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

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

[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, China*

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

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

*Corresponding author at Key Laboratory of Monitoring and Estimate for Environment and Disaster of Hubei Province, Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan 430077, China

E-mail address: lingf@whigg.ac.cn (F. Ling)

ABSTRACT

Studies of land cover dynamics would benefit greatly from the generation of land cover maps at both fine spatial and temporal resolutions. Fine spatial resolution images are usually acquired relatively infrequently, whereas coarse spatial resolution images may be acquired with a high repetition rate but may not capture the spatial detail of the land cover mosaic of the region of interest. Traditional image spatial–temporal fusion methods focus on the blending of pixel spectra reflectance values and do not directly provide land cover maps or information on land cover dynamics. In this research, a novel Spatial–Temporal remotely sensed Images and land cover Maps Fusion Model (STIMFM) is proposed to produce land cover maps at both fine spatial and temporal resolutions using a series of coarse spatial resolution images together with a few fine spatial resolution land cover maps that pre- and post-date the series of coarse spatial resolution images. STIMFM integrates both the spatial and temporal dependences of fine spatial resolution pixels and outputs a series of fine spatial–temporal resolution land cover maps instead of reflectance images, which can be used directly for studies of land cover dynamics. Here, three experiments based on simulated and real remotely sensed images were undertaken to evaluate the STIMFM for studies of land cover change. These experiments included comparative assessment of methods based on single-date image such as the super-resolution approaches (e.g., pixel swapping-based super-resolution mapping) and the state-of-the-art spatial–temporal fusion approach that used the Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model (ESTARFM) and the Flexible Spatiotemporal DAta Fusion model (FSDAF) to predict

the fine-resolution images, in which the maximum likelihood classifier and the automated land cover updating approach based on integrated change detection and classification method were then applied to generate the fine-resolution land cover maps. Results show that the methods based on single-date image failed to predict the pixels of changed and unchanged land cover with high accuracy. The land cover maps that were obtained by classification of the reflectance images outputted from ESTARFM and FSDAF contained substantial misclassification, and the classification accuracy was lower for pixels of changed land cover than for pixels of unchanged land cover. In addition, STIMFM predicted fine spatial–temporal resolution land cover maps from a series of Landsat images and a few Google Earth images, to which ESTARFM and FSDAF that require correlation in reflectance bands in coarse and fine images cannot be applied. Notably, STIMFM generated higher accuracy for pixels of both changed and unchanged land cover in comparison with other methods.

Keywords: Spatial temporal fusion; Super-resolution mapping; Endmember extraction.

1. Introduction

Land cover maps are one of the most fundamental datasets used in many scientific fields and are often produced from remotely sensed images (Bartholome and Belward 2005; Friedl et al. 2002). A wide variety of remote sensing systems have been developed, and hence, images are available with different spatial and temporal resolutions, thereby allowing the production of land cover maps at different spatial and temporal scales. With most satellite remote sensing systems, a trade-off typically exists between spatial and temporal resolution. In general, fine spatial resolution remote sensors can acquire images that provide spatially detailed land cover information, but their relatively coarse temporal resolution limits their usage in monitoring rapid land cover changes. By contrast, coarse spatial resolution remotely sensed images can often be acquired at a fine temporal resolution that provides a repetition rate suitable for the detection of rapid land cover changes but are unable to represent the spatial detail of the land cover mosaic. To realize the full potential of remote sensing as a source of information on land cover change, a method that allows the production of land cover maps with both fine spatial and temporal resolutions is required. Such maps could be obtained by combining all available remotely sensed images of varying spatial and temporal resolution to form a series of fine-resolution land cover maps.

Recently, spatial–temporal image fusion, which aims to produce fine spatial and temporal resolution remotely sensed images from images with different spatial and temporal resolutions, has become a promising means to address the trade-off between spatial and temporal resolution (Gevaert and Garcia-Haro 2015; Zhu et al. 2016).

Spatial–temporal data fusion methods can be categorized into weighted function based methods, unmixing-based methods, and dictionary-pair learning based methods (Zhu et al. 2016). Among the weighted function based methods, the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) proposed by Gao et al. (2006) was developed first and is one of the most popular spatial–temporal image fusion methods. By fusing coarse spatial resolution Moderate Resolution Imaging Spectroradiometer (MODIS) and fine spatial resolution Landsat sensor images, STARFM can predict Landsat-like reflectance images with the spatial resolution of Landsat and the temporal resolution of MODIS. A number of studies have suggested improvements to STARFM, including studies of forest disturbance (Hilker et al. 2009a), and in heterogeneous regions (Zhu et al. 2010), as well as in gap filling to reduce the negative effects of cloud (Gevaert and Garcia-Haro 2015). STARFM and the improved models based on it have been mainly used to detect reflectance changes caused by processes such as phenology over large areas, and used to generate dense time series of Landsat-like data (Hilker et al. 2009b), enhance land cover classification (Jia et al. 2014), and predict key environmental variations such as evapotranspiration (Anderson et al. 2011) and temperature (Hilker et al. 2009b). Other spatial–temporal image fusion models, such as the unmixing-based algorithm that extracts endmembers on the basis of linear spectral mixture model and assigns the unmixed reflectance to fine spatial resolution pixels (Huang and Zhang 2014; Zhukov et al. 1999; Zurita-Milla et al. 2009) and the dictionary-pair learning based methods, which capture features from the coarse- and fine-resolution image pairs used for predicting fine-resolution image (Huang and Song

2012), have also been proposed and applied to Landsat and MODIS images in recent years (Amoros-Lopez et al. 2013; Wu et al. 2012).

Generally, spatial–temporal image fusion models aim to generate a series of continuous reflectance values instead of discrete categorical values. A further image classification step is needed to produce from the reflectance images a corresponding series of land cover maps for the study of land cover class dynamics (Jia et al. 2014). The use of these methods for generating land cover maps and monitoring land cover changes often suffers from two important limitations.

First, most spatial–temporal image fusion algorithms assume that land cover type does not change during the data observation period (Fu et al. 2013; Gao et al. 2006; Zhu et al. 2010). Previous research has shown that STARFM does not deal well with abrupt land cover changes. Song and Huang (2013) showed that STARFM failed to fuse the pixel reflectance accurately in a study of land cover change in an urban area. The Enhanced STARFM (ESTARFM) is often better than STARFM for studies of heterogeneous landscapes (Zhu et al. 2010) but can be worse than STARFM for predicting abrupt changes of land cover type (Emelyanova et al. 2013). The Spatial Temporal Adaptive Algorithm for mapping Reflectance CHange (STAARCH) improves STARFM's performance when land cover type change and disturbance exist, but it is more suitable for spatial–temporal fusion of forest land cover (Hilker et al. 2009a). The Flexible Spatiotemporal DAta Fusion model (FSDAF) can predict Landsat-like reflectance values with both gradual change and land cover type change, but it cannot capture tiny changes in land cover type, such as when only a few fine

pixels experienced land cover type change and the change is invisible in the coarse-resolution image (Zhu et al. 2016). Similar to STARFM, the unmixing-based spatial–temporal reflectance fusion methods consider only the change in endmember spectra but not in land cover types (Huang and Zhang 2014; Zhukov et al. 1999; Zurita-Milla et al. 2009).

Second, most spatial–temporal image fusion methods need one or more observed pairs of coarse- and fine-resolution images for training and require the coarse- and fine-resolution remotely sensed data from different satellite sensors to be mutually comparable and correlated. All the weighted function based methods, including STARFM, ESTARFM, STAARCH, and all the dictionary-pair learning-based methods need one or more observed pairs of coarse- and fine-resolution images, which have comparable reflectance bands, for training (Gao et al. 2006; Gevaert and Garcia-Haro 2015; Zhu et al. 2010). These methods mainly focus on predicting Landsat-like remotely sensed images with MODIS repetition rates. However, these methods cannot deal with other satellite images, which have uncorrelated reflectance bands, and are thus limited in the use of land cover change analysis. For instance, in regional-scale land cover analysis, the detection of very-high-resolution land cover changes at high temporal resolutions is required. In general, we can obtain a series of Landsat images and a few very-high-resolution images such as panchromatic aerial photograph. The weighted function based and dictionary-pair learning based methods cannot fuse these data because the very-high-resolution images usually have different reflectance bands compared with Landsat images.

The spatial–temporal image fusion methods aim to produce fine spatial–temporal resolution reflectance images rather than land cover maps. The fused fine-resolution images have many applications, such as phenology analysis. If the aim is to generate a sequence of land cover maps from the reflectance images from which land cover change trajectories may be extracted, then a further image classification analysis is still required, which may introduce uncertainty and error in the land cover maps. First, the classification of an image series can be complex and laborious. Training statistics are required to inform classification analysis, and these may vary in quality in time due to issues such as phenology. Moreover, the classification is also problematic, with the potential for different classifiers to generate dissimilar land cover maps from the same training data. Traditional classifiers applied to mono-temporal image may also ignore the temporal information contained in a series of images and thereby produce a classification of sub-optimal accuracy. The spatial–temporal–based image classifier has the advantage in taking both the spatial and temporal links between neighboring pixels (Cai et al. 2014), but is challenging to use for voluminous image series (Liu and Cai 2012; Liu et al. 2006). Finally, the spatial–temporal image fusion models generate a large volume of fine spatial–temporal resolution reflectance images as the intermediate data to be used for the production of land cover maps. This situation may represent practical challenges in terms of data access and storage.

Given the concerns with the spatial–temporal reflectance fusion model for producing land cover maps, a more appropriate fusion approach could be based on directly downscaling the coarse spatial resolution image series to fine spatial resolution

land cover maps rather than reflectance images, with the aid of information derived from a few fine spatial resolution images that may be available. Chen et al. (2015) updated land cover maps from downscaled Normalized Difference Vegetation Index (NDVI) time-series data from MODIS, a current Landsat image, and a Landsat image that pre-dates it. The NDVI time-series data are used as ancillary data to extract changed pixels in the Landsat images, and the labels of changed pixels are determined using the current Landsat image. Thus, this method can update fine-resolution land cover maps with Landsat repetition rates based on available Landsat images, but cannot predict fine-resolution land cover maps with MODIS repetition rates. In addition, a major problem with this approach is that a large proportion of coarse spatial resolution image pixels may be of mixed land cover composition. A possible solution of this problem is to use the fractional land cover class composition images that can be extracted via a spectral unmixing analysis. A comparison of the obtained fraction images indicates the change, if any, in land cover that has occurred in the time period between the dates of image acquisition (Lu et al. 2004). This approach can potentially reveal important temporal land cover information, such as land cover modification and conversion (Foody 1999; Lu et al. 2011). Unfortunately, these approaches show only the change in the fraction of the area that is represented by each coarse-resolution pixel and not the geographical location of the change. Information on the location of change might be obtained through a super-resolution analysis (Li et al. 2016; Wang et al. 2015).

Super-resolution land cover mapping (SRM) is a promising technique used to generate land cover maps with a finer spatial resolution than the input data and is

typically viewed as a post-processing approach applied after a spectral unmixing analysis. SRM predicts the spatial distribution of each land cover class in the area represented by each coarse spatial resolution pixel and provides more fine spatial resolution land cover information than spectral unmixing (Atkinson 2009). Various approaches have been proposed (Atkinson 2005; Ge et al. 2016; Kasetkasem et al. 2005; Ling et al. 2016; Tatem et al. 2003), and SRM has been used in many fields, including the extraction of waterlines (Foody et al. 2005), rural land cover (Tatem et al. 2003), refining the estimation of ground control point location (Foody 2002), land cover change detection (Wang et al. 2015), land cover map updating (Li et al. 2015b), and wetland inundation analysis (Li et al. 2015a).

