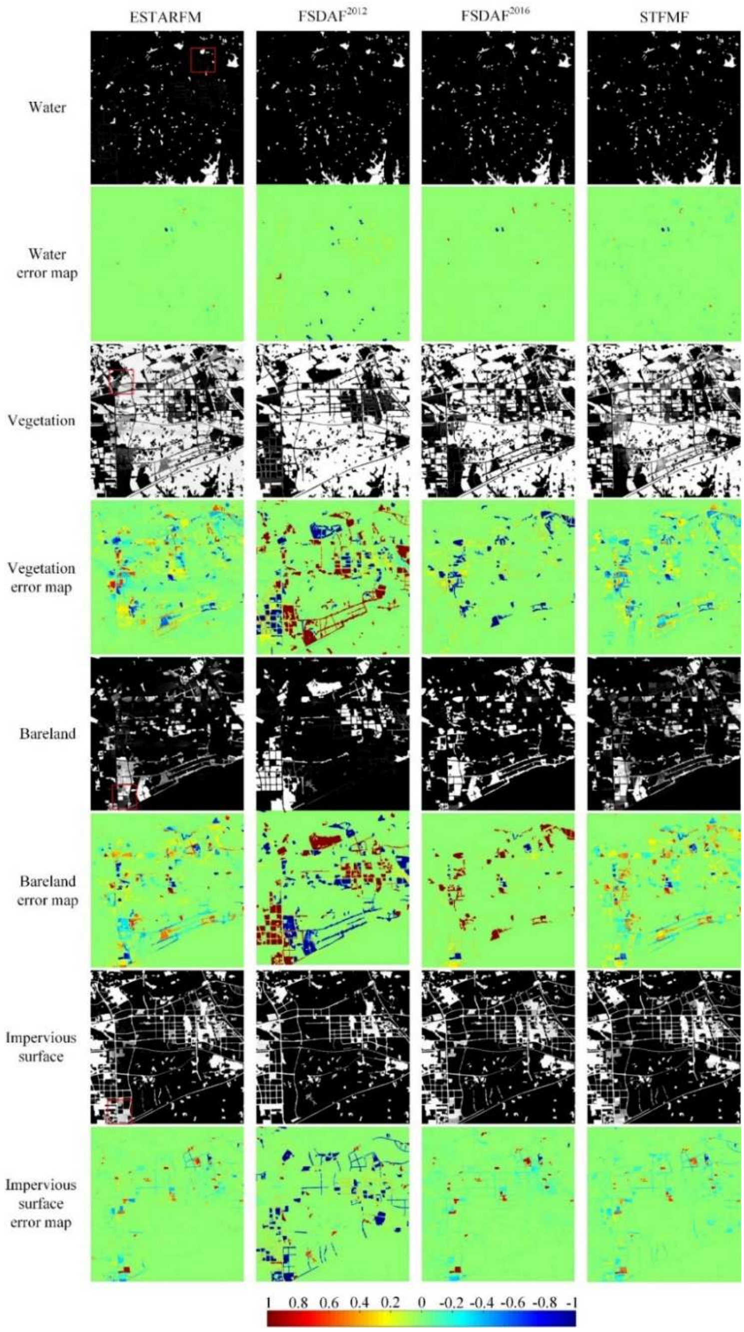


Figure 5. FR fraction maps and corresponding fraction difference images produced by ESTARFM, FSDAF^{2012}, FSDAF^{2016} and STFMF in the GEI experiment.

With the input multi-scale fraction maps shown in Fig. 4, the FR fraction maps and corresponding fraction error images produced by ESTARFM, FSDAF^{2012}, FSDAF^{2016} and STFMF are presented in Fig. 5. The fraction error images were generated by comparing the resultant FR fraction maps with the reference FR fraction maps at 2014. Additionally, four enlarged subarea images with spatial size of 50 × 50 pixels were shown in Fig. 6 to provide a clearer visual comparison of the results, and the red boxes in

Fig. 5 indicate the locations of the four enlarged subarea images.

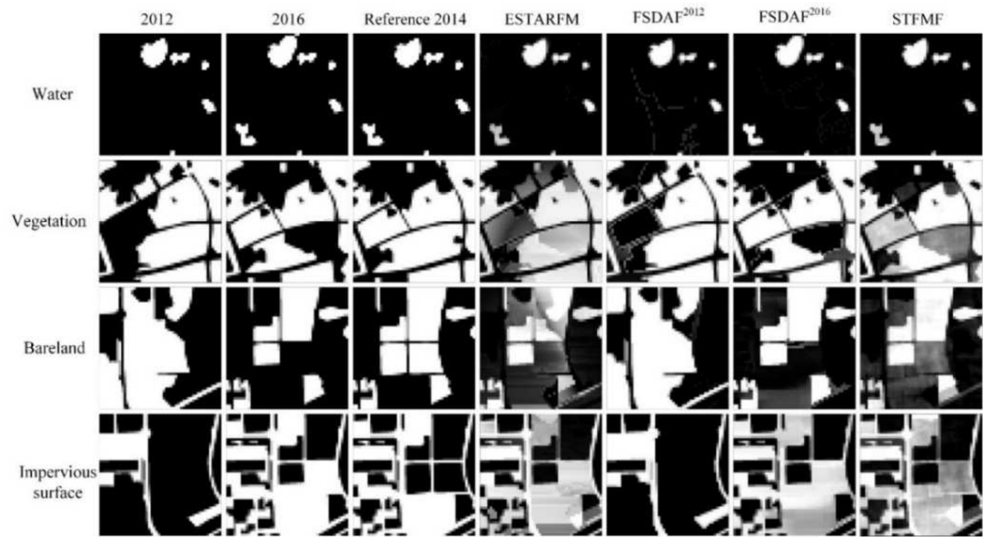


Figure 6. FR fraction maps at the subarea of the rectangle as shown in Fig. 5 (with spatial size of 50 × 50 pixels) for four land coves in the GEI experiment.

For the results of ESTARFM shown in Fig. 5, there were many pixels with mis-estimated fractional cover in the vegetation and bareland classes, and many pixels were over-estimated in the fraction maps of impervious surface. These errors arose because ESTARFM assumes that there were no land cover changes during the period spanned by the prediction process. Any areas that had undergone change would not be accurately estimated in the results. For FSDAF^{2012} and FSDAF^{2016}, there are more pixels with mis-estimated fractional cover. As presented in Fig. 6, the results of FSDAF^{2012} and FSDAF^{2016} are almost the same as the subarea fraction maps at 2012 and 2016 respectively. This is because FSDAF is mathematically based on the linear spectral mixture theory to detect temporal land cover change (Zhu et al. 2016). However, the input data of this GEI experiment are already the fraction maps that assumed to be perfectly generated by spectral unmixing, and the results of FSDAF would be similar to the pre- or post-time FR fraction maps. Focusing on the result of STFMF, it is evident that there are relatively few pixels with large mis-estimation errors indicated by dark blue and red colours in Fig. 5 and Fig. 6. Overall, it was evident that of the methods investigated the STFMF produced the FR fraction maps that were visually closest to the reference FR fraction maps.

Table 1. Accuracy assessment of the FR fraction maps generated by different methods in the GEI experiment. (The bold means the best value)

		Ideal	ESTARFM	FSDAF^{2012}	FSDAF^{2016}	STFMF
CC	Water	1	**0.9941**	0.9428	0.9810	0.9908
	Vegetation	1	0.9408	0.6702	0.9099	**0.9603**
	Bareland	1	0.8710	0.2859	0.8192	**0.9000**
	Impervious surface	1	0.9724	0.7896	0.9678	**0.9774**
	Mean	1	0.9446	0.6721	0.9195	**0.9571**
RMSE	Water	0	**0.0224**	0.0693	0.0406	0.0284
	Vegetation	0	0.1555	0.3686	0.1944	**0.1289**
	Bareland	0	0.1554	0.3864	0.2086	**0.1383**
	Impervious surface	0	0.0905	0.2483	0.0976	**0.0827**
	Mean	0	0.1060	0.2681	0.1353	**0.0946**
AAD	Water	0	**0.0019**	0.0090	0.0039	0.0026
	Vegetation	0	0.0637	0.1619	0.0552	**0.0524**
	Bareland	0	0.0567	0.1697	0.0539	**0.0526**
	Impervious surface	0	0.0218	0.0817	0.0217	**0.0231**
	Mean	0	0.0361	0.1056	0.0337	**0.0327**
UIQI	Water	1	**0.9941**	0.9423	0.9799	0.9899
	Vegetation	1	0.9388	0.6666	0.9090	**0.9577**
	Bareland	1	0.8633	0.2830	0.7852	**0.8904**
	Impervious surface	1	0.9720	0.7738	0.9674	**0.9760**
	Mean	1	0.9420	0.6664	0.9104	**0.9535**

Table 1 exhibits the accuracy assessment of the FR fraction maps produced by four spatial-temporal fusion methods. FSDAF^{2012} was associated with the worst accuracy values, particularly for the fraction maps of vegetation and bareland. The FSDAF^{2016} results were better than those from the FSDAF^{2012}, because FSDAF failed to estimate temporal land cover change, and land cover change between 2014 and 2012 was larger than that between 2016 and 2014. ESTARFM produced fraction maps with higher accuracy values than those of FSDAF^{2012} and FSDAF^{2016}, as it can take advantage of both pre- and post-prediction date CR and FR fraction maps. Consistent with the abovementioned visual comparison, the FR fraction maps produced by STFMF achieved almost the largest CC and UIQI values and smallest RMSE and AAD values and had an obvious improvement by comparing with the results of ESTARFM and FSDAF. This is because STFMF can not only take the best advantages of both the CR and FR fraction maps at 2012 and 2014, but also effectively deal with the temporal land cover change.

3.2 The MODIS-Landsat experiment

In order to have a comprehensive and rigorous validation of the performance of STFMF for different landscapes, both synthetic and real MODIS-Landsat images covering areas with heterogeneous (urban area) and homogeneous (rainforest area) landscapes were used. In addition, this experiment sought to show that a dense time series of FR fraction maps could be produced.

In the following MODIS-Landsat experiments, all of the Landsat Operational Land Imager (OLI, path 123 and row 039) and Enhanced Thematic Mapper Plus (ETM+, path 226 and row 069) images were collected as the land surface reflectance products from the USGS Earth Explorer (http://earthexplorer.usgs.gov). Additionally, the MODIS/Terra Surface Reflectance Daily L2G Global composite product of MOD09GA images (MODIS tile: h12v10) were obtained from the NASA's Earth Observing System Data and Information System (EOSDIS, http://reverb.echo.nasa.gov/reverb). MODIS images based on MOD09GA product have a spatial resolution of nearly 480 m, and the spatial ratio between MODIS and Landsat images is 16. As MODIS and Landsat images have different geographic reference systems, all of the MODIS images were reprojected into the geographic reference system of the original Landsat OLI and ETM+ images: UTM, WGS 84.

3.2.1 The urban area experiment

In this experiment, synthetic MODIS-Landsat images located for the urban area of Xianning city, China, were used to validate the performance of STFMF for a region with a heterogeneous land cover mosaic. Three cloud-free Landsat-8 Operational Land Imager (OLI) multispectral images acquired on

December 6, 2013, October 25, 2015 and October 30, 2017 were used as the FR Landsat images at times T_i, T_p and T_j, respectively. As shown in the first row of Fig. 7, each of the three time-series Landsat OLI images has a spatial size of 28.8 km × 28.8 km (960 × 960 pixels). For the CR images, synthetic MODIS images [see the second raw of Fig. 7], comprising 60 × 60 coarse pixels, were used; they were downscaled from the three Landsat-8 OLI images by a spatial averaging process. It is noteworthy that the synthetic MODIS images could represent greater control on the errors caused by satellite sensor difference and could thus be used objectively to assess and comprise the quality of FR fraction maps produced by different methods.

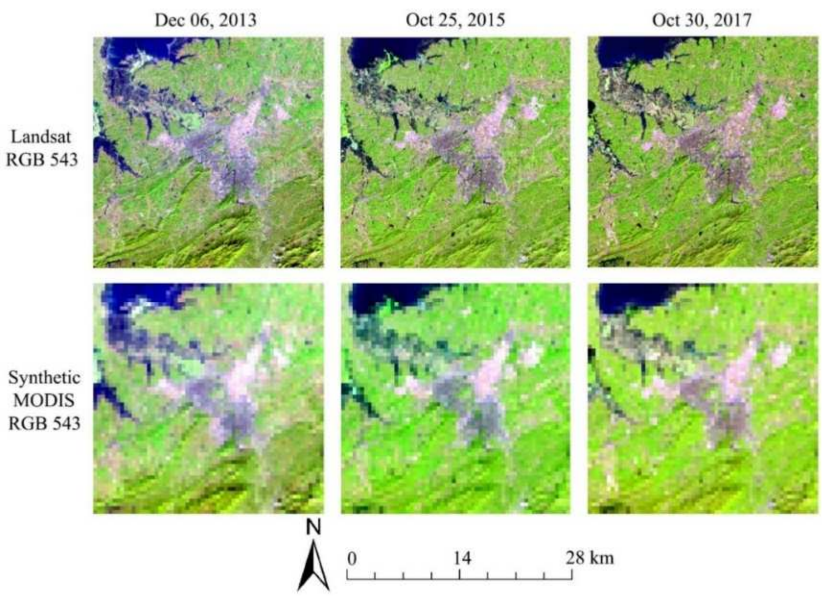


Figure 7. Landsat and downscaled MODIS images in the synthetic MODIS-Landsat experiment on urban area.

Time-series fine and coarse spatial resolution fraction maps of four land covers, water, vegetation, bareland and impervious surface, were then produced from the Landsat-8 OLI and synthetic MODIS images. With the generated MODIS and Landsat fraction maps at 2013 and 2017 and synthetic MODIS fraction maps at 2015, the FR fraction maps at 2015 were produced by the proposed STFMF approach. For ESTARFM and FSDAF, the inputs were the original Landsat-8 OLI reflectance images at 2013 and

2017, and corresponding synthetic and real MODIS reflectance images at 2015, the output was the predicted Landsat-8 OLI images at 2015, which were then used to generate the final FR fraction maps of four land cover classes at 2015. Fig. 8 shows the fraction maps produced by different methods for the synthetic MODIS images and also presents the fraction error maps by comparing with the reference FR fraction maps. Table 3 reports the accuracy assessment of the resultant FR fraction maps.

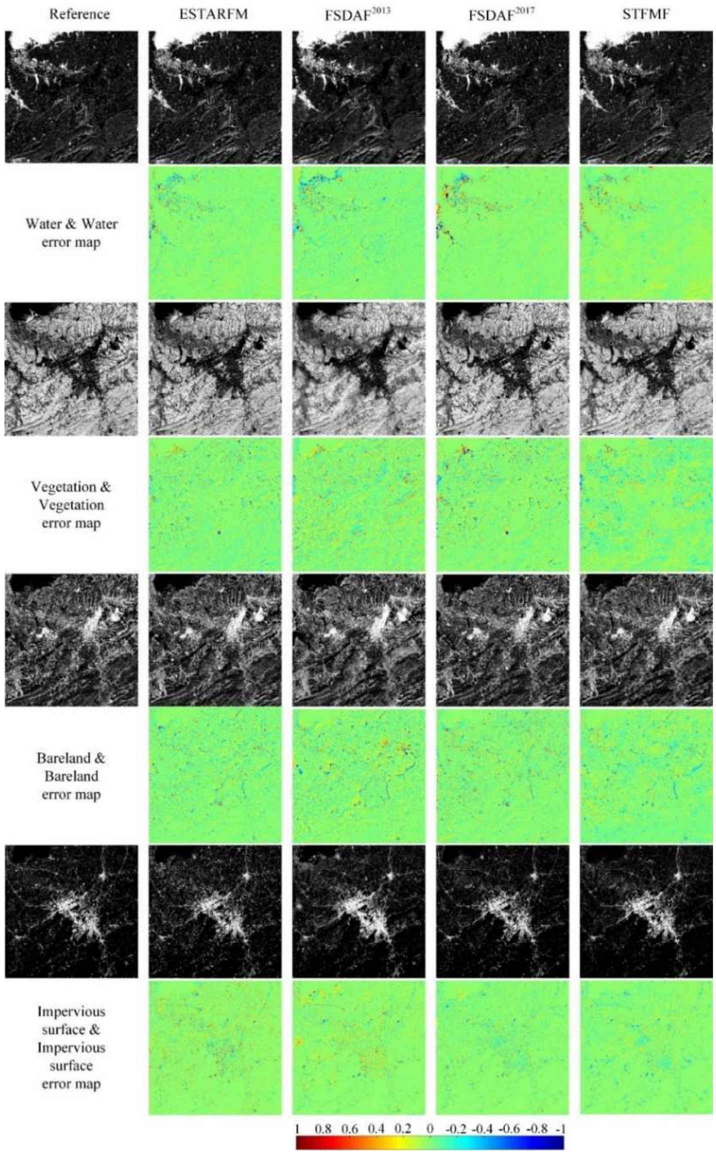


Figure 8. Reference FR fraction maps, resultant FR fraction maps and fraction error maps in the synthetic MODIS-Landsat experiment on urban area.

For ESTARFM, as shown in the second column of Fig. 8, the fraction maps of water and bareland have many pixels with under-estimated fractional value (blue pixels in the error map), while the vegetation and impervious surface fraction maps have many pixels with over-estimated fractional values (red pixels in the error map) around the boundaries. Compared with ESTARFM, more pixels with mis-estimated fractional cover can be found in the results of FSDAF^{2013}, especially for the fraction maps of vegetation and bareland. By contrast, for the vegetation and impervious surface fraction maps of FSDAF^{2017}, the result was superior to those from ESATRFM and FSDAF^{2013}. Although FSDAF has the ability to deal with land cover change to some extent, it is still sensitive to land cover change. Focusing on the results of the proposed STFMF approach, it was evident that there are fewer pixels with large mis-estimation of fractional cover in the error maps in comparison to those from the other methods. In addition, more spatial detail, such as of the linear water feature, was evident in the results, and the boundaries of different land cover features were represented most clearly. The FR fraction maps produced by STFMF are visually closest to the reference FR fraction maps.

Table 2 reports the accuracy assessment, although the water, vegetation and bareland fraction maps of ESTARFM were more accurate than those from FSDAF^{2013} and FSDAF^{2017}, it had the smallest CC and UIQI values and largest RMSE and AAD values for the fraction map of the impervious surface. Compared with FSDAF^{2013} and FSDAF^{2017}, it can be found that the fraction maps of FSDAF^{2017} have smaller CC and UIQI values and larger RMSE and AAD values than those of FSDAF^{2013}. This is because the land cover change of fraction maps between 2013 and 2015 is larger than that between 2015 and 2017. Consistent with visual comparison, by taking advantages of both the fine and coarse spatial resolution fraction maps at 2013 and 2017, the proposed STFMF approach produced the FR fraction maps with the largest CC and UIQI values and smallest RMSE and AAD values.