Traditionally, SRM is applied to a mono-temporal coarse spatial resolution image dataset. The SRM solution space is large because SRM predicts land cover maps with finer spatial resolution than the input data, and it can provide multiple plausible solutions that satisfy the constraints of the SRM analysis. A fine spatial resolution land cover map that pre- or post-dates the coarse spatial resolution image could be used to provide fine spatial resolution information to constrain and enhance the SRM solution (Li et al. 2015b; Ling et al. 2011; Wang et al. 2015; Xu and Huang 2014). Although the accuracy of SRM may be increased through the use of multi-temporal data, challenges remain, especially if a time series of images are used. Often, a sequence of coarse spatial resolution images together with a few fine spatial resolution images that pre- and post-date the coarse-resolution images are available. Applying existing SRMs to each coarse-resolution image without or with only one fine spatial resolution map

211 that pre- or post-dates it fails to account for the temporal dependence in the image series
212 and fails to fully exploit the available information. The construction of SRM that
213 considers the temporal dependences of a coarse-resolution image from fine-resolution
214 maps that pre- and post-date it is necessary for a fuller reconstruction of land cover
215 dynamics.

216 The objective of this paper is to propose a Spatial–Temporal remotely sensed
217 Images and land cover Maps Fusion Model (STIMFM). The inputs to STIMFM are a
218 series of coarse spatial resolution multi-spectral remotely sensed images and few fine
219 spatial resolution land cover maps that pre- and post-date the coarse spatial resolution
220 image series. The fine spatial resolution land cover maps can be obtained from various
221 data sources, such as through classification of remotely sensed images or maps
222 produced conventionally from field survey. As a result, the input to STIMFM is more
223 general than that of other spatial–temporal image fusion models. Critically, STIMFM
224 outputs a series of fine spatial–temporal resolution land cover maps, not reflectance
225 images. In addition, STIMFM takes information on class temporal dependence that
226 exists in different images into account and is able to deal with land cover change.
227 STIMFM was compared with a set of alternative methods. The latter includes two SRM
228 methods that use a mono-temporal coarse spatial resolution image as input and two
229 spatial–temporal image fusion methods, namely, the ESTARFM which adopts the
230 coarse spatial resolution image and two fine and coarse spatial resolution image pairs
231 that pre- and post-date the coarse-resolution image as input, and the FSDAF which
232 adopts the coarse spatial resolution image and one fine and coarse spatial resolution

image pair that pre- or post-date the coarse-resolution image as input.

2. Methods

2.1. STIMFM

A coarse spatial resolution image **Y** contains $I \times J$ pixels. Fine spatial resolution land cover maps of the same geographical region are $\mathbf{X}_{pre}$ and $\mathbf{X}_{post}$, which are temporally close to and pre- or post-date **Y**, respectively. $\mathbf{X}_{pre}$ and $\mathbf{X}_{post}$ contain $I \times s \times J \times s$ pixels, where s is the scale or zoom factor and each coarse spatial resolution pixel contains $s \times s$ fine spatial resolution pixels. C land cover classes are present in $\mathbf{X}_{pre}$ and $\mathbf{X}_{post}$. The STIMFM predicts a fine spatial resolution land cover map **X** at the time of coarse-resolution image **Y** observation, and has $I \times s \times J \times s$ pixels, each of which has a land cover class label in C. STIMFM produces a series of fine spatial and fine temporal resolution land cover maps. It uses a series of coarse spatial resolution remotely sensed images and a few fine spatial resolution land cover maps as input (Fig. 1). STIMFM comprises several main steps, including spectral endmember estimation, analysis of land cover class fraction temporal change, objective function construction, and model optimization. The STIMFM flowchart is shown in Fig. 2.

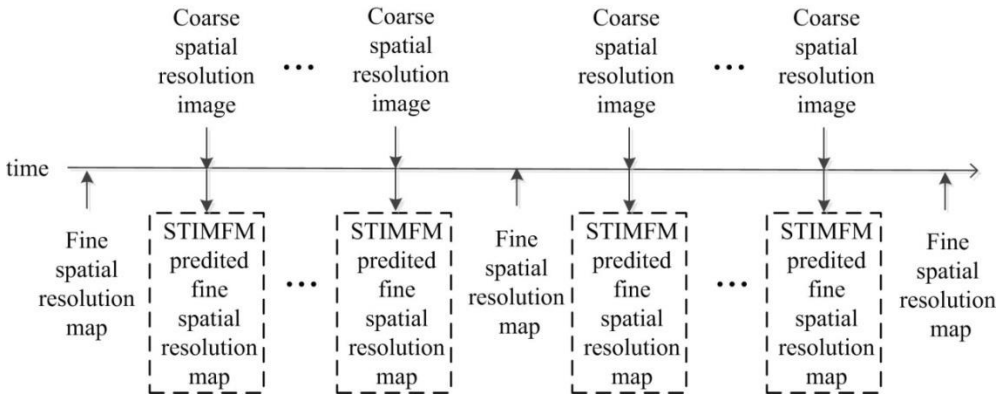


Fig. 1 Production of a series of fine spatial and temporal resolution land cover maps from a series of coarse spatial resolution remotely sensed images and a few fine spatial resolution land cover maps in STIMFM.

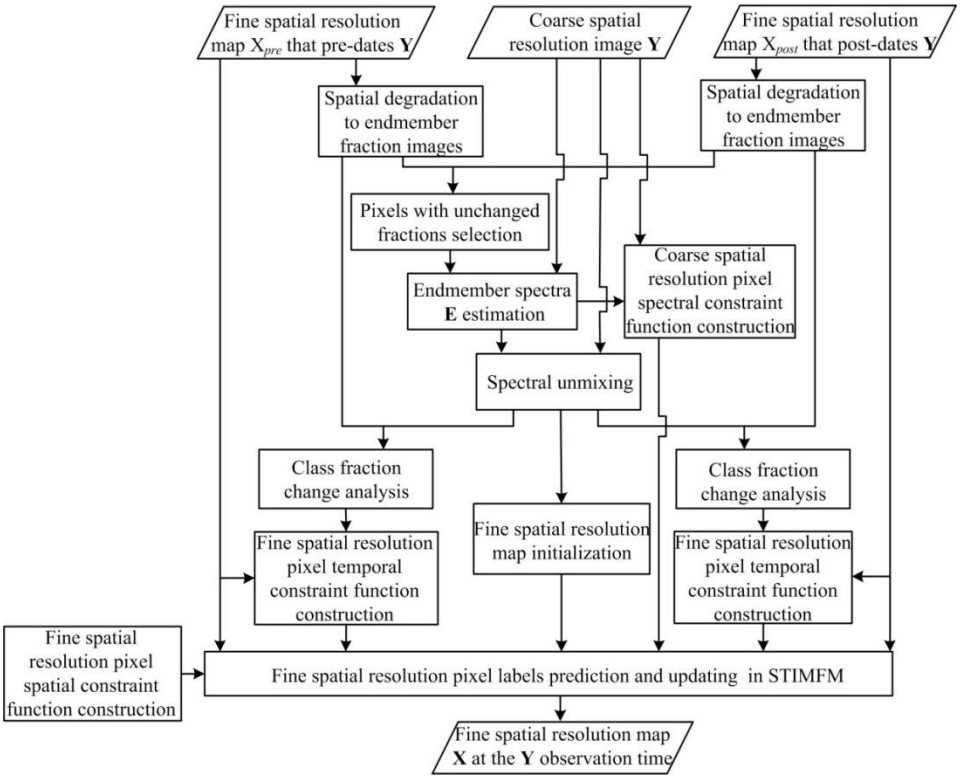


Fig. 2 Flowchart of STIMFM.

2.2. Spectral endmember estimation

In STIMFM, endmembers that are representative of the spectra of pure land cover classes are estimated for coarse spatial resolution remotely sensed images. Endmember spectra need to be extracted for each coarse spatial resolution image in the dataset as differences may be expected in a time series due to issues such as phenology or variation in image acquisition properties (e.g., angular viewing geometry). Although many endmember extraction algorithms are available, they are not directly used in STIMFM because spectral endmembers are difficult to extract accurately from coarse spatial resolution remotely sensed images due to the small proportion of pure pixels that are typically contained. Information for the estimation of endmembers is instead provided by the fine spatial resolution land cover maps that pre- and post-date the coarse spatial resolution image time series.

The land cover classes are defined in the fine spatial resolution land cover maps. For each coarse spatial resolution remotely sensed image, the linear mixture model (LMM) is applied in STIMFM to estimate endmember spectra. With the LMM, the spectral response of each coarse spatial resolution pixel is viewed as being composed of a weighted linear sum of the endmember spectra within that pixel, in which the weights are determined by the relative areal proportions of each endmember (Settle and Drake 1993). On the basis of the linear mixing assumption, the spectral signature $\boldsymbol{y}_{ij}$ for the coarse spatial resolution pixel (i,j) in $\mathbf{Y}$ can be represented by

$$\boldsymbol{y}_{ij} = \mathbf{E}\boldsymbol{f}_{ij} \quad (1)$$

where $\boldsymbol{y}_{ij}$ is a $B \times 1$ spectral vector. B is the number of spectral bands. $\mathbf{E}$ is a $B \times C$ matrix that represents the endmembers used for $\mathbf{Y}$. $\boldsymbol{f}_{ij}$ is the $C \times 1$ vector that represents

fractions of all endmembers in the pixel (i,j) in **Y**.

Theoretically, to solve for **E** with $B \times C$ unknown variables, at least $B \times C$ equations are required. l (l>C) coarse pixels are collected to compose a system of linear mixture equations

$$[\boldsymbol{y}_1,\ \boldsymbol{y}_2,\cdots,\boldsymbol{y}_l] = \mathbf{E}[\boldsymbol{f}_1,\ \boldsymbol{f}_2,\cdots,\boldsymbol{f}_l] \tag{2}$$

where $\boldsymbol{y}_l$ is the spectral signature for the l-th coarse spatial resolution pixel in **Y**, and $\boldsymbol{f}_l$ is the fraction vector of different classes in the l-th coarse spatial resolution pixel in **Y**. **E** can be solved on the basis of the inversion of Eq. (2) by computing a least squares best fit solution

$$[\tilde{\mathbf{E}}] = \arg\min\left(\sum_{n=1}^{l}\left\|\boldsymbol{y}_n - \mathbf{E}f_n\right\|^2\right) \tag{3}$$

where $\boldsymbol{y}_n$ is the n-th coarse spatial resolution pixel's spectral signature in **Y**, and $\boldsymbol{f}_n$ is the fraction vector in the n-th coarse spatial resolution pixel in **Y**. $\|\cdot\|^2$ is the L2 norm of the residual vector. "argmin" means the minimizing argument of the function.

A number of coarse spatial resolution pixels in **Y** with known endmember fractions are sought to solve Eq. (3). For each class, the focus is a set of coarse-resolution pixels that have the least changed fractions of that class during the time period covered by $\mathbf{X}_{pre}$ and $\mathbf{X}_{post}$. To avoid the collinearity problem in the use of LMM (van der Meer and Jia 2012), m coarse-resolution pixels that have the highest fraction of a given class (i.e., the m purest coarse-resolution pixels of the class) among the selected set of coarse-resolution pixels are used. All the $m \times C$ coarse spatial resolution pixels are used for endmember estimation in Eq. (3), which can be solved by computing a least squares best fit solution. Assuming the fractions of the $m \times C$ coarse spatial resolution pixels

are unchanged, the fractions of these coarse pixels in $\mathbf{X}_{pre}$ or $\mathbf{X}_{post}$ are used as a substitute of those in $\mathbf{Y}$. The fractions in $\mathbf{X}_{pre}$ and $\mathbf{X}_{post}$ are produced through a spatial degradation process by dividing the number of fine spatial resolution pixels of each class by the total number of fine spatial resolution pixels in a coarse spatial resolution pixel (i.e., s^2).

2.3. Analysis of land cover class fraction temporal change

With the estimated endmembers, class fraction images that represent the area percentage of a pixel occupied by each endmember can be extracted from coarse spatial resolution image $\mathbf{Y}$ using the estimated endmember spectra $\mathbf{E}$ and on the basis of the mean square error minimization criterion of the LMM

$$[\tilde{\boldsymbol{f}}_{ij}] = \arg\min \left\| \boldsymbol{y}_{ij} - \mathbf{E}\boldsymbol{f}_{ij} \right\|^2 \tag{4}$$

$$0 \le f_{ijc} \le 1, \quad c = 1, \cdots, C \tag{5}$$

$$\sum_{c=1}^{C} f_{ijc} = 1 \tag{6}$$

where $\boldsymbol{f}_{ij} = \left[f_{ij1}, f_{ij2}, \cdots, f_{ijC} \right]^{\mathrm{T}}$, and f_{ijc} is the fraction value of the c-th endmember in coarse spatial resolution pixel (i,j) in $\mathbf{Y}$.

The fraction images produced from the coarse spatial resolution image by spectral unmixing, as well as those produced by spatially degrading the fine spatial resolution land cover maps that pre- and post-date the coarse spatial resolution image, provide the land cover trajectory at the acquisition times of $\mathbf{X}_{pre}$, $\mathbf{Y}$, and $\mathbf{X}_{post}$. The change of class fractions in each coarse spatial resolution pixel represents the temporal transitions between classes in the period between the dates of image acquisition. If the class fractions remain unchanged between the coarse-resolution image and fine-resolution

map that pre- or post-dates it, then the fine spatial resolution pixel class labels may probably also be unchanged during this period. In this case, the images are temporally correlated. By contrast, if the class fractions changed drastically between two images, then the fine spatial resolution pixels may have changed during this period. Thereby, the images are temporally uncorrelated. As a result, the temporal dependence between different images can be analyzed on the basis of the change in fractions in each coarse spatial resolution pixel.

Assume a_{ijk} is the k-th ($k = 1, \cdots, s^2$) fine spatial resolution pixel in the coarse spatial resolution pixel (i,j) ($i = 1, \cdots, I$, $j = 1, \cdots, J$) in the land cover map $\mathbf{X}$, $a_{ijk,pre}$ and $a_{ijk,post}$ are the k-th fine spatial resolution pixel in coarse spatial resolution pixel (i,j) in the maps $\mathbf{X}_{pre}$ and $\mathbf{X}_{post}$, and $c(a_{ijk})$, $c(a_{ijk,pre})$, and $c(a_{ijk,post})$ are land cover class labels for fine spatial resolution pixels a_{ijk}, $a_{ijk,pre}$, and $a_{ijk,post}$, respectively. The temporal dependence or correlation between fine spatial resolution pixels $a_{ijk,pre}$ and a_{ijk} during $\mathbf{X}_{pre}$ and $\mathbf{Y}$ observation period or between fine spatial resolution pixels a_{ijk} and $a_{ijk,post}$ during $\mathbf{Y}$ and $\mathbf{X}_{post}$ observation period, which is dependent on the class labels of $a_{ijk,pre}$ and a_{ijk} or the class labels of a_{ijk} and $a_{ijk,post}$ [Eqs. (7)–(8)] and the change in fractions in this coarse pixel measured by $w_{ij,pre}$ and $w_{ij,post}$ [Eqs. (9)–(10)], can be characterized as $w_{ij,pre} \times \delta(c(a_{ijk}), c(a_{ijk,pre}))$ or $w_{ij,post} \times \delta(c(a_{ijk}), c(a_{ijk,post}))$.

$$\delta(c(a_{ijk}), c(a_{ijk,pre})) = \begin{cases} 1 & \text{if} \quad c\left(a_{ijk}\right) = c\left(a_{ijk,pre}\right) \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

$$\delta(c(a_{ijk}), c(a_{ijk,post})) = \begin{cases} 1 & \text{if} \quad c\left(a_{ijk}\right) = c(a_{ijk,post}) \\ 0 & \text{otherwise} \end{cases} . \tag{8}$$

On the basis of the Kronecker delta function, Eqs. (7)–(8) return a value of 1 if the fine

spatial resolution pixel in different images have an unchanged class label, thereby indicating that different image pixels are temporally correlated, and a value of 0 if the fine spatial resolution pixel in different images have changed class labels, thereby indicating that the different image pixels are temporally uncorrelated.

The changes in fractions in coarse-resolution pixel (i,j) during $\mathbf{X}_{pre}$ and $\mathbf{Y}$ observation period and during $\mathbf{Y}$ and $\mathbf{X}_{post}$ observation period are measured by $w_{ij,pre}$ and $w_{ij,post}$ on the basis of the Gaussian model in Eqs. (9)–(10)

$$w_{ij,pre} = \exp\left(-\left\| \boldsymbol{f}_{ij} - \boldsymbol{f}_{ij,pre} \right\|^2\right) \tag{9}$$

$$w_{ij,post} = \exp\left(-\left\| \boldsymbol{f}_{ij} - \boldsymbol{f}_{ij,post} \right\|^2\right) \tag{10}$$

where $\boldsymbol{f}_{ij,pre}$ and $\boldsymbol{f}_{ij,post}$ are the land cover fraction vector in coarse pixel (i,j) in $\mathbf{X}_{pre}$ and $\mathbf{X}_{post}$ produced by spatially degrading $\mathbf{X}_{pre}$ and $\mathbf{X}_{post}$ according to the scale factor s. $w_{ij,pre}$ and $w_{ij,post}$ indicate the strength of temporal dependence between fine pixels in coarse pixel (i,j) during $\mathbf{X}_{pre}$ and $\mathbf{Y}$ observation period or during $\mathbf{Y}$ and $\mathbf{X}_{post}$ observation period. $w_{ij,pre}$ and $w_{ij,post}$ decrease with the increase in the change of fractions in Eqs. (9)–(10).

2.4. Spatial–temporal SRM model

Given the coarse spatial resolution image $\mathbf{Y}$, the fine spatial resolution maps $\mathbf{X}_{pre}$ and $\mathbf{X}_{post}$, STIMFM aims to predict the fine spatial resolution land cover map $\mathbf{X}$ at the time of $\mathbf{Y}$ observation. The optimal STIMFM result $\mathbf{X}$, given $\mathbf{Y}$, $\mathbf{X}_{pre}$, and $\mathbf{X}_{post}$, can be formulated by applying the maximum a posteriori rule in Bayesian framework, i.e., by solving the maximization problem:

$$\begin{aligned}\tilde{\mathbf{X}} &= \arg\max\left\{P^{posterior}\left(\mathbf{X}\middle|\mathbf{Y},\mathbf{X}_{pre},\mathbf{X}_{post}\right)\right\} \\ &= \arg\max\left\{\frac{1}{Z}\exp\left[-U^{posterior}\left(\mathbf{X}\middle|\mathbf{Y},\mathbf{X}_{pre},\mathbf{X}_{post}\right)\right]\right\}\end{aligned} \tag{11}$$

where Z is a normalizing constant. $P^{posterior}\left(\mathbf{X}\middle|\mathbf{Y},\mathbf{X}_{pre},\mathbf{X}_{post}\right)$ is the posterior probability of **X**, given **Y**, $\mathbf{X}_{pre}$, and $\mathbf{X}_{post}$. $U^{posterior}\left(\mathbf{X}\middle|\mathbf{Y},\mathbf{X}_{pre},\mathbf{X}_{post}\right)$ is the posterior energy function of **X**, given **Y**, $\mathbf{X}_{pre}$, and $\mathbf{X}_{post}$. The solving of (11) is complicated because it involves the optimization of a global distribution model of the entire image. Based on the Markov random field approach, the searching of the optimal **X** is equivalent to minimization the posterior energy function, which can be specified to model the spatial and temporal dependencies of pixel on its spatial and temporal neighborhoods (Cai et al. 2014; Li et al. 2014).

$$U^{posterior}\left(\mathbf{X}\middle|\mathbf{Y},\mathbf{X}_{pre},\mathbf{X}_{post}\right)=U^{spectral}(\mathbf{Y}\middle|\mathbf{X})+U^{spatial}(\mathbf{X})+U^{temporal}(\mathbf{X}\middle|\mathbf{X}_{pre},\mathbf{X}_{post}) \tag{12}$$

where $U^{spectral}(\mathbf{Y}\middle|\mathbf{X})$ is spectral constraint function that represents the inconsistency between **Y** and **X**, $U^{spatial}(\mathbf{X})$ and $U^{temporal}(\mathbf{X}\middle|\mathbf{X}_{pre},\mathbf{X}_{post})$ are the spatial and temporal constraint functions, respectively.

2.4.1 Spectral constraint function

The spectral constraint function is used to link the fine spatial resolution land cover map **X** with the observed coarse spatial resolution image **Y**. The spectral response of a coarse spatial resolution pixel in **Y** is composed of a weighted linear sum of endmember spectral responses within that pixel in the fine spatial resolution map **X** on the basis of the LMM. A synthetic coarse spatial resolution pixel spectral signature is developed for a coarse spatial resolution pixel on the basis of the endmember spectral signatures and the fraction of each endmember according to the LMM. The STIMFM spectral

385 constraint function aims to minimize the L2 norm of the residual vector between the
386 observed and synthetic coarse spatial resolution spectral signatures

387 $$U^{spectral}(\mathbf{Y}|\mathbf{X}) = \sum_{i=1}^{I}\sum_{j=1}^{J}\left\|\boldsymbol{y}_{ij} - \mathbf{E}\boldsymbol{f}_{ij}\right\|^2 \quad (13)$$

388 where the class fraction vector $\boldsymbol{f}_{ij}$ is calculated by dividing the number of fine-
389 resolution pixels of different classes in coarse-resolution pixel (i,j) by s^2 in **X**, which is
390 estimated and iteratively updated from STIMFM. $\mathbf{E}\boldsymbol{f}_{ij}$ is the synthetic spectra for
391 coarse-resolution pixel (i,j) on the basis of the LMM.

392 *2.4.2 Spatial constraint function*

393 The spatial constraint function is used to describe the spatial pattern of land cover
394 distribution. In STIMFM, the maximal spatial dependence model that aims to maximize
395 the spatial dependence between neighboring fine spatial resolution pixels was used for
396 its simplicity and effectiveness (Atkinson 2009). For a fine spatial resolution pixel a_{ijk},
397 the spatial dependence is quantified with respect to its neighboring fine spatial
398 resolution pixels. The STIMFM spatial constraint function is computed as

399 $$U^{spatial}(\mathbf{X}) = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{s^2}\left(-\lambda_s \times \sum_{l\in \mathrm{N}(a_{ijk})} \frac{1}{d\left(a_{ijk}, a_l\right)} \cdot \delta\left(c(a_{ijk}), c(a_l)\right)\right) \quad (14)$$

400 where N(a_{ijk}) is the spatial neighborhood that includes all fine spatial resolution pixels
401 inside a square window whose center is a_{ijk} (a_{ijk} itself is not included), and a_l is a
402 neighboring fine spatial resolution pixel of a_{ijk} in N(a_{ijk}). The size of the neighborhood
403 N(a_{ijk}) is W. $d\left(a_{ijk}, a_l\right)$ is the Euclidean distance between a_{ijk} and a_l. $c(a_l)$ is the land
404 cover class label for fine spatial resolution pixel a_l. $\delta\left(c(a_{ijk}), c(a_l)\right)$ equals 1 if $c(a_{ijk})$
405 and $c(a_l)$ are the same and 0 otherwise. λ_S is the spatial weight parameter. -λ_S is

multiplied because the STIMFM objective function seeks the minimal value as the optimal solution.