Table 2. Accuracy assessment of the FR fraction maps generated by different methods in the synthetic MODIS-Landsat experiment of an urban area. (The bold means the best value)

		Ideal	ESTARFM	FSDAF^{2013}	FSDAF^{2017}	STFMF
CC	Water	1	0.8974	0.8684	0.8780	**0.9107**
	Vegetation	1	0.8878	0.8374	0.8741	**0.8957**
	Bareland	1	0.8149	0.7624	0.7975	**0.8337**
	Impervious surface	1	0.6986	0.7274	0.7892	**0.8095**
	Mean	1	0.8247	0.7989	0.8347	**0.8624**
RMSE	Water	0	0.1086	0.1205	0.1200	**0.1023**
	Vegetation	0	0.1336	0.1565	0.1429	**0.1271**
	Bareland	0	0.1353	0.1516	0.1440	**0.1282**
	Impervious surface	0	0.1393	0.1244	0.1092	**0.1017**
	Mean	0	0.1292	0.1382	0.1290	**0.1148**
AAD	Water	0	0.0621	0.0722	0.0643	**0.0607**
	Vegetation	0	0.0888	0.1138	0.0926	**0.0873**
	Bareland	0	0.0846	0.0991	0.0883	**0.0821**
	Impervious surface	0	0.0675	0.0630	0.0498	**0.0491**
	Mean	0	0.0757	0.0870	0.0738	**0.0698**
UIQI	Water	1	0.8973	0.8596	0.8777	**0.9072**
	Vegetation	1	0.8874	0.8288	0.8741	**0.8926**
	Bareland	1	0.8105	0.7566	0.7970	**0.8286**
	Impervious surface	1	0.6861	0.7214	0.7850	**0.7990**
	Mean	1	0.8203	0.7916	0.8334	**0.8568**

3.2.2 The rainforest area experiment

Real MODIS-Landsat images of a region of rainforest were used to further validate the performance of the proposed STFMF approach for a relatively homogeneous landscape. A time-series cloud-free Landsat ETM+ images (path 226 and row 069) acquired on July 28, 2002 (T_i), August 13, 2002 (T_p) and August 29, 2002 (T_j) were used as the FR remotely sensed images. The corresponding real MOD09GA images (MODIS tile: h12v10) acquired at almost the same time as that of Landsat ETM+ images were used as the CR remotely sensed image. As shown in Fig. 9, each band of the Landsat ETM+ images includes 432 × 432 pixels, and each band of the MOD09GA images contains 27 × 27 pixels. Three land covers, forest, bareland and burned area, were studied and the fine and coarse spatial resolution fraction maps.

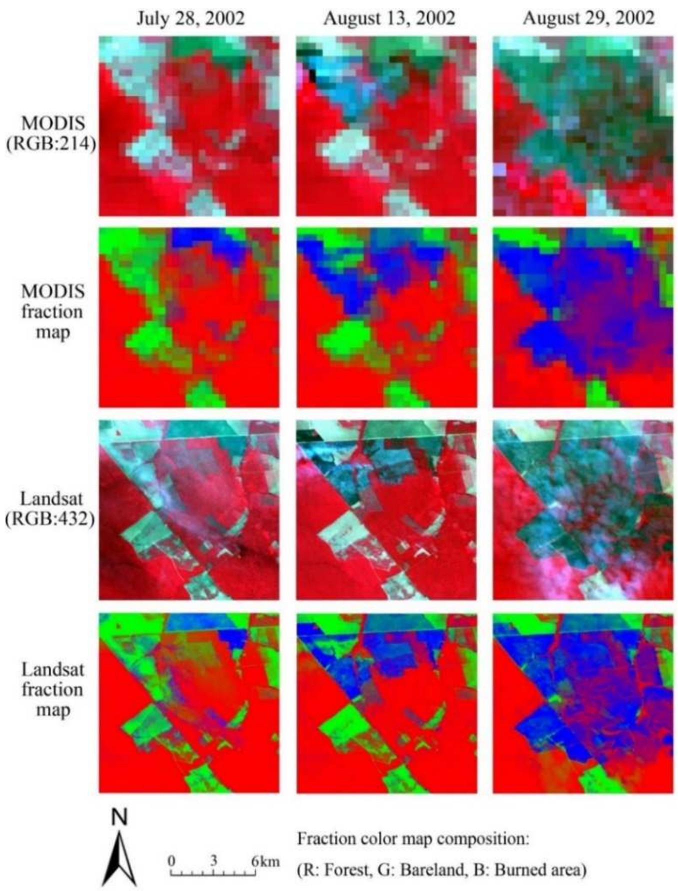


Figure 9. MODIS, Landsat reflectance images and fraction maps of forest, bareland and burned area in the MODIS-Landsat experiment on rainforest area.

The Landsat ETM+ image acquired on August 13, 2002 was used to produce the reference FR fraction maps. ESTARFM and FSDAF were applied for the original time-series Landsat and MODIS reflectance images to predict the FR Landsat-like multispectral images. Specially, FSDAF is based on the MODIS-Landsat images pair at T_i, as the fractional land cover change between T_p and T_j is larger than that between T_i and T_p. As shown in the second and third rows of Fig. 10, the fused FR reflectance images were used as the inputs of LSM to produce the Landsat-like fraction maps. With the time-series Landsat and MODIS fraction maps, the proposed STFMF approach was used to produce the Landsat-like fraction maps as shown in the last row of Fig. 10. Moreover, the corresponding fraction error maps for different methods were generated by comparing with the reference FR fraction maps. The accuracy assessment of the results generated by different fusion methods is listed in Table 3.

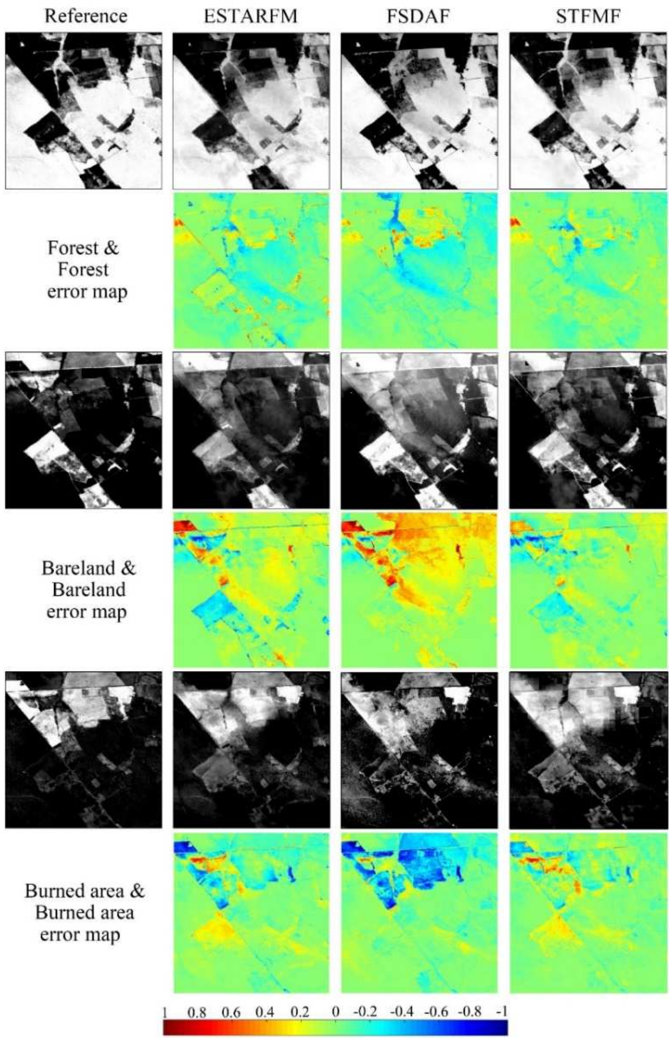


Figure 10. Reference FR fraction maps, resultant FR fraction maps and corresponding fraction error maps in the MODIS-Landsat experiment on rainforest area.

Similar trends as those observed in the MODIS-Landsat experiment on urban area can also be found in this MODIS-Landsat experiment for the rainforest area. As shown in Fig. 10, due to the inability of ESTARFM to deal with land cover changes, many pixels with over-estimated (red pixels in the error maps) and under-estimated (blue pixels in the error maps) fractions appear in the results. Compared with the ESTARFM results, there were fewer over-estimated forest fraction features in the output of FSDAF, but more over-estimated bareland fraction features and under-estimated burned area fraction features appear across the results. Overall, the results of FSDAF have the lowest accuracy values, as shown in Table 3. This demonstrates that FSDAF is not able to deal with land cover change well in a real situation.

Notably, the results of the proposed STFMF approach have fewer pixels with large fraction mis-estimation in the error maps, and the under-estimated and over-estimated fraction features decrease significantly. STFMF produced FR fraction maps that were visually closer to the reference FR fraction maps shown in Fig. 10. For the accuracy assessment reported in Table 3, consistently with the above images experiments, STFMF produced the FR fraction maps with the largest CC and UIQI values and smallest RMSE and AAD values, which highlights its potential for the production of FR fraction maps for a relatively homogeneous landscape even if land cover change may have occurred within a short time.