2.4.3 Temporal constraint function

The STIMFM temporal constraint function is used to measure the temporal dependence between the predicted fine spatial resolution map $\mathbf{X}$ and the input fine spatial resolution maps $\mathbf{X}_{pre}$ and $\mathbf{X}_{post}$. The class label of fine spatial resolution pixel a_{ijk} is temporally correlated to fine spatial resolution pixel $a_{ijk,pre}$ and $a_{ijk,post}$ in the maps $\mathbf{X}_{pre}$ and $\mathbf{X}_{post}$ depending on the class labels of a_{ijk}, $a_{ijk,pre}$ and $a_{ijk,post}$ and the strength of temporal dependences measured by the weights $w_{ij,pre}$ and $w_{ij,post}$. λ_T is the temporal weight parameter. The STIMFM temporal constraint function is written as

$$U^{temporal}(\mathbf{X}|\mathbf{X}_{pre},\mathbf{X}_{post})=\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{s^2}\left(-\lambda_T\times\left(w_{ij,pre}\times\delta\left(c(a_{ijk}),c(a_{ijk,pre})\right)+w_{ij,post}\times\delta\left(c(a_{ijk}),c(a_{ijk,post})\right)\right)\right) \quad (15)$$

2.5. Model initialization and optimization

An initial fine spatial resolution land cover map is used as input to STIMFM at the outset. The initialization map is produced according to the land cover class fraction images estimated from a coarse spatial resolution image. The fine spatial resolution pixels are randomly allocated class labels in a manner that maintains the class proportional information conveyed by a prior spectral unmixing analysis (Kasetkasem et al. 2005). The class labels in the initial fine spatial resolution land cover map are then updated iteratively. Here, the Iterative Conditional Mode (ICM) was applied to update the fine spatial resolution pixel class labels. ICM converges when no pixel class

426 labels change during two successive iterations or when a predefined number of
427 iterations have been undertaken.

428 3. Experiments and results

429 The proposed STIMFM was evaluated in three experiments. The first used Landsat
430 multi-spectral images and the National Land Cover Database (NLCD) land cover maps
431 (Landsat–NLCD). The second used MODIS and Landsat multi-spectral images
432 (MODIS–Landsat). The third used Landsat and Google Earth Images (Landsat–GEI).
433 For a rigorous assessment, several traditional approaches were used for comparison,
434 including the Pixel Swapping based SRM (PS_SRM) (Atkinson 2005), the Spatial
435 Regularization based SRM (SR_SRM) (Ling et al. 2014), the ESTARFM (Zhu et al.
436 2010), and the FSDAF model (Zhu et al. 2016).

437 PS_SRM and SR_SRM use only a mono-temporal coarse spatial resolution image
438 as input and hence do not exploit the temporal information in the land cover. By contrast,
439 ESTARFM uses a coarse spatial resolution image and pairs of coarse and fine spatial
440 resolution images that pre-and post-date it as input. ESTARFM is based on the
441 assumption that remotely sensed data from different satellite sensors observed on the
442 same, or at least very close, date are mutually comparable and correlated, and uses the
443 correlation to blend multi-source data and minimize the system biases. The FSDAF,
444 which is based on spectral unmixing analysis and a thin plate spline interpolator, is also
445 used for comparison. It requires only one pair of fine and coarse spatial resolution
446 images that pre- or post-date the coarse-resolution image.

447 ESTARFM and FSDAF output a fine spatial resolution reflectance image rather

than a land cover map. The fine spatial resolution image produced may then be classified. Two types of classification methods were applied. The first one is the Maximum Likelihood Classifier (MLC), which is one of the statistical classifiers that relies on the second-order statistics of a Gaussian probability density function for the distribution of the feature vector of each class. In MLC, a pixel is allocated to the class with which it has the highest likelihood of membership (Richards and Jia 1999). The produced final fine spatial resolution land cover maps by MLC, which are referred to as ESTARFM_MLC and FSDAF_MLC, were compared with STIMFM.

The second used classification method is the automated Land Cover updating approach based on integrated change detection and classification methods (LCupdating) produced by Chen et al. (2012). MLC used only the fused fine-resolution image as input but ignored the available fine-resolution image and land cover map that pre- or post-dates the coarse-resolution image. LCupdating was applied to the fused image from ESTARFM and FSDAF by incorporating the fine-resolution remotely sensed image and land cover map that pre-date the coarse-resolution image. LCupdating first detects changes between the input and fused fine-resolution images from ESTARFM or FSDAF and then predicts the changed pixel labels in the fused image based on the Markov random field based classifier. The ESTARFM and FSDAF incorporating LCupdating methods (ESTARFM_LCupdating and FSDAF_LCupdating) were compared with STIMFM.

The parameters of these different methods were set according to results reported in the literature and through trial and error. The STIMFM spatial weight parameter λ_S

and temporal weight parameter λ_T were set to 0.05. The neighborhood window size *W* in the STIMFM spatial constraint function was set to 2×*s*-1 (Tolpekin and Stein 2009). The number of unchanged coarse pixels, *m*, in STIMFM endmember estimation was set to 100.

3.1 Landsat–NLCD experiment

3.1.1 Data preparation

This experiment used Landsat Thematic Mapper (TM) multi-spectral images and NLCD land cover maps. The NLCD is a land cover classification scheme of Albers Equal Area projection, which has been applied consistently at a spatial resolution of 30 m across the conterminous USA primarily on the basis of Landsat satellite data. NLCD maps for the years 2001, 2006, and 2011 were used in this experiment. The NLCD 2001 was based primarily on a decision tree classification of 2001 Landsat satellite data. The NLCD 2006 and 2011 were based primarily on a decision tree classification from 2006 and 2011 Landsat satellite data, and also quantified land cover change from 2001 to 2006 and 2006 to 2011 (Homer et al. 2015; Jin et al. 2013; Xian et al. 2009). The original sixteen classes were reclassified into eight classes (Fig. 3). Subset land cover maps, each with a size of 2000 × 2000 pixels (centered at 34°4′00″N and 79°27′00″W), were acquired from NLCD 2001, 2006, and 2011 [Fig. 3(*b*–*d*)].

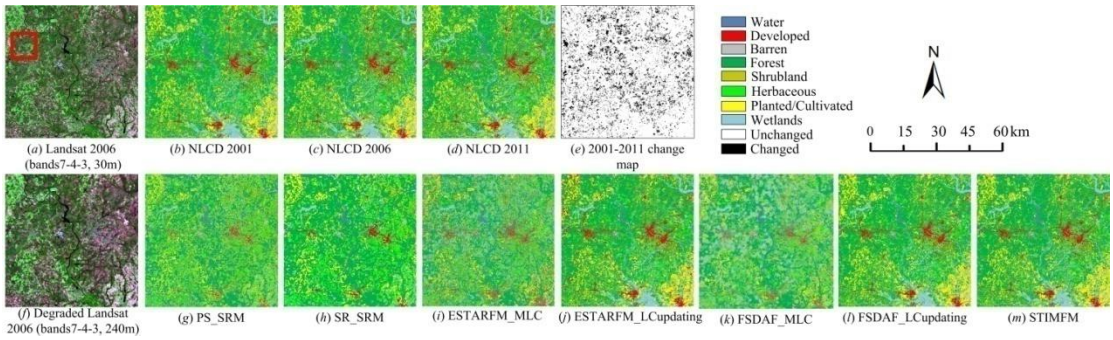


Fig. 3 Input and result maps for the entire study area in the Landsat–NLCD experiment.

A Landsat TM image (path 016, row 036) acquired on April 9, 2006 in the study area was downloaded from the United States Geological Survey (USGS). This Landsat image was re-projected to the Albers Equal Area projection, and six spectral bands at the spatial resolution of 30 m (the 120 m thermal infrared band was excluded) were used to extract the same 2000 × 2000 pixel area that was identified in the NLCD maps [Fig. 3(*a*)]. The subset image was calibrated to surface reflectance (Gao et al. 2006; Masek et al. 2006) and then spatially degraded to simulate a coarse spatial resolution multi-spectral image using a scale factor *s*=8 [Fig. 3(*f*), 240 m] with a mean filter. The NLCD 2006 [Fig. 3(*c*)] was used as the reference map used for accuracy assessment. The pixels that changed land cover class from 2001 to 2011 accounted for 12.08% of all fine spatial resolution pixels.

For analyses with the PS_SRM and SR_SRM, only the degraded multi-spectral image [Fig. 3(*f*)] was needed as input. For the STIMFM, the required input included the degraded multi-spectral image [Fig. 3(*f*)] and the NLCD 2001 and NLCD 2011 land cover maps [Fig. 3(*b*), 3(*d*)]. For ESTARFM, pairs of fine and coarse spatial resolution multi-spectral images that temporally pre- and post-date the 2006 coarse-resolution

remotely sensed image were needed. To obtain the required data, a Landsat TM image acquired on April 17, 2001 and a Landsat TM image acquired on April 7, 2011 were also downloaded, re-projected, subsetted, and calibrated. The original 30 m spatial resolution reflectance images with six spectral bands (the 120 m thermal infrared band was excluded) were spatially degraded to simulate their corresponding coarse spatial resolution multi-spectral images at scale factors *s*=8, respectively. Therefore, the input to the ESTARFM_MLC and ESTARFM_LCupdating included fine and coarse spatial resolution multi-spectral image pairs in 2001 and 2011 and the coarse spatial resolution multi-spectral image for 2006. The input to FSDAF_MLC and FSDAF_LCupdating included fine and coarse spatial resolution multi-spectral image pairs in 2001 and the coarse spatial resolution multi-spectral image in 2006. In ESTARFM_LCupdating and FSDAF_LCupdating, the NLCD 2001 fine-resolution land cover map was also used as the base data.

3.1.2 Results

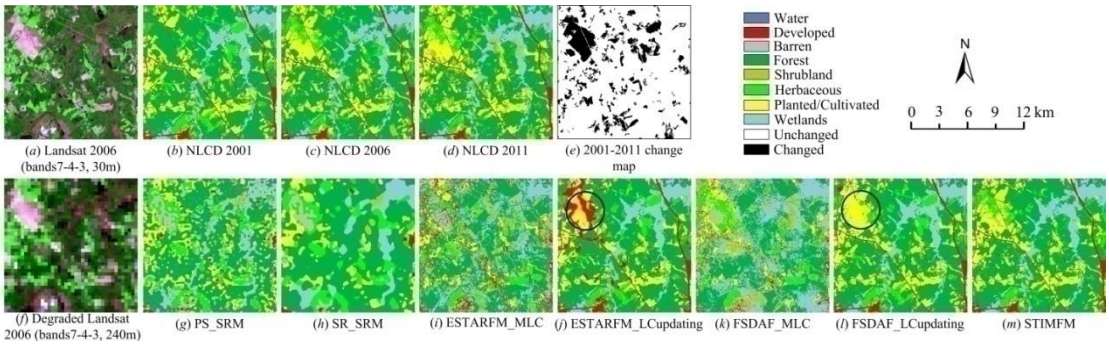
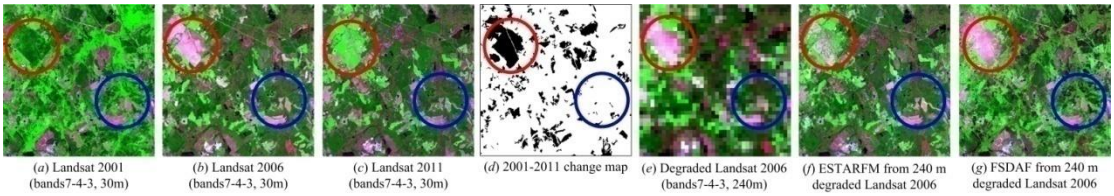


Fig. 4 Input and result maps for the zoomed area in the Landsat–NLCD experiment.

526 The land cover maps produced from the different methods are shown in Fig. 3 for
527 the entire area and in Fig. 4 for the zoomed area [320 × 320 pixel area in Fig. 3(*a*)]. In
528 the zoomed area, the PS_SRM map contained many speckle-like artifacts [Fig. 4(*g*)].
529 This situation arises because the spectral unmixing may determine a small fractional
530 cover of a class that is actually absent in a coarse-resolution pixel, and this fraction
531 must be maintained in the result. The SR_SRM map contained fewer speckle-like
532 artifacts than PS_SRM, because SR_SRM relaxed the constraint of land cover fraction
533 maintenance [Fig. 4(*h*)]. However, the maximal spatial dependence model used in
534 SR_SRM also led to rounded land cover patches. Compared with PS_SRM and
535 SR_SRM, more spatial detail of the land cover mosaic was retained in the ESTARFM,
536 FSDAF, and STIMFM maps. Many speckle-like artifacts in the ESTARFM_MLC [Fig.
537 4(*i*)] and FSDAF_MLC [Fig. 4(*k*)] maps existed because MLC is a per-based
538 classification method, and the spatial context information was not used.
539 ESTARFM_MLC and FSDAF_MLC incorrectly classified cases that have similar
540 reflectance values, such as "forest", "herbaceous", and "wetlands", in the result maps
541 [Figs. 4(*i*), (*k*)].



542

543 **Fig. 5** Landsat, ESTARFM and FSDAF images in the zoomed area for the Landsat–NLCD experiment.

544 In contrast to ESTARFM_MLC and FSDAF_MLC, ESTARFM_LCupdating [Fig.
545 4(*j*)] and FSDAF_LCupdating [Fig. 4(*l*)] quantified the land cover changes between
546 2001 and 2006 and generated land cover maps that were more similar to the reference

map [Fig. 4(*c*)]. The labels of pixels that were detected as unchanged by LCupdating were preserved in the ESTARFM_LCupdating and FSDAF_LCupdating maps. The labels of changed pixels were determined based on the Markov random field based classifier, which considers the contextual information in classification. Thus, most speckle-like artifacts were eliminated in ESTARFM_LCupdating and FSDAF_LCupdating. However, many changed pixel labels were incorrectly predicted by ESTARFM_LCupdating and FSDAF_LCupdating. "Herbaceous" was incorrectly labeled as "developed" in the ESTARFM_LCupdating highlighted by the black circle [Fig. 4(*j*)], and the linear-shaped "developed" in the FSDAF_LCupdating highlighted by the black circle was disconnected [Fig. 4(*l*)]. The predicted reflectance for pixels of changed land cover for "herbaceous" and "planted/cultivated" in ESTARFM [e.g., those highlighted by the red circle in Fig. 5(*f*)] was dissimilar to that in the Landsat 2006 reference image [Fig. 5(*b*)] because ESTARFM cannot capture abrupt land cover changes (Zhu et al. 2010), and the predicted reflectance of linear-shaped "developed" land cover was similar to that of "planted/cultivated" in the FSDAF image highlighted by the red circle in Fig. 5(*g*), because FSDAF cannot capture tiny land cover changes (Zhu et al. 2016). By contrast, the STIMFM land cover map as shown in Fig. 4(*m*) was quite similar to the reference map, and the detailed land cover patterns were well represented. STIMFM correctly predicted the class labels not only for almost all pixels of unchanged land cover but also for most of those pixels for which land cover class had changed, such as those highlighted in the red circle in Fig. 5(*d*).

569 **Table 1**

570 Overall accuracies (OAs) and accuracies of different methods in predicting PULC and PCLC in the

571 Landsat–NLCD experiment.

	OA	PULC	PCLC
PS_SRM	41.61	41.54	42.13
SR_SRM	49.10	49.01	49.77
ESTARFM_MLC	33.57	35.10	22.44
ESTARFM_LCupdating	88.28	94.89	40.18
FSDAF_MLC	33.60	34.29	28.60
FSDAF_LCupdating	89.50	96.15	41.08
STIMFM	94.89	99.24	63.27

572

573 The overall accuracies of different methods are shown in Table 1. The result maps
574 were compared with the NLCD 2006. The overall accuracy of STIMFM is higher than
575 those obtained from the other methods. Table 1 also shows the accuracies of pixels of
576 changed and unchanged land cover (PULC means the percentage of correctly labeled
577 pixels of unchanged land cover among all pixels of unchanged land cover, and PCLC
578 means the percentage of correctly labeled pixels of changed land cover among all pixels
579 of changed land cover) obtained from the different methods. For PS_SRM and
580 SR_SRM, which applied a mono-temporal remotely sensed image, no obvious
581 difference was found between PULC and PCLC values. For ESTARFM, FSDAF, and
582 STIMFM applied to multi-temporal data, the PULC values were higher than the PCLC
583 values. These results indicate that extracting changed land cover information is more
584 difficult than extracting unchanged land cover information from ESTARFM, FSDAF,
585 and STIMFM. STIMFM integrates the temporal dependence model in its objective
586 function, and the fine spatial resolution pixel class labels are temporally dependent on

587 those in the pre- and post-dated fine-resolution land cover maps. If the fine-resolution
588 pixel class labels are unchanged during the observation period, then STIMFM could
589 make the best use of pixel class labels in the fine-resolution maps that pre- and post-
590 date the coarse-resolution image. Thus, the accuracies for classes with unchanged class
591 labels are high. By contrast, if the fine-resolution pixel class labels have changed during
592 the observation period, then STIMFM could not make the best use of pixel class labels
593 in the fine-resolution maps that pre- and post-date the coarse-resolution image, and the
594 accuracies for classes with changed class labels are relatively low. The PULC was
595 higher than 99%, and the PCLC was higher than 63% for STIMFM; these values are
596 higher than those obtained from the other methods.
597

598 *3.2 MODIS–Landsat experiment*

599 *3.2.1 Data preparation*

600 The study area was located near Sorriso (12°33′00″S and 55°42′00″W) in Mato
601 Grosso State, Brazil. This area was mainly covered by tropical forests but has suffered
602 from deforestation in recent years (Hansen et al. 2008). This experiment used eleven
603 coarse spatial resolution MODIS images and two fine spatial resolution land cover
604 maps that pre- and post-date the coarse spatial resolution image series as input and
605 outputs eleven fine-resolution land cover maps with MODIS repetition rates to show
606 the fine spatial and temporal deforestation process in the study area. Landsat Enhanced
607 Thematic Mapper Plus (ETM+) images (path 226, row 069) acquired on 2002/06/08
608 and 2002/09/12 were downloaded from USGS [Fig. 6(*d*) and (*f*)]. Data in six bands (the

120 m thermal infrared band was excluded) at the 30 m spatial resolution with the Universal Transverse Mercator projection were used and calibrated to surface reflectance values (Gao et al. 2006; Masek et al. 2006). One cloud-free Landsat ETM+ image acquired on 2002/07/10 was used for accuracy assessment [Fig. 6(*e*)]. A total of thirteen eight-day surface reflectance MODIS product (MOD09A1) datasets that comprise seven spectral bands (620 nm–2055 nm) with a spatial resolution of 463 m acquired from 2002/06/02 to 2002/09/13 were downloaded from USGS (Walker et al. 2012). The MODIS images were re-projected into the UTM coordinate system and resampled to a spatial resolution of 450 m using the nearest neighbor interpolation, and were adopted as the coarse spatial resolution multi-spectral images required for the analyses. The study area covers 300 × 300 MODIS pixels, which correspond to 4500 × 4500 Landsat pixels, with a scale factor s=15.

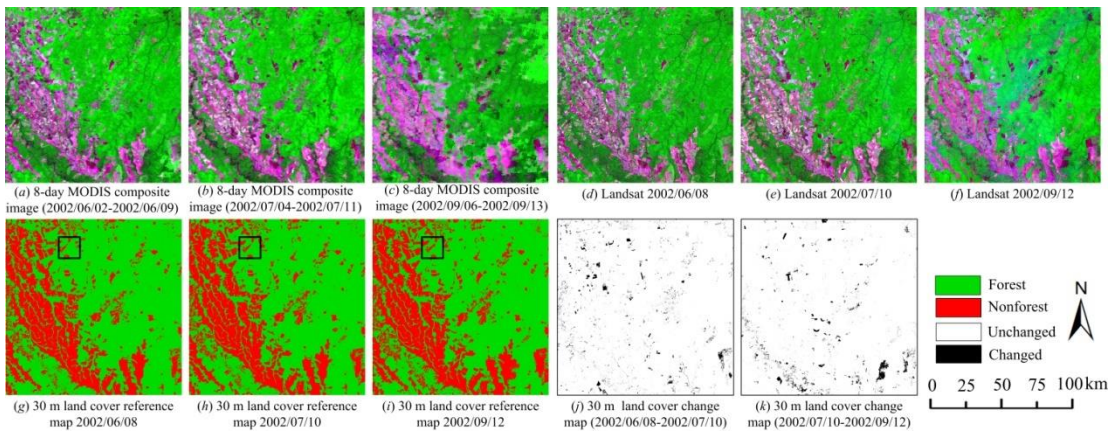


Fig. 6 MODIS, Landsat images, and reference maps in the MODIS–Landsat experiment from 2002/06/08 to 2002/09/12.

The three Landsat images were classified to produce land cover maps with a 30 m

627 spatial resolution [Fig. 6(*g*)–(*i*)]. Two land cover classes, forest and nonforest, were
628 considered in this experiment. The endmembers of each class were manually selected
629 from each Landsat image, and MLC was applied to generate the fine spatial resolution
630 forest/nonforest reference maps. The fine-resolution change maps that produced by a
631 per-pixel comparison of maps in Fig. 6(*g*)–(*i*) are shown in Fig. 6(*j*)–(*k*). The pixels that
632 changed land cover class from 2002/06/08 to 2002/09/12 accounted for 4.30% of all
633 fine spatial resolution pixels.

634 STIMFM used the MODIS multi-spectral image series from 2002/06/10 to
635 2002/09/05 and the 2002/06/08 and 2002/09/12 fine spatial resolution land cover maps
636 in Fig. 6 (*g*) and (*i*) as input and predicted a series of land cover maps at 30 m spatial
637 resolution with MODIS repetition rates during this period. The accuracy was assessed
638 using the 2002/07/10 land cover map [Fig. 6(*h*)]. The STIMFM was compared with
639 PS_SRM, SR_SRM, ESTARFM_MLC, ESTARFM_LCupdating, FSDAF_MLC, and
640 FSDAF_LCupdating using the 2002/07/10 land cover map in Fig. 6(*h*) for assessment.
641 In these methods, the eight-day composite MODIS image [2002/07/04–2002/07/11 in
642 Fig. 6(*b*)] was used as the coarse-resolution image. Aside from this data, ESTARFM
643 used the eight-day composite MODIS images [2002/06/02–2002/06/09 in Fig. 6(*a*) and
644 2002/09/06–2002/09/13 in Fig. 6(*c*)] and Landsat multi-spectral images [2002/06/08 in
645 Fig. 6(*d*) and 2002/09/12 in Fig. 6(*f*)] as input, and FSDAF used the eight-day
646 composite MODIS image [2002/06/02–2002/06/09 in Fig. 6(*a*)] and Landsat multi-
647 spectral image [2002/06/08 in Fig. 6(*d*)] as input. In ESTARFM and FSDAF, the
648 MODIS bands 1–4 and 6–7 were used in ESTARFM and FSDAF, because no similar

649 spectral band of Landsat image was observed from MODIS band 5. The 2002/06/08
650 fine spatial resolution land cover map in Fig. 6(*g*) was also inputted in the
651 ESTARFM_LCupdating and FSDAF_LCupdating.