Table 3. Accuracy assessment of the fraction maps generated by different spatial-temporal fusion methods applied to the MODIS-Landsat experiment on rainforest area. (The bold means the best value)

		Ideal	ESTARFM	FSDAF	STFMF
CC	Forest	1	0.9564	0.9522	**0.9721**
	Bareland	1	0.8360	0.8554	**0.9143**
	Burned area	1	0.8484	0.7634	**0.9042**
	Mean	1	0.8802	0.8570	**0.9302**
RMSE	Forest	0	0.1239	0.1337	**0.0971**
	Bareland	0	0.1697	0.2193	**0.1218**
	Burned area	0	0.1460	0.2048	**0.1177**
	Mean	0	0.1465	0.1859	**0.1122**
AAD	Forest	0	0.0800	0.0863	**0.0597**
	Bareland	0	0.1091	0.1462	**0.0725**
	Burned area	0	0.0845	0.1152	**0.0686**
	Mean	0	0.0912	0.1159	**0.0669**
UIQI	Forest	1	0.9519	0.9491	**0.9716**
	Bareland	1	0.8145	0.7469	**0.9099**
	Burned area	1	0.8260	0.5268	**0.9025**
	Mean	1	0.8642	0.7409	**0.9280**

Finally, STFMF was used to generate a time series of FR fraction maps for the experiment focused on the rainforest. During the period from July 28, 2002 (T_i) to August 29, 2002 (T_j), as shown in the first row of Fig. 11, we collected four other scenes of MOD09GA images (cloud-free images covering the study site); however, there is only one scene of Landsat ETM+ image (acquired on August 13, 2002) covering the study site during T_i and T_j. To provide a greater understanding of the forest fraction changes that occurred between T_i and T_j, it is of interest to obtain time-series fine spatial and temporal forest fraction maps between T_i and T_j from the CR MODIS images applying STFMF.

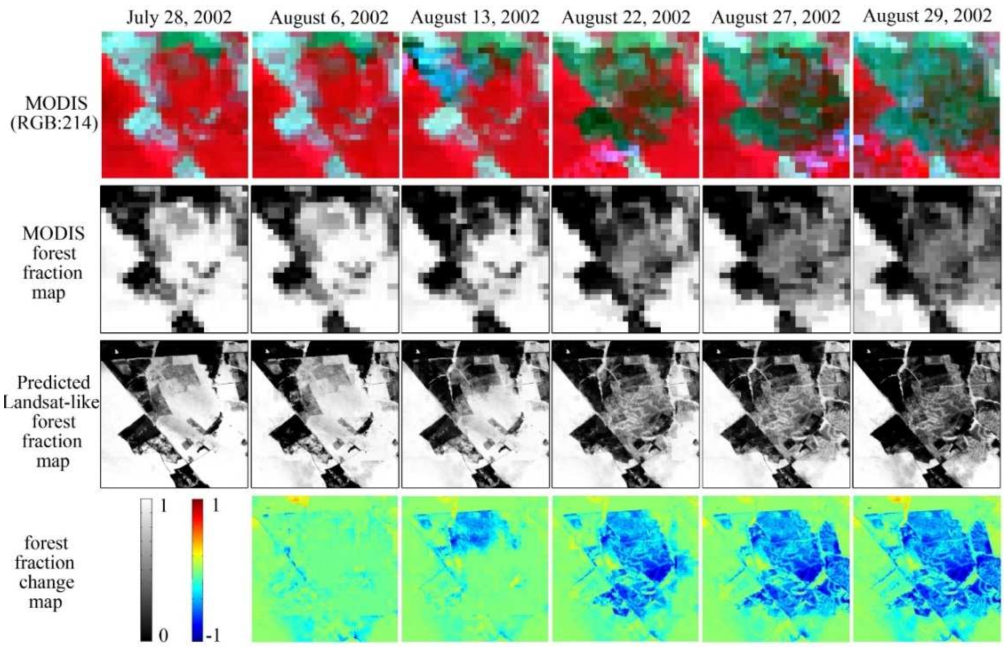


Figure 11. Time-series MODIS reflectance images, MODIS forest fraction maps, predicted Landsat-like forest fraction maps and forest fraction change maps between July 28, 2002 and August 29, 2002.

With the collected subarea MOD09GA images acquired on August 6, 13, 22 and 27 of 2002 and the endmembers of three land cover classes of forest, bareland and burned area, the time-series MODIS forest fraction maps were then generated by using LSM. Since the MODIS-Landsat forest fraction map pairs at T_i and T_j are already known, four time-series FR forest fraction maps shown in Fig. 11 can, thus, be reconstructed from the MODIS forest fraction maps at August 6, 13, 22 and 27 of 2002 (T_p) by using STFMF. Moreover, the last row of Fig. 11 shows the FR (Landsat-like) forest fraction change maps at August 6, 13, 22, 27 and 29 of 2002 by comparing with the Landsat image-based forest fraction map acquired on July 28, 2002.

From July 28, 2002 to August 29, 2002 (which is almost one month), there were substantial land cover changes that occurred. By observing the time-series MODIS forest fraction maps shown in the second row of Fig. 11, it is possible to observe the trend of forest fraction change that happened within the one-month period; however, due to the coarse spatial resolution of MODIS images, the detail about the spatial patterns of forest fraction change was almost lost. By contrast, the predicted time-series Landsat-like forest fraction maps contain greater spatial detail, especially some small-sized linear forest cover features. Simultaneously, the forest fraction change maps generated by using the predicted Landsat-like forest fraction maps exploit more spatial detail information about the forest cover change, in which the change of forest cover started at the north central part, and then spread from the northwest to the southeast. This experiment demonstrates the potential of STFMF for generating a dense time-series of fine spatial and temporal forest fraction maps, which will provide more accurate information about where, when and how forest fraction changes occur through time. Critically, it allows exploitation of the high temporal resolution of CR MODIS imagery to provide FR land cover information.

4. Discussion

In above synthetic and real experiments, STFMF achieved the most accurate FR fraction maps in both terms of visual and quantitative comparisons. In addition, STFMF showed great potential to produce a time series of FR land cover fraction maps from the high temporal resolution of CR images.

4.1 Influence of satellite sensor difference

In order to assess the influence of satellite sensor difference on the performance of the proposed STFMF model, the synthetic MODIS images were replaced by the real MODIS/Terra Surface Reflectance 8-Day L3 Global composite product of MOD09A1 images (Terra MODIS tile: h27v06), as shown in Fig. 12, in the MODIS-Landsat urban area experiment. Table 4 reports the accuracies of the FR fraction maps generated by different methods with real MODIS-Landsat images.

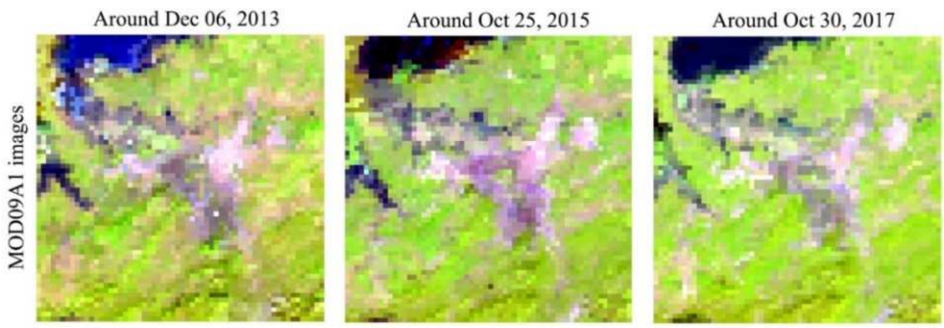


Figure 12. Real MODIS reflectance images (MOD09A1) in the MODIS-Landsat experiment on urban area.

The accuracy values [see Table 4] of the predicted FR fraction maps produced using real MODIS images were worse than those obtained through the use of synthetic MODIS images [see Table 2]. In particular, $FSDAF^{2013}$ and $FSDAF^{2017}$ showed a greater decline in accuracy relative to ESTARFM and STFMF. The mean CC values of $FSDAF^{2013}$ and $FSDAF^{2017}$ results decreased by 0.0785 and 0.0393, while those of ESTARFM and STFMF were 0.0162 and 0.0088 respectively. This indicates that the satellite sensor difference would have a negative impact on the results of all spatial-temporal fusion methods, and especially for FSDAF. This is because there are no registration error and bandwidth difference between MODIS and Landsat images when the synthetic MODIS images were used while with the real MODIS images, errors associated with mis-registration and the bandwidth differences would be inherited into the results. However, when all spatial-temporal fusion methods are compared, a similar trend as that in the synthetic MODIS-Landsat experiment can also be observed. The fraction maps produced by STFMF had the better accuracy values in comparison to those from ESTARFM and FSDAF. Moreover, the decrease of CC, UIQI values and the increase of RMSE, AAD values for STFMF results were smaller than that of ESTARFM and FSDAF. This demonstrates that STFMF is more accurate, and less sensitive to the errors caused by differences in the satellite sensor data used.

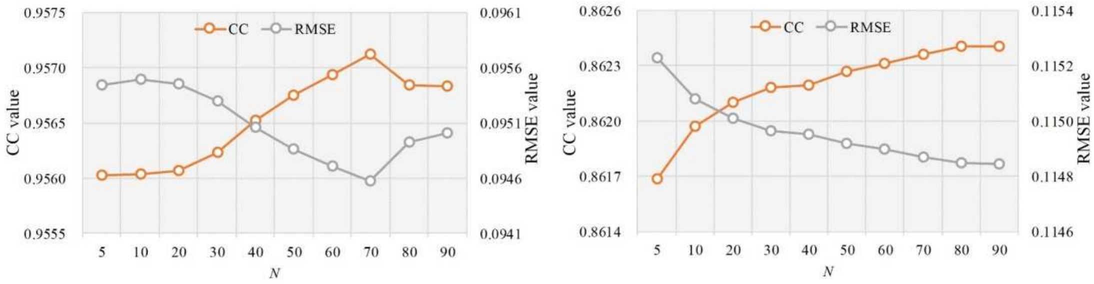
Table 4. Accuracy assessment of the FR fraction maps generated by different methods in the real MODIS-Landsat experiment on urban area.

		ESTARFM	FSDAF^{2013}	FSDAF^{2017}	STFMF
CC	Water	0.8677(-0.0297)	0.8034(-0.0650)	0.8479(-0.0301)	0.9021(-0.0086)
	Vegetation	0.8670(-0.0208)	0.8028(-0.0346)	0.8486(-0.0255)	0.8890(-0.0066)
	Bareland	0.8005(-0.0143)	0.6946(-0.0678)	0.7431(-0.0545)	0.8243(-0.0094)
	Impervious surface	0.6985(-0.0001)	0.5808(-0.1465)	0.7420(-0.0471)	0.7989(-0.0106)
	Mean	0.8084(-0.0162)	0.7204(-0.0785)	0.7954(-0.0393)	0.8536(-0.0088)
RMSE	Water	0.1233(0.0148)	0.1468(0.0263)	0.1386(0.0186)	0.1074(0.0051)
	Vegetation	0.1520(0.0183)	0.1739(0.0174)	0.1615(0.0185)	0.1310(0.0039)
	Bareland	0.1536(0.0183)	0.1724(0.0208)	0.1678(0.0238)	0.1316(0.0033)
	Impervious surface	0.1444(0.0051)	0.1747(0.0504)	0.1272(0.0180)	0.1046(0.0028)
	Mean	0.1433(0.0141)	0.1669(0.0287)	0.1488(0.0197)	0.1186(0.0038)
AAD	Water	0.0739(0.0118)	0.0896(0.0174)	0.0805(0.0162)	0.0631(0.0024)
	Vegetation	0.1057(0.0169)	0.1296(0.0158)	0.1109(0.0182)	0.0900(0.0027)
	Bareland	0.1015(0.0169)	0.1175(0.0184)	0.1122(0.0239)	0.0851(0.0030)
	Impervious surface	0.0721(0.0046)	0.1021(0.0391)	0.0610(0.0112)	0.0510(0.0018)
	Mean	0.0883(0.0126)	0.1097(0.0227)	0.0911(0.0174)	0.0723(0.0025)
UIQI	Water	0.8673(-0.0300)	0.7919(-0.0678)	0.8394(-0.0383)	0.8978(-0.0093)
	Vegetation	0.8584(-0.0290)	0.7887(-0.0401)	0.8472(-0.0268)	0.8858(-0.0067)
	Bareland	0.7589(-0.0516)	0.6910(-0.0656)	0.7143(-0.0827)	0.8189(-0.0097)
	Impervious surface	0.6726(-0.0135)	0.5061(-0.2153)	0.7395(-0.0455)	0.7905(-0.0085)
	Mean	0.7893(-0.0310)	0.6944(-0.0972)	0.7851(-0.0483)	0.8483(-0.0086)

Note: The values in brackets indicate the difference between the real and synthetic MODIS-Landsat experiments on urban area, negative value means decreasing and positive value mean increasing.

4.2 Influence of the number of candidate neighboring patch pairs

The number of candidate neighboring patch pairs (*N*) is a critical parameter in the KRR model used in STFMF. In order to evaluate how *N* influences the results of STFMF, the value of *N* was set at values varying from 5 to 90, and Fig. 13 reports the corresponding mean CC and RMSE values of four land cover fraction maps in the GEI experiment and the MODIS-Landsat urban area experiment. Generally, when *N* was very small, such as 5, the FR fraction maps of STFMF in both experiments had the lowest CC and highest RMSE values. This is because the use of few neighboring patch pairs results in a failure to provide enough FR spatial feature information for the prediction process. The CC values increased rapidly when *N* increased from 5 to 70 in the GEI experiment and 5 to 80 in the MODIS-Landsat experiment. With the continuous increase of *N* (e.g. larger than 70), the results of STFMF in the GEI experiment achieved decreasing CC values and increasing RMSE values. But for the MODIS-Landsat experiment, there was almost no obvious increase when *N* was larger than 80. Compared with the MODIS-Landsat experiment, the changes of CC and RMSE values for the results in GEI experiment are more sensitive to the variation of *N*, but similar trend of CC and RMSE values can be observed from them. Fig. 13 indicates that a larger value of *N* is suggested, but the STFMF results would have no obvious improvement when the value of *N* is set at a very large value (such as larger than 80). Moreover, it is noteworthy that the computation cost would increase rapidly with the increment of *N*. Therefore, in practice, if there is a specific limitation of the computation cost, it is suggested to set *N* as a relative small value, such as between 60 to 80.



(a) GEI experiment (b) MODIS-Landsat experiment on urban area

Figure 13. Influence of the candidate neighboring patch pairs number (*N*) on the STFMF results in the GEI experiment and MODIS-Landsat experiment on urban area.

4.3 Influence of fraction errors

In order to have a quantitative analysis of the influence of fraction errors on the resultant FR fraction maps of STFMF, the Gaussian noise was added into the synthetic time-series Landsat-like fraction maps in the GEI experiment to simulate errors caused by different spectral unmixing methods. Table 5 lists the accuracy assessment of STFMF results with different fraction error levels ranging from 0 to 0.2 with an interval of 0.02. The corresponding input MODIS-like fraction maps in STFMF were downscaled from the Gaussian noise-based Landsat-like fraction maps by spatially averaging. It is evident from table 5 that with the increment of fraction errors, the CC values of STFMF results had a continuous decrease, while the RMSE values had a continuous increase. Moreover, the decrease of CC values and the increase of RMSE values became larger with the increasing of fraction error. This illustrates that errors in fraction maps would have a serious impact on the STFMF results. In practice, the fraction errors caused by spectral unmixing vary from method to method, and more powerful spectral unmixing methods should be applied to provide more accurate fraction maps, in order to finally improve the STFMF results.

Table 5. Accuracy assessment of the STFMF results with different fraction error levels in the GEI experiment.

Fraction error		0	0.02	0.04	0.06	0.080	0.10	0.12	0.14	0.16	0.18	0.20
CC	Water	0.9908	0.9888	0.9844	0.9768	0.9654	0.9497	0.9293	0.9050	0.8770	0.8401	0.7992

	Vegetation	0.9603	0.9596	0.9572	0.9537	0.9476	0.9407	0.9299	0.9185	0.9033	0.8849	0.8614
	Bareland	0.9000	0.8994	0.8957	0.8913	0.8826	0.8727	0.8586	0.8436	0.8212	0.7970	0.7669
	IS	0.9774	0.9769	0.9749	0.9706	0.9656	0.9581	0.9472	0.9337	0.9173	0.8972	0.8733
	Mean	0.9571	0.9562	0.9531	0.9481	0.9403	0.9303	0.9163	0.9002	0.8797	0.8548	0.8252
	Water	0.0284	0.0328	0.0409	0.0512	0.0632	0.0758	0.0888	0.1018	0.1145	0.1287	0.1425
	Vegetation	0.1289	0.1327	0.1406	0.1509	0.1651	0.1797	0.1977	0.2152	0.2344	0.2542	0.2752
RMSE	Bareland	0.1383	0.1391	0.1424	0.1462	0.1527	0.1594	0.1682	0.1765	0.1872	0.1979	0.2096
	IS	0.0827	0.0853	0.0915	0.1010	0.1113	0.1238	0.1380	0.1532	0.1684	0.1839	0.2000
	Mean	0.0946	0.0975	0.1038	0.1123	0.1231	0.1347	0.1482	0.1617	0.1761	0.1912	0.2068

Note: IS indicates impervious surface.

4.4 Comparisons of three satellite images spatial-temporal fusion models

Benefiting from the free availability, wide swath, short revisit-rate and long-term archiving of CR satellite images and amount of spatial details in FR satellite images, spatial-temporal fusion methods can reconstruct time-series fine spatial and temporal resolution images for large areas and over long-time frames. As shown in Fig. 14, current satellite images spatial-temporal fusion models could be summarized into three different levels: surface reflectance level, land cover class level and land cover fraction level. Surface reflectance level includes the popular spatial-temporal fusion methods of STARFM, ESTARFM and FSDAF, and the output of them is the FR surface reflectance multispectral images. Although the predicted FR multispectral images can be used to produce FR land cover map as that of STMRF and FR fraction maps as that of STFMF, they were designed particularly for the prediction of reflectance multispectral images, and most of them are sensitive to the land cover change.

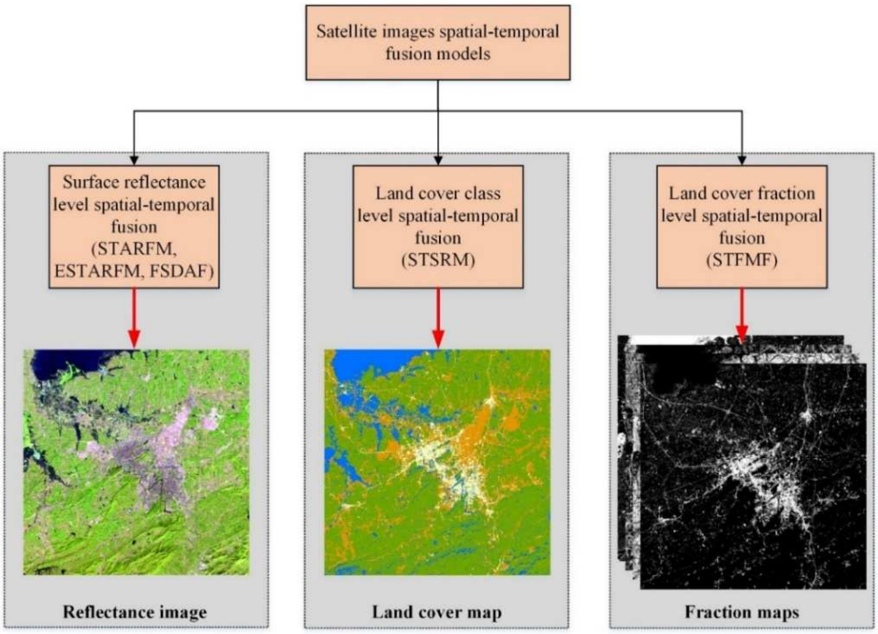


Figure 14. Illustration of the three different levels of satellite images spatial-temporal fusion models.