652 *3.2.2 Results*

653

654

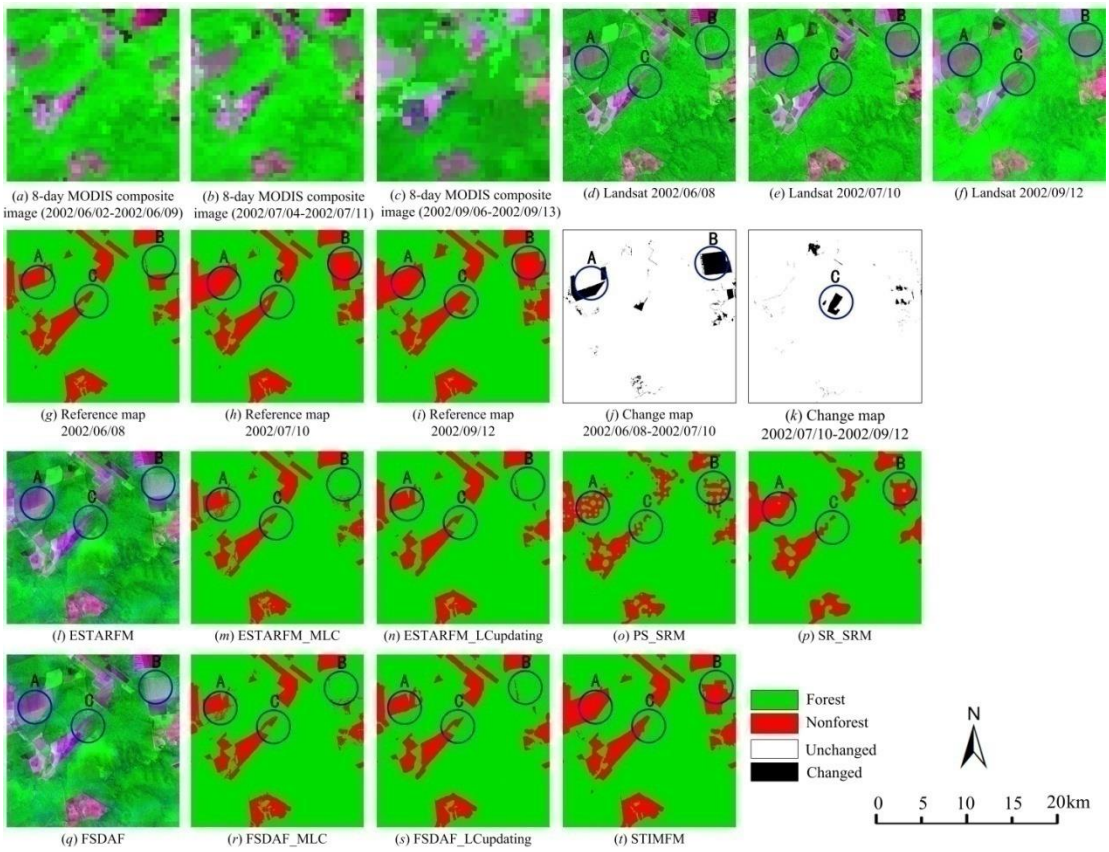
655 **Table 2**

656 Overall accuracies (OAs) and accuracies of different methods in predicting PULC and PCLC in the
657 MODIS–Landsat experiment. The MODIS image used in different methods was the eight-day composite
658 data from 2002/07/04 to 2002/07/11.

	OA	PULC	PCLC
PS_SRM	88.14	89.29	62.52
SR_SRM	89.28	90.42	63.73
ESTARFM_MLC	95.17	96.90	56.71
ESTARFM_LCupdating	96.32	98.07	57.42
FSDAF_MLC	95.07	96.72	58.23
FSDAF_LCupdating	96.61	98.37	57.38
STIMFM	98.27	99.69	66.67

659

660 The OA, PULC, and PCLC values obtained from the application of the different
661 methods are shown in Table 2. The overall accuracies obtained from the PS_SRM and
662 SR_SRM were lower than 90%, whereas the overall accuracies of ESTARFM_MLC,
663 ESTARFM_LCupdating, FSDAF_MLC, and FSDAF_LCupdating were higher than
664 95%. These findings indicate that the classification from fine-resolution image
665 extracted by spatial–temporal fusing of coarse and fine-resolution images can better
666 improve the accuracy compared with SRM applied to a mono-temporal coarse-
667 resolution image. The OA value for STIMFM was 98.27%, which is higher than all the
668 other methods. The PCLC values were lower than the PULC values for ESTARFM,
669 FSDAF, and STIMFM methods, which is similar to those in the Landsat–NLCD
670 experiment. The STIMFM has the highest PULC value, which was 99.69%, and the
671 highest PCLC value, which was 66.67%, among all the methods.



672

673 **Fig. 7** Input, reference, and result images and maps for the zoomed area at different years for the MODIS–

674 Landsat experiment. The MODIS image used in different methods was the eight-day composite data

675 from 2002/07/04 to 2002/07/11.

676

677 The reference, input, and result images and maps in the zoomed area are shown in

678 Fig. 7. A part of the forest patch (highlighted by a blue circle in regions A and B in Fig.

679 7) changed to nonforest from 2002/06/08 to 2012/07/10 [Fig. 7(*j*)], and a part of the

680 forest patch (highlighted by a blue circle in region C in Fig. 7) changed to nonforest

681 from 2002/07/10 to 2012/09/12 [Fig. 7(*k*)]. The PS_SRM map contained many speckle-

682 like artifacts [Fig. 7(*o*)], and SR_SRM contained land cover patches with oversmoothed

683 rounded boundaries [Fig. 7(*p*)]. In the ESTARFM and FSDAF fused images [Fig. 7(*l*),

684 (*q*)], the pixels of unchanged land cover considerably resemble those in the reference

685 Landsat image, whereas the pixels of changed land cover (highlighted by blue circles
686 in Fig. 7) were noticeably different from those in the reference Landsat image [Fig.
687 7(*e*)]. As a result, these pixels of changed land cover were erroneously classified in the
688 ESTARFM_MLC, ESTARFM_LCupdating, FSDAF_MLC, and FSDAF_LCupdating
689 results [Figs. 7(*m*), (*n*), (*r*), and (*s*)]. By contrast, most of the changed and unchanged
690 pixels are correctly allocated by STIMFM [Fig. 7(*t*)], thereby showing the ability of the
691 proposed STIMFM model in the reconstruction of land cover trajectories for pixels of
692 both changed and unchanged land cover. The land cover changes in Fig. 8 were
693 extracted by comparing the STIMFM predicted maps and input fine-resolution land
694 cover map that pre-dates the coarse images [Fig. 6(*g*)]. The colors in Fig. 8 indicate the
695 date when the pixels begin to change. The forest area decreased gradually, whereas the
696 nonforest area increased in Fig. 9. With STIMFM, the detailed spatial extent
697 information and the change of areas for different classes can be extracted, thereby
698 showing the effectiveness of the proposed method.
699

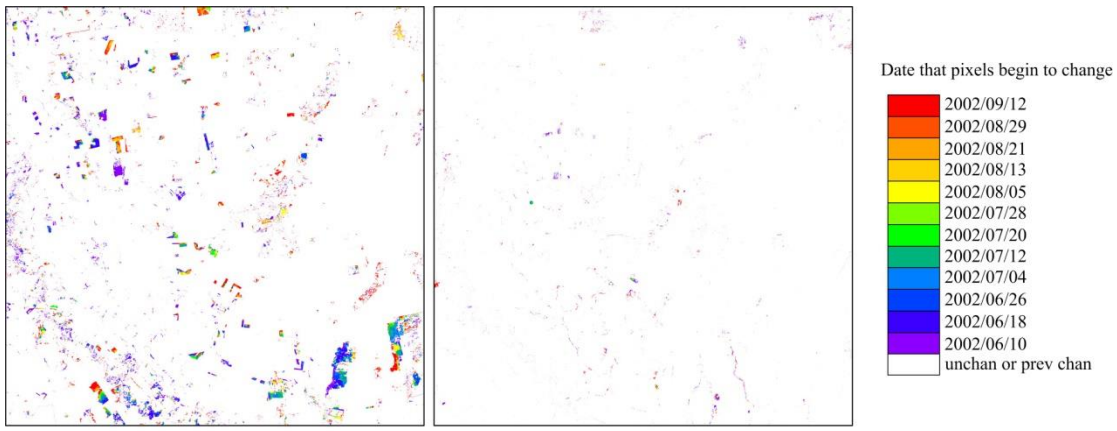


Fig. 8 30 m spatial extent of land cover change with MODIS repetition rates derived from STIMFM. The colors represent the date when pixels begin to change. "unchan or prev chan" marked as white color means unchanged or previously changed before 2002/06/08.

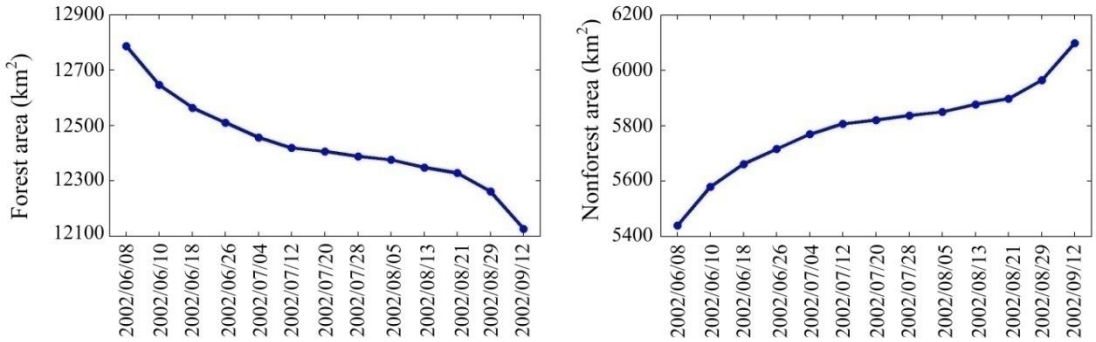
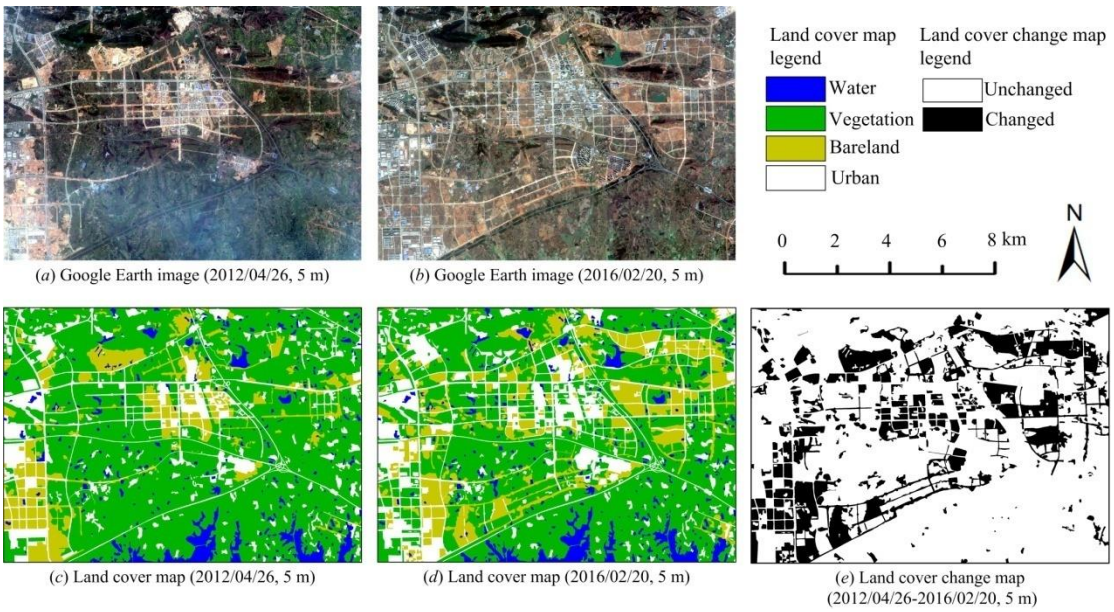


Fig. 9 Forest and nonforest areas extracted using STIMFM in the MODIS–Landsat experiment.