Since STSRM takes into account land cover change information, land cover change would have less impact on the final FR land cover map than is observed with the other methods. The main disadvantage of STSRM is the pure pixel assumption of the input and output FR land cover maps. A major limitation of using the ‘pure’ pixel assumption in STSRM is that land cover change information occurring at the sub-pixel scale cannot be considered fully. An example shown in Fig. 15 is used to further illustrate the limitation. Assume that the fraction values of land cover class A for one fine pixel are 95% and 65 % at time T_1 and T_2 respectively, and the class labels of the fine pixel are the same land cover class A at both time T_1 and T_2. If we focus on the class label, there would be no land cover change for the fine pixel; but in fact, there is 30% loss of fraction values (land cover class A) at the sub-pixel scale between time T_1 and T_2. By contrast, for the proposed STMFM approach, the 30% loss of fraction values can be observed.

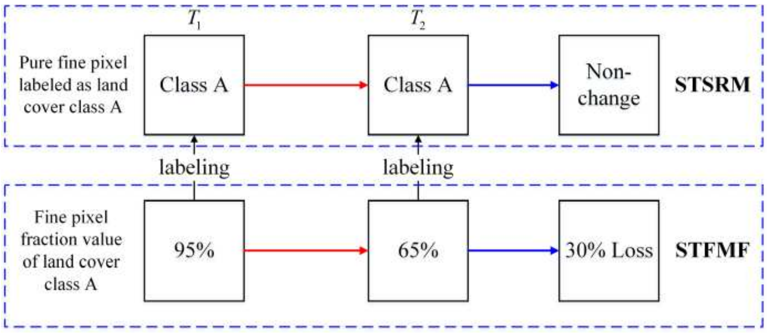


Figure 15. An example used to illustrate the FR land cover change of pure labeled pixel and fine pixel fraction values in STSRM and STFMF models.

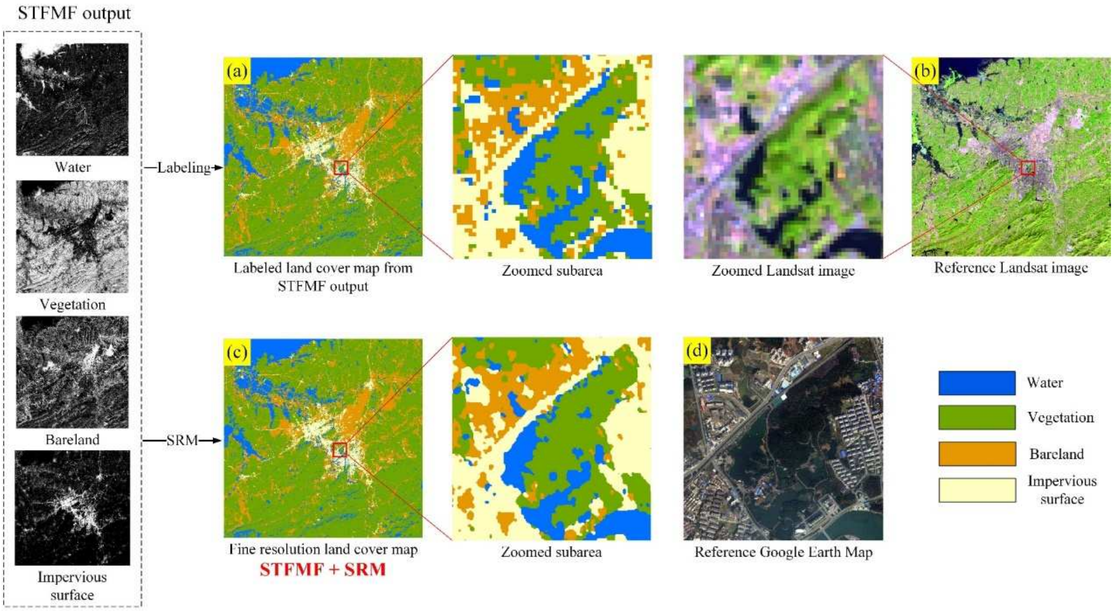


Figure 16. An example used to show the combination of STFMF and super resolution mapping (SRM). (a) Land cover map generated by labeling the resultant fraction maps of STFMF at per-pixel scale; (b) 30 m Landsat OLI image; (c) FR land cover map generated by the combination of STFMF and SRM at sub-pixel scale; (d) Reference Google Earth Map covering the zoomed subarea of Landsat images.

Generally, with the resultant FR fraction maps of STFMF, it is instinctive and readily to obtain a land cover map. An example shown in Fig. 16 is used to indicate the land cover mapping process. With the fraction maps of open water, vegetation, bareland and impervious surface generated by STFMF in the MODIS-Landsat images experiment, we can obtain a land cover map, as shown in Fig. 16(a), by using the tradition classification labeling strategy, where the class of a pixel is labeled as the land cover class which has the largest fraction values. The resultant FR land cover map shown in Fig. 16(a) shows the potential to present more spatial details about the four land covers than the original CR (MODIS) fraction maps and could have advantages to monitor the land cover changes occurred at Landsat image pixel scale.

However, a more effective way is to combine STFMF and super resolution mapping (SRM), to make the most use of resultant FR fraction maps. As a post-processing of spectral unmixing, SRM is a technique to predict the sub-pixel spatial locations of different land cover classes by using fraction maps as input, and can produce land cover maps at a finer spatial resolution than the input data (Atkinson 2005). Motivated by this, it is possible to use the fraction maps generated by STFMF as the input of SRM to further produce a land cover map with finer spatial resolution than that of the output fraction maps of STFMF. As shown in Fig, 16(c), the finer spatial resolution land cover map was produced by a spatial regularization-based SRM model (Ling et al. 2014; Zhong et al. 2015) with a spatial ratio of 6. Comparing Fig. 16(c) with Fig. 16 (a), it is observed that the 5 m land cover map generated by the combination of STFMF and SRM has more spatial smooth boundaries and presents more spatial details about different land covers. In addition, the land cover map shown in Fig. 16(c) is closer to the reference Google Earth Map shown in Fig. 16(d). This demonstrates the great potential of the combination of STFMF and SRM in the field of land cover mapping, and they could be integrated to provide finer spatial resolution land cover map.

4.5 Computation efficiency

In order to validate the computation efficiency of the proposed STFMF against ESTARFM and FSDAF, table 6 reports the computation cost of the ESTARFM, FSDAF and STFMF methods in real MODIS-Landsat experiments on urban and rainforest areas. The implementations of ESTARFM and FSDAF were performed by the IDL code (Zhu et al. 2010; Zhu et al. 2016), while STFMF was implemented by the MATLAB platform (MATLAB R2017b version). All of the algorithms used in this research were implemented on an Intel(R) Core(TM) i7-7700K Processer at 4.20 GHz. From table 6, it can be found that ESTARFM is the most time consuming, while FSDAF takes the least time in both of the two experiments. The computation cost of STFMF was more than that of FSDAF and lower than that of ESTARFM. For STFMF, most of the computation time is spent on KRR, in which the candidate neighboring patch pairs searching, the training and the predicting processes are time-consuming. A possible improvement is to take into account various spatial patterns of fraction changes during the training, in order to avoid repeatedly building training model for each prediction process. By this way, the computation cost of STFMF is expected to be obviously decreased, as the training process is the most time-consuming step.

Table 6. Computation cost of the ESTARFM, FSDAF and STFMF methods in real MODIS-Landsat experiment on urban area and rainforest area.

	Spatial size	ESTARFM	FSDAF	STFMF
Urban area	960 × 960 pixels	1359s	295s	364s
Rainforest area	432 × 432 pixels	304s	85s	242s

4.6 Limitations and future work

The input data are crucial to the performance of the proposed STFMF method. At first, STFMF aims to use the fraction change information with different spatial resolutions between fraction maps at T_i and T_j as the training dataset to predict the fraction change maps at T_p. The implicit assumption is that for any CR fraction change pattern, a similar CR and FR fraction change pattern can be found from the training dataset and they can be used to predict the final FR fraction change map. However, if the fraction maps at T_i and T_j are similar to each other, there will be not enough representative fraction change pattern information contained in the training dataset, and the predicting accuracy in

STFMF would be, therefore, decreased. For example, when there are land cover changes like floods on T_p, STFMF will be difficult to capture the changes on T_p, because data on T_i and T_j contain no information about floods. This phenomenon can be found in the GEI experiment, in which for the produced FR fraction map of water, ESTARFM had better accuracy than that of the proposed STFMF. This is because fraction maps of water at 2012 and 2016 were similar to each other and contained little information about the change of water. Therefore, it is suggested that the fraction maps collected at T_i and T_j should not be similar to each other, in order to contain more important information about the fraction change of various classes. Moreover, with a large study area, there will be higher possibilities to contain more fraction change spatial patterns for different classes. Secondly, fraction errors caused by spectral unmixing would limited the performance of STFMF. This issue arises because STFMF uses directly the fraction maps generated by a spectral unmixing analysis as input, and the accuracy of the fraction maps, therefore, affects the accuracy of the final result. In the experiments, the linear spectral mixture model was used to produce the fraction maps. Although linear spectral mixture modeling has physical significance, the actual spectral mixtures of the land surfaces are often non-linear (Keshava and Mustard 2002). To estimate the fraction maps more accurately from remotely sensed images, alternative non-linear spectral mixture models, such as artificial neural networks (Foody et al. 1997) and support vector machines (Brown et al. 2000) could be used.

The method used to predict the FR fraction change maps from CR fraction change maps is another key problem for STFMF. It is noteworthy that predicting FR image from the CR image is a pathological inversion problem, and there are possibilities that similar CR fraction change maps would produce different FR fraction change maps, especially when the spatial ratio between CR and FR images is high. A popular solution for this problem is to use the learning based methods by assuming that similar CR fraction change maps would be corresponding to similar FR fraction change maps. This has been successfully applied in the field of image supper-resolution (Freeman et al. 2002) and land cover supper-resolution mapping (Ling et al. 2016b). In this research, KRR is used as the learning algorithm, as it has used widely in the field of image super-resolution and has less number of parameters to be determined (Kim and Kwon 2010). But there are limitations when using KRR and further improvement exists. The normalization operation in equation (21) should be implemented for the output FR fraction maps predicted by KRR to ensure that the sum of the fraction values for all classes is 1. However, this will change the original values of the resultant FR fraction maps, and biases are likely to happen for the normalized fraction values. Generally, a better way of keeping the sum of the fraction values for all classes at 1 is to add constraints when deriving the FR fraction maps but not after all the fraction maps have been calculated. But it is hard for KRR to globally constrain the resultant fraction values of all classes at the same time. Besides KRR, there are some more powerful machine learning algorithms, such as deep learning convolutional neural networks (Dong et al. 2016; Zhang et al. 2016), which are expected to have a better performance than KRR. The future introduction of a framework based on deep learning algorithms into the proposed approach is of great interest, and will help improve the performance of STFMF.

There exists uncertainty for the weights of each prediction calculated globally in equations (18)-(20). Generally, a better way is to calculate these weights at local scale, as the temporal similarity between CR fraction maps will change site by site. However, given that the fraction error caused by the spectral unmixing is always inevitable in real applications, if the local weights are applied, the fraction error at the local scale would most likely be introduced into the final result. This is the reason why only global weights were applied in this research, but it is of great interest to develop more suitable approaches to incorporate local weights in the model.

5. Conclusion

In this paper, a novel approach, termed as STFMF, was proposed to generate fine spatial and temporal resolution fraction maps by fusing multiscale coarse-spatial-fine-temporal and fine-spatial-coarse-temporal remotely sensed images. Compared with the STSRM method, the proposed approach considers the mixed pixel problem at the fine spatial scale and can produce FR fraction maps instead of a FR land cover map. Compared with the traditional reflectance image spatial-temporal fusion methods, the proposed approach does not use directly the original remotely sensed images as inputs, but focuses on the multi-scale fraction maps generated by spectral unmixing and, thus, is theoretically more able to deal with any land cover change occurring at the sub-pixel scale. STFMF is good for spatial-temporal fusion, because it (1) can accommodate for the mixed pixel problem in FR remotely sensed images, (2) can use fraction maps generated from a range of satellite images or other suitable data sources, (3) focuses on the accurate estimation of fraction cover changes happened through time.

The performance of STFMF was assessed with several experiments including both synthetic and real images, and was also compared with two popular image spatial-temporal fusion methods: ESTARFM and FSDAF. The results show that the proposed approach is able to produce FR fraction maps with the greatest visual performance compared with the two benchmark methods, and contains more spatial detail about the land cover features in the regions of study. In both the synthetic and real image experiments, the proposed approach typically produced the largest CC and UIQI and smallest RMSE and AAD values. Moreover, the proposed approach was used to generate a time-series of FR forest fraction maps, which demonstrates the potential of STFMF in the production of a time-series of fine spatial and temporal forest fraction maps for real-world application. In addition, it is of great interest to combine STFMF and SRM to produce finer spatial resolution land cover maps than the resultant fraction maps produced by the proposed STFMF approach in future research.

Acknowledgment

The authors are grateful to the editors and three anonymous referees for their constructive comments and suggestions, which helped to improve this paper. This work was supported by the Strategic Priority Research Program of Chinese Academy of Sciences (Grant No. XDA2003030201), the Youth Innovation Promotion Association CAS (Grant No. 2017384), the Natural Science Foundation of China (Grant No. 61671425), and the State Key Laboratory of Resources and Environmental Informational System of China.

References

Adams, J.B., Smith, M.O., & Johnson, P.E., 1986. Spectral Mixture Modeling - a New Analysis of Rock and Soil Types at the Viking Lander-1 Site. *J. Geophys. Res-Solid 91*:8098-8112. http://dx.doi.org/10.1029/JB091iB08p08098.

Atkinson, P.M., 2005. Sub-pixel target mapping from soft-classified, remotely sensed imagery. *Photogramm. Eng. Remote Sens. 71*:839-846. http://dx.doi.org/10.14358/PERS.71.7.839.

Bentley, J.L., 1975. Multidimensional Binary Search Trees Used for Associative Searching. *Commun. Acm 18*:509-517. http://dx.doi.org/10.1145/361002.361007.

Brown, M., Lewis, H.G., & Gunn, S.R., 2000. Linear spectral mixture models and support vector machines for remote sensing. *IEEE Trans. Geosci. Remote Sens. 38*:2346-2360. http://dx.doi.org/10.1109/36.868891.

Chang, C.I., & Plaza, A., 2006. A fast iterative algorithm for implementation of pixel purity index. *IEEE Geosci. Remote Sens. Lett. 3*:63-67. http://dx.doi.org/10.1109/Lgrs.2005.856701.

Dong, C., Loy, C.C., He, K.M., & Tang, X.O., 2016. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans. Pattern Anal. Mach. Intell. 38*:295-307. http://dx.doi.org/10.1109/Tpami.2015.2439281.

Foody, G.M., 2001. Monitoring the magnitude of land-cover change around the southern limits of the Sahara. *Photogramm. Eng. Remote Sens. 67*:841-847. http://dx.doi.org/

Foody, G.M., 2002. Hard and soft classifications by a neural network with a non-exhaustively defined set of classes. *Int. J. Remote Sens. 23*:3853-3864. http://dx.doi.org/10.1080/01431160110109570.

Foody, G.M., & Doan, H.T.X., 2007. Variability in soft classification prediction and its implications for sub-pixel scale change detection and super resolution mapping. *Photogramm. Eng. Remote Sens. 73*:923-933. http://dx.doi.org/10.14358/PERS.73.8.923.

Foody, G.M., Lucas, R.M., Curran, P.J., & Honzak, M., 1997. Non-linear mixture modelling without end-members using an artificial neural network. *Int. J. Remote Sens. 18*:937-953. http://dx.doi.org/10.1080/014311697218845.

Freeman, W.T., Jones, T.R., & Pasztor, E.C., 2002. Example-based super-resolution. *IEEE Comput. Graph. Appl. 22*:56-65. http://dx.doi.org/10.1109/38.988747.

Gao, F., Masek, J., Schwaller, M., & Hall, F., 2006. On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance. *IEEE Trans. Geosci. Remote Sens. 44*:2207-2218. http://dx.doi.org/10.1109/Tgrs.2006.872081.

Gevaert, C.M., & Garcia-Haro, F.J., 2015. A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion. *Remote Sens. Environ. 156*:34-44. http://dx.doi.org/10.1016/j.rse.2014.09.012.

Gong, P., Wang, J., Yu, L., Zhao, Y.C., Zhao, Y.Y., Liang, L., et al., 2013. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. *Int. J. Remote Sens. 34*:2607-2654. http://dx.doi.org/10.1080/01431161.2012.748992.

Goodwin, N., Coops, N.C., & Stone, C., 2005. Assessing plantation canopy condition from airborne imagery using spectral mixture analysis and fractional abundances. *Int. J. Appl. Earth Obs. 7*:11-28. http://dx.doi.org/10.1016/j.jag.2004.10.003.

Hansen, M.C., & Loveland, T.R., 2012. A review of large area monitoring of land cover change using Landsat data. *Remote Sens. Environ. 122*:66-74. http://dx.doi.org/10.1016/j.rse.2011.08.024.

He, D., Zhong, Y.F., Feng, R.Y., & Zhang, L.P., 2016. Spatial-Temporal Sub-Pixel Mapping Based on Swarm Intelligence Theory. *Remote Sens. 8*:894. http://dx.doi.org/10.3390/rs8110894.

Heylen, R., Burazerovic, D., & Scheunders, P., 2011. Fully Constrained Least Squares Spectral Unmixing by Simplex Projection. *IEEE Trans. Geosci. Remote Sens. 49*:4112-4122. http://dx.doi.org/10.1109/TGRS.2011.2155070.

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

Huang, B., & Song, H.H., 2012. Spatiotemporal Reflectance Fusion via Sparse Representation. *IEEE Trans. Geosci. Remote Sens. 50*:3707-3716. http://dx.doi.org/10.1109/Tgrs.2012.2186638.

Keshava, N., & Mustard, J.F., 2002. Spectral unmixing. *IEEE Signal Proc. Mag. 19*:44-57. http://dx.doi.org/10.1109/79.974727.

Kim, K.I., & Kwon, Y., 2010. Single-Image Super-Resolution Using Sparse Regression and Natural Image Prior. *IEEE Trans. Pattern Anal. Mach. Intell. 32*:1127-1133. http://dx.doi.org/10.1109/Tpami.2010.25.

Li, X.D., Du, Y., & Ling, F., 2015. Sub-pixel-scale Land Cover Map Updating by Integrating Change Detection and Sub-Pixel Mapping. *Photogramm. Eng. Remote Sens. 81*:59-67. http://dx.doi.org/10.14358/PERS.81.1.59.

Li, X.D., Ling, F., Foody, G.M., & Du, Y., 2016. A Superresolution Land-Cover Change Detection Method Using Remotely Sensed Images With Different Spatial Resolutions. *IEEE Trans. Geosci. Remote Sens. 54*:3822-3841. http://dx.doi.org/10.1109/Tgrs.2016.2528583.