3.3 Landsat–GEI experiment

The study area was located in Wuhan (30°27′30″N and 114°32′30″E), Hubei province, China. This area underwent rapid urbanization in 2010–2016. This experiment used eleven cloud-free 30 m spatial resolution Landsat-8 Operational Land Imager (OLI) multi-spectral images (path 123, row 039) from 2013 to 2015 and two

5 m spatial resolution land cover maps acquired in 2012 and 2016 as input. Eleven 5 m resolution land cover maps during 2013–2015 were predicted to show the fine spatial and temporal urbanization process in the study area. The acquired eleven Landsat OLI images were downloaded from USGS. The first seven bands of OLI image with a spatial resolution of 30 m were selected. Two GEIs acquired on 2012/04/26 and on 2016/02/20 [Figs. 10(*a*), (*b*)] with a spatial resolution of 5 m were re-projected into the UTM coordinate system and digitized into the 5 m land cover maps [Figs. 10(*c*), (*d*)]. Four land cover classes, namely, water, vegetation, bareland, and urban, were found in the fine-resolution maps. The study area covers 320 × 450 Landsat pixels, which correspond to 1920 × 2700 fine-resolution pixels in Figs. 10(*c*) and (*d*), with a scale factor $s = 6$. The land cover change map from 2012/04/26 to 2016/02/20 is shown in Fig. 10(*e*). The pixels that changed land cover class accounted for 23.49% of all fine spatial resolution pixels from 2012/04/26 to 2016/02/20.



728

729 **Fig. 10** Google Earth images, land cover maps and change maps in the Landsat–GEI experiment.

730

731 STIMFM was used to produce the eleven 5 m resolution land cover maps with
732 Landsat repetition rates during 2013–2015 using the eleven cloud-free Landsat images
733 and two 5 m land cover maps on 2012/04/26 and 2016/02/20 as input. The STIMFM
734 accuracy was assessed using a 5 m fine-resolution land cover map, which was produced
735 according to a GEI at the spatial resolution of 5 m acquired on 2014/12/06 [Fig. 11(*b*)].
736 This GEI is the only fine-resolution one available in the study area during 2012–2016
737 and was re-projected into the UTM coordinate system and digitized to the reference
738 land cover map [Fig. 11(*c*)]. STIMFM was compared with PS_SRM and SR_SRM,
739 which were applied to a single-date Landsat OLI image acquired on 2014/10/06 [Fig.
740 11(*a*)]; this image is temporally closest to the GEI in 2014 [Fig. 11(*b*)]. ESTARFM and
741 FSDAF were not used for comparison because they require the coarse- and fine-
742 resolution images to have comparable and correlated reflectance bands, whereas
743 Landsat and GEI have different spectral bands and the GEI can hardly be transformed

744 into reflectance images, which are correlated to the Landsat images.

745

746 **Table 3**

747 Overall accuracies (OAs) and accuracies of different methods in predicting PULC and PCLC in the

748 Landsat–GEI experiment. The Landsat image used for assessment in the different methods was acquired

749 on 2014/10/06.

	OA	PULC	PCLC
PS_SRM	72.71	76.22	61.13
SR_SRM	73.73	77.29	61.99
STIMFM	94.31	99.61	76.81

750

751 The OA accuracies were lower than 74% for PS_SRM and SR_SRM and increased

752 to 94.31% for STIMFM (Table 3). The PULC value was higher than 99%, and the

753 PCLC value was higher than 76% for STIMFM; these values were obviously higher

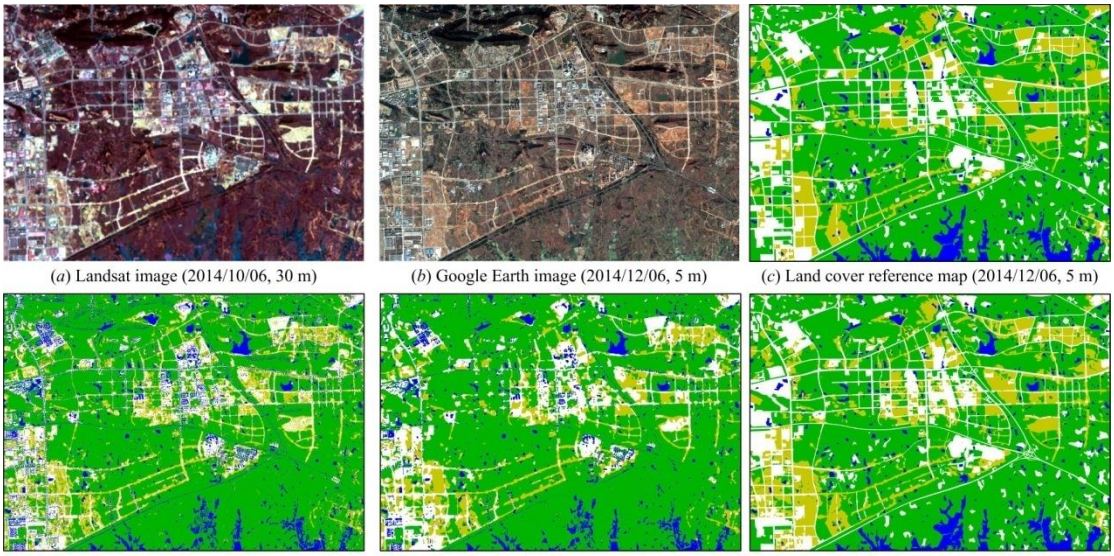
754 than those for PS_SRM and SR_SRM. The PULC values were higher than the PCLC

755 values for STIMFM because STIMFM could make the best use of unchanged pixel

756 labels in the fine-resolution maps that pre- and post-date the Landsat images in land

757 cover mapping.

758



759

760 **Fig. 11** Landsat image, Google Earth image, reference and result maps in the Landsat–GEI experiment.

761 The Landsat image used in the different methods was acquired on 2014/10/06.

762 The PS_SRM contained many speckle-like artifacts [Fig. 11(*d*)]. Many speckle-

763 like artifacts were smoothed to rounded patches in SR_SRM [Fig. 11(*e*)]. The linear-

764 shaped urban objects were discrete in PS_SRM and SR_SRM and connected in

765 STIMFM [Fig. 11(*f*)]. In STIMFM, most speckle-like artifacts and rounded patches

766 were eliminated, and the spatial pattern of most patches was close to the reference map

767 [Fig. 11(*c*)]. The 5 m spatial extent of land cover change with Landsat repetition rates

768 derived from STIMFM is shown in Fig. 12, in which the explicit time of land cover

769 change and the detailed spatial extent of urbanization process at fine spatial and fine

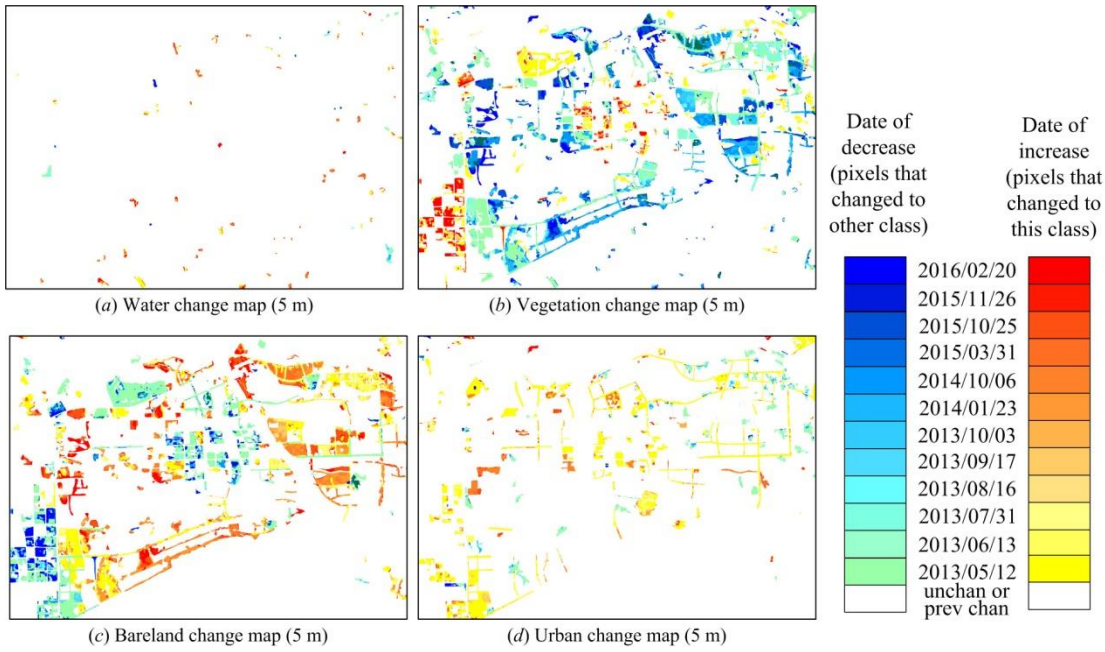
770 temporal resolutions are obvious. Fig. 13 shows the areas of different classes extracted

771 using STIMFM. The water, vegetation, and bareland areas decreased, whereas the

772 urban area increased from April 2012 to May 2013. The areas of different classes

773 remained almost unchanged from May to September 2013. Since October 2013, the

774 vegetation area decreased, whereas the water, bareland, and urban areas increased.



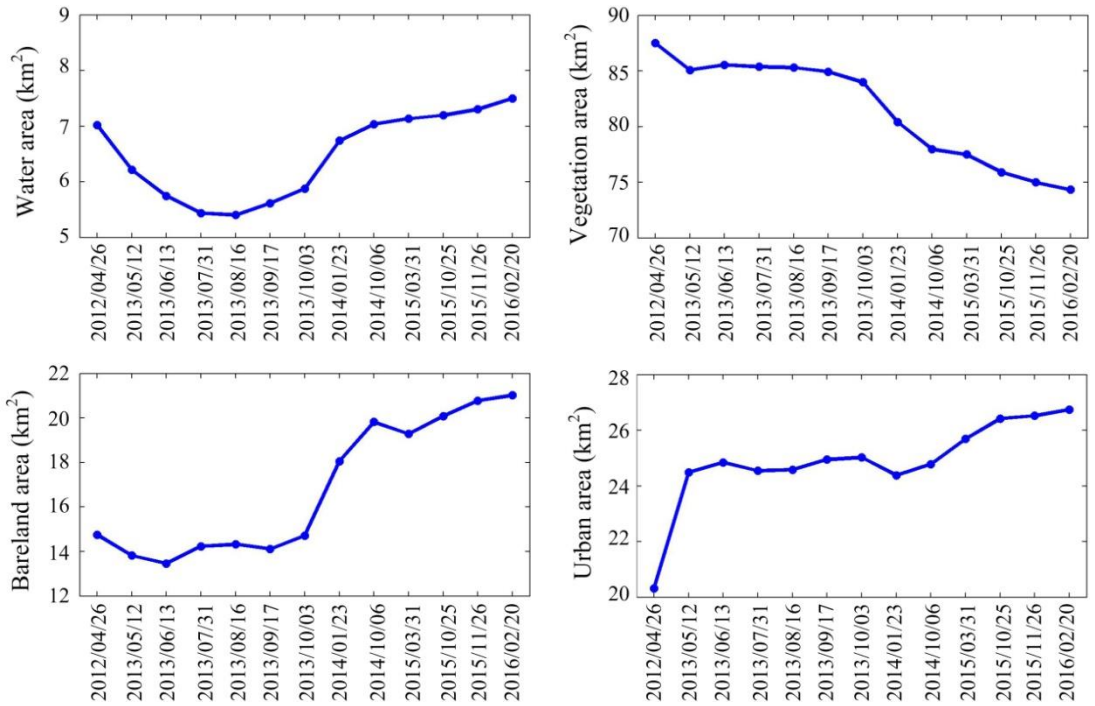
775

776 **Fig. 12** 5 m spatial extent of land cover change with Landsat repetition rates derived from STIMFM. The

777 colors represent the date when pixels begin to change. "unchan or prev chan" marked as white color

778 means unchanged or previously changed before 2012/04/26.

779



780

781 **Fig. 13** Areas of different classes extracted using STIMFM in the Landsat–GEI experiment.