Li, X.D., Ling, F., Foody, G.M., Ge, Y., Zhang, Y.H., & 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 Sens. Environ. 196*:293-311. http://dx.doi.org/http://dx.doi.org/10.1016/j.rse.2017.05.011.

Ling, F., Foody, G.M., Li, X.D., Zhang, Y.H., & Du, Y., 2016a. Assessing a Temporal Change Strategy for Sub-Pixel Land Cover Change Mapping from Multi-Scale Remote Sensing Imagery. *Remote Sens. 8*:642. http://dx.doi.org/10.3390/rs8080642.

Ling, F., Li, W.B., Du, Y., & Li, X.D., 2011. Land Cover Change Mapping at the Subpixel Scale With Different Spatial-Resolution Remotely Sensed Imagery. *IEEE Geosci. Remote Sens. Lett. 8*:182-186. http://dx.doi.org/10.1109/Lgrs.2010.2055034.

Ling, F., Li, X.D., Xiao, F., & Du, Y., 2014. Superresolution Land Cover Mapping Using Spatial Regularization. *IEEE Trans. Geosci. Remote Sens. 52*:4424-4439. http://dx.doi.org/10.1109/TGRS.2013.2281992.

Ling, F., Zhang, Y.H., Foody, G.M., Li, X.D., Zhang, X.H., Fang, S.M., et al., 2016b. Learning-Based Superresolution Land Cover Mapping. *IEEE Trans. Geosci. Remote Sens. 54*:3794-3810. http://dx.doi.org/10.1109/Tgrs.2016.2527841.

Lu, D.S., Hetrick, S., Moran, E., & Li, G.Y., 2010. Detection of urban expansion in an urban-rural landscape with multitemporal QuickBird images. *J. Appl. Remote Sens. 4*:041880. http://dx.doi.org/10.1117/1.3501124.

Lu, D.S., & Weng, Q.H., 2004. Spectral mixture analysis of the urban landscape in Indianapolis with landsat ETM+ imagery. *Photogramm. Eng. Remote Sens. 70*:1053-1062. http://dx.doi.org/10.14358/PERS.70.9.1053.

Lu, D.S., & Weng, Q.H., 2009. Extraction of urban impervious surfaces from an IKONOS image. *Int. J. Remote Sens. 30*:1297-1311. http://dx.doi.org/10.1080/01431160802508985.

Michishita, R., Jiang, Z.B., & Xu, B., 2012. Monitoring two decades of urbanization in the Poyang Lake area, China through spectral unmixing. *Remote Sens. Environ. 117*:3-18. http://dx.doi.org/10.1016/j.rse.2011.06.021.

Ni, K.S., & Nguyen, T.Q., 2007. Image superresolution using support vector regression. *IEEE Trans. Image Process. 16*:1596-1610. http://dx.doi.org/10.1109/Tip.2007.896644.

Powell, R.L., Roberts, D.A., Dennison, P.E., & Hess, L.L., 2007. Sub-pixel mapping of urban land cover using multiple endmember spectral mixture analysis: Manaus, Brazil. *Remote Sens. Environ. 106*:253-267. http://dx.doi.org/10.1016/j.rse.2006.09.005.

Pu, R., Xu, B., & Gong, P., 2003. Oakwood crown closure estimation by unmixing Landsat TM data. *Int. J. Remote Sens. 24*:4433-4445. http://dx.doi.org/10.1080/0143116031000095989.

Rosenthal, W., & Dozier, J., 1996. Automated mapping of montane snow cover at subpixel resolution from the Landsat Thematic Mapper. *Water Resour. Res. 32*:115-130. http://dx.doi.org/10.1029/95WR02718.

Shen, H.F., Wu, P.H., Liu, Y.L., Ai, T.H., Wang, Y., & Liu, X.P., 2013. A spatial and temporal reflectance fusion model considering sensor observation differences. *Int. J. Remote Sens. 34*:4367-4383. http://dx.doi.org/10.1080/01431161.2013.777488.

Song, H.H., & Huang, B., 2013. Spatiotemporal Satellite Image Fusion Through One-Pair Image Learning. *IEEE Trans. Geosci. Remote Sens. 51*:1883-1896. http://dx.doi.org/10.1109/Tgrs.2012.2213095.

Townshend, J., Justice, C., Li, W., Gurney, C., & Mcmanus, J., 1991. Global Land Cover Classification by Remote-Sensing - Present Capabilities and Future Possibilities. *Remote Sens. Environ. 35*:243-255. http://dx.doi.org/10.1016/0034-4257(91)90016-Y.

Vikhamar, D., & Solberg, R., 2003. Subpixel mapping of snow cover in forests by optical remote sensing. *Remote Sens. Environ. 84*:69-82. http://dx.doi.org/10.1016/S0034-4257(02)00098-6.

Vivone, G., Alparone, L., Chanussot, J., Dalla Mura, M., Garzelli, A., Licciardi, G.A., et al., 2015. A Critical Comparison Among Pansharpening Algorithms. *IEEE Trans. Geosci. Remote Sens. 53*:2565-2586. http://dx.doi.org/10.1109/TGRS.2014.2361734.

Wang, Q.M., Atkinson, P.M., & Shi, W.Z., 2015. Fast Subpixel Mapping Algorithms for Subpixel Resolution Change Detection. *IEEE Trans. Geosci. Remote Sens. 53*:1692-1706. http://dx.doi.org/10.1109/Tgrs.2014.2346535.

Wang, Q.M., Shi, W.Z., & Atkinson, P.M., 2016. Spatiotemporal Subpixel Mapping of Time-Series Images. *IEEE Trans. Geosci. Remote Sens. 54*:5397-5411. http://dx.doi.org/10.1109/Tgrs.2016.2562178.

Wang, Z., & Bovik, A.C., 2002. A universal image quality index. *IEEE Signal Process. Lett. 9*:81-84. http://dx.doi.org/10.1109/97.995823.

Weng, Q.H., Hu, X.F., & Liu, H., 2009. Estimating impervious surfaces using linear spectral mixture analysis with multitemporal ASTER images. *Int. J. Remote Sens. 30*:4807-4830. http://dx.doi.org/10.1080/01431160802665926.

Wu, C.S., & Murray, A.T., 2003. Estimating impervious surface distribution by spectral mixture analysis. *Remote Sens. Environ. 84*:493-505. http://dx.doi.org/10.1016/S0034-4257(02)00136-0.

Wu, K., Du, Q., Wang, Y., & Yang, Y.T., 2017. Supervised Sub-Pixel Mapping for Change Detection from Remotely Sensed Images with Different Resolutions. *Remote Sens. 9*:284.

http://dx.doi.org/10.3390/rs9030284.

Xu, Y., & Huang, B., 2014. A Spatio-emporal Pixel-Swapping Algorithm for Subpixel Land Cover Mapping. *IEEE Geosci. Remote Sens. Lett. 11*:474-478. http://dx.doi.org/10.1109/LGRS.2013.2268153.

Xu, Y., Lin, L., & Meng, D., 2017. Learning-Based Sub-Pixel Change Detection Using Coarse Resolution Satellite Imagery. *Remote Sens. 9*:709. http://dx.doi.org/10.3390/rs9070709.

Yang, X., & Lo, C.P., 2002. Using a time series of satellite imagery to detect land use and land cover changes in the Atlanta, Georgia metropolitan area. *Int. J. Remote Sens. 23*:1775-1798. http://dx.doi.org/10.1080/01431160110075802.

Zhang, L.P., Zhang, L.F., & Du, B., 2016. Deep Learning for Remote Sensing Data A technical tutorial on the state of the art. *IEEE Geosci. Remote Sens. Mag. 4*:22-40. http://dx.doi.org/10.1109/Mgrs.2016.2540798.

Zhang, Y.H., Atkinson, P.M., Li, X.D., Ling, F., Wang, Q.M., & Du, Y., 2017. Learning-Based Spatial-Temporal Superresolution Mapping of Forest Cover With MODIS Images. *IEEE Trans. Geosci. Remote Sens. 55*:600-614. http://dx.doi.org/10.1109/TGRS.2016.2613140.

Zhang, Y.H., Du, Y., Ling, F., Fang, S.M., & Li, X.D., 2014. Example-Based Super-Resolution Land Cover Mapping Using Support Vector Regression. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 7*:1271-1283. http://dx.doi.org/10.1109/Jstars.2014.2305652.

Zhang, Y.H., Ling, F., Li, X.D., & Du, Y., 2015. Super-Resolution Land Cover Mapping Using Multiscale Self-Similarity Redundancy. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 8*:5130-5145. http://dx.doi.org/10.1109/Jstars.2015.2480120.

Zhong, Y.F., Wu, Y.Y., Xu, X., & Zhang, L.P., 2015. An Adaptive Subpixel Mapping Method Based on MAP Model and Class Determination Strategy for Hyperspectral Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens. 53*:1411-1426. http://dx.doi.org/10.1109/TGRS.2014.2340734.

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

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

Zhu, Z., & Woodcock, C.E., 2014. Continuous change detection and classification of land cover using all available Landsat data. *Remote Sens. Environ. 144*:152-171. http://dx.doi.org/10.1016/j.rse.2014.01.011.

Zhukov, B., Oertel, D., Lanzl, F., & Reinhackel, G., 1999. Unmixing-based multisensor multiresolution image fusion. *IEEE Trans. Geosci. Remote Sens. 37*:1212-1226. http://dx.doi.org/10.1109/36.763276.

Zurita-Milla, R., Clevers, J.G.P.W., & Schdepman, M.E., 2008. Unmixing-based Landsat TM and MERIS FR data fusion. *IEEE Geosci. Remote Sens. Lett. 5*:453-457. http://dx.doi.org/10.1109/LGRS.2008.919685.