4. Discussion

Results show that STIMFM is a promising approach for the production of a series of fine spatial–temporal resolution land cover maps, which were achieved by fusing a series of coarse spatial resolution remotely sensed images with a limited set of fine spatial resolution land cover maps. The mapping accuracies of STIMFM reached relatively high levels in all three experiments. Compared with popular state-of-the-art SRM algorithms that are generally applied on mono-temporal remotely sensed image, STIMFM can produce land cover maps of a much higher accuracy as expected, because fine spatial resolution land cover temporal information is incorporated into its analysis. Compared with ESTARFM and FSDAF, STIMFM predicted the labels of both changed and unchanged pixels with higher accuracy in the Landsat–NLCD and MODIS–Landsat experiments. In the Landsat–GEI experiment, STIMFM predicted a sequence of fine spatial–temporal resolution land cover maps from eleven Landsat images and two GEIs, to which ESTARFM and FSDAF that require correlation in reflectance bands in coarse and fine images cannot be applied.

Although ESTARFM, FSDAF, and STIMFM aim to extract high spatial–temporal resolution information, they have important differences that affect practical application. First, they have different assumptions and thus use different inputs. ESTARFM and FSDAF require coarse- and fine-resolution remotely sensed images from different satellite sensors observed at the same or similar date to have comparable and highly

803 correlated reflectance bands. Thus, only a limited set of images can be used in
804 ESTARFM and FSDAF, such as Landsat and MODIS, thereby limiting the application
805 of ESTARFM and FSDAF. For instance, panchromatic aerial photographs with a very
806 high spatial resolution cannot be used with Landsat image with a 30 m resolution to
807 generate fine-resolution land cover maps with Landsat repetition rates from ESTARFM
808 and FSDAF because the aerial photographs and Landsat images have different spectral
809 bands. By contrast, STIMFM does not require similar coarse and fine spatial resolution
810 images, but directly considers the relationship between the land cover classes
811 themselves and not their spectral response. The coarse spatial resolution images are
812 unmixed to land cover fractions, and STIMFM is built on the analysis of land cover
813 spatial and temporal dependences in the different images instead of analyzing the
814 relationship of pixel spectral values in different images. In addition, ESTARFM and
815 FSDAF require one or more observed pairs of coarse- and fine-resolution images
816 acquired at the same or similar date for training, whereas STIMFM does not need the
817 coarse-resolution images at the acquisition data of the fine-resolution maps as input.

818 Second, ESTARFM, FSDAF, and STIMFM have different outputs; the output of
819 ESTARFM and FSDAF are multi-spectral reflectance images, whereas the output of
820 STIMFM are land cover maps. If the aim is to generate spectral images, then
821 ESTARFM and FSDAF are suitable. For instance, unlike the STIMFM result, the
822 ESTARFM and FSDAF result can be used in the analysis of phenology change.
823 STIMFM produces land cover maps with discrete class labels and is more suitable in
824 monitoring the spatial distribution pattern and temporal change trajectory of land cover

825 classes at a fine spatial and temporal resolution. Although the reflectance images output
826 from ESTARFM and FSDAF can be further classified to produce land cover maps,
827 problems still exist. To generate fine spatial resolution land cover map series,
828 ESTARFM and FSDAF should generate a series of fine spatial resolution multi-spectral
829 reflectance images, thereby requiring massive storage for these intermediate data when
830 the study area is large. In addition, the generation of land cover maps from these image
831 series requires a large amount of training data, which are difficult to collect in practice.
832 The classification of reflectance images is also often underdetermined and contains a
833 large solution space. By contrast, STIMFM is modeled based on the spatial–temporal
834 character of pixel class labels. It does not produce intermediate fine spatial resolution
835 multi-spectral image series, and the endmembers could be automatically estimated for
836 each coarse spatial resolution image on the basis of optimization approach. The
837 STIMFM has a simple objective function and comprises only few parameters and is
838 thus relatively easy to use. As a result, STIMFM is more suitable in the reconstruction
839 of fine spatial and temporal resolution land cover maps compared with ESTARFM and
840 FSDAF.

841 Although STIMFM provides a great opportunity to enhance studies of land cover
842 and its dynamics, its performance is dependent on several factors. In STIMFM, the
843 analysis of land cover class fraction temporal change is conducted by comparing the
844 coarse-resolution fraction images produced from spatial degrading the input fine-
845 resolution maps and from spectral unmixing of the coarse-resolution image. First,
846 fraction images extracted from spectral unmixing probably have errors and

847 uncertainties, which will affect the land cover fraction temporal change analysis in
848 STIMFM. The linear mixture model is used to estimate the endmembers and generate
849 fraction images for class fraction change analysis in STIMFM. However, this approach
850 is not always ideal, because the mixing may be nonlinear. Some nonlinear mixture
851 models may be applied to decrease the fraction image error and improve the class
852 fraction temporal change accuracy. Second, the accuracy of the fine-resolution land
853 cover maps that pre- and post-dated the coarse-resolution images also affect the
854 STIMFM accuracy. The overall accuracy of STIMFM decreases with the increase in
855 the number of incorrect pixel labels in the input fine-resolution land cover maps because
856 STIMFM labeled the unchanged pixels according to the labels in the fine-resolution
857 maps. In addition, the incorrect pixel labels in the fine-resolution maps would decrease
858 the accuracy in the class fraction temporal change analysis and thus decrease the
859 STIMFM accuracy. Advanced classifiers such as object-based classifiers should be
860 used to extract accurate land cover maps from the fine-resolution images used as
861 STIMFM input. Third, the co-registration between the fine-resolution land cover maps
862 and the coarse-resolution images plays a key role because misregistration would lead
863 to inaccurate detection of fraction changes of each class in each coarse pixel. Advanced
864 methods such as the sub-pixel scale co-registration method should be developed and
865 applied in STIMFM.

866 The STIMFM performance is also affected by the model functions and parameters.
867 First, in the STIMFM spatial constraint function, the *a priori* land cover spatial
868 distribution model has a major role in the prediction of fine spatial resolution land cover

869 spatial pattern. The land cover maximal spatial dependence model is used as the *a priori*
870 land cover spatial pattern information in STIMFM for its simplicity. However, this *a*
871 *priori* information is used for all classes although they may actually have different
872 spatial patterns and is most suitable for the situation in which land cover patches are
873 larger than the coarse spatial resolution pixel size. More *a priori* information could be
874 introduced to characterize the spatial pattern of the classes in STIMFM. Second, the
875 STIMFM performance is dependent on the spatial and temporal weights λ_S and λ_T.
876 When the spatial weight λ_S is relatively large, STIMFM would decrease the influence
877 of temporal information, and the STIMFM result would be dominated by the spatial
878 constraint function and resemble the SR_SRM result. By contrast, when the temporal
879 weight λ_T is relatively large, the STIMFM result would be dominated by the fine
880 spatial resolution maps that temporally pre- and post-dated the coarse-resolution
881 images, and the spatial pattern of land cover patches would be difficult to reconstruct
882 in the result maps. The optimal λ_S value can be automatically estimated through
883 quantification of the effects of land cover class spectral separability (Li et al. 2016;
884 Tolpekin and Stein 2009), whereas the estimation of optimal λ_T value has not been
885 studied to our knowledge. In this paper, the optimal λ_S and λ_T values were
886 determined through many trials. In practice, a subset of coarse spatial resolution images
887 and fine spatial resolution maps are usually available, and these data can be used to
888 estimate the optimal λ_S and λ_T values in STIMFM. Finally, a selected number ($m \times$
889 C) of purest coarse-resolution pixels are used to estimate the endmembers **E** in **Y**. In
890 practice, m can be set in the range about 100–200 if **Y** is a multi-spectral image.

5. Conclusion

In this paper, a novel spatial–temporal remotely sensed images and land cover maps fusion model was proposed. This model aims to produce a series of fine spatial–temporal resolution land cover maps from a series of coarse spatial resolution remotely sensed images and a few fine spatial resolution land cover maps. In STIMFM, the endmember spectra of different land cover classes are estimated automatically for each coarse spatial resolution image with the aid of available fine spatial resolution land cover maps. Using the estimated endmember spectra, an objective function, which incorporates the pixel spectral and land cover spatial and temporal information, is constructed. The output of STIMFM is achieved by solving the optimization problem.

The performance of STIMFM was explored using three experiments and compared with that of several popular state-of-the-art algorithms. The STIMFM has comparable efficiency with ESTARFM and FSDAF in terms of computing time. STIMFM can produce land cover maps with higher accuracies than those algorithms used for comparison. The overall accuracies of STIMFM are higher than 94% in all experiments reported. Results indicate that STIMFM is a promising approach for generating land cover maps and estimating land cover change at both fine spatial and temporal resolutions. Although issues that would benefit from further research exist, this novel land cover fusion method provides a great opportunity to enhance studies of land cover and its dynamics.

Acknowledgements

This work was supported in part by the Youth Innovation Promotion Association CAS (Grant No. 2017384), in part by the Natural Science Foundation of China (Grant No. 61671425), in part by the Distinguished Young Scientist Grant of the Chinese Academy of Sciences, and in part by the State Key Laboratory of Resources and Environmental Informational System. The authors would like to thank Xiaolin Zhu for providing the ESTARFM and FSDAF programs and Xuehong Chen for providing the LCupdating program.

References

Amorós-López J., Gómez-Chova L., Alonso L., Guanter L., Zurita-Milla R., Moreno J., et al., Multitemporal fusion of Landsat/TM and ENVISAT/MERIS for crop monitoring, *Int. J. Appl. Earth Obs. Geoinf.* **23**, 2013, 132–141, http://dx.doi.org/10.1016/j.jag.2012.12.004.

Anderson M. C.M.C., Kustas W. P.W.P., Norman J. M.J.M., Hain C. R.C.R., Mecikalski J. R.J.R., Schultz L., et al., Mapping daily evapotranspiration at field to continental scales using geostationary and polar orbiting satellite imagery, *Hydrol. Earth Syst. Sci*. **15**, 2011, 223–239, http://dx.doi.org/10.5194/hess-15-223-2011.

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

Atkinson P. M.P.M., Issues of uncertainty in super-resolution mapping and their implications for the design of an inter-comparison study, *Int. J. Remote Sens.* **30**, 2009, 5293–5308, http://dx.doi.org/10.1080/01431160903131034.

Bartholomé E. and Belward A. S.A.S., GLC2000: a new approach to global land cover mapping from Earth observation data, *Int. J. Remote Sens.* **26**, 2005, 1959–1977, http://dx.doi.org/10.1080/01431160412331291297.

Cai S., Liu D., Sulla-Menashe D. and Friedl M. A.M.A., Enhancing MODIS land cover product with a spatial-temporal modeling algorithm, *Remote Sens. Environ.* **147**, 2014, 243–255, http://dx.doi.org/10.1016/j.rse.2014.03.012.

Chen X., Chen J., Shi Y. and Yamaguchi Y., An automated approach for updating land cover maps based on integrated change detection and classification methods, *ISPRS J. Photogramm. Remote Sens.* **71**, 2012, 86–95, http://dx.doi.org/10.1016/j.isprsjprs.2012.05.006.

943 Chen X., Yang D., Chen J. and Cao X., An improved automated land cover updating approach by
944 integrating with downscaled NDVI time series data, *Remote Sens. Lett.* **6**, 2015, 29–38,
945 http://dx.doi.org/10.1080/2150704X.2014.998793.
946 Emelyanova I. V.I.V., McVicar T. R.T.R., Van Niel T. G.T.G., Li L. T.L.T. and van Dijk A.,
947 Assessing the accuracy of blending Landsat-MODIS surface reflectances in two landscapes with
948 contrasting spatial and temporal dynamics: a framework for algorithm selection, *Remote Sens.*
949 *Environ.* **133**, 2013, 193–209, http://dx.doi.org/10.1016/j.rse.2013.02.007.
950 Foody G. M.G.M., The continuum of classification fuzziness in thematic mapping, *Photogramm.*
951 *Eng. Remote Sens.* **65**, 1999, 443–451.
952 Foody G. M.G.M., The role of soft classification techniques in the refinement of estimates of ground
953 control point location, *Photogramm. Eng. Remote Sens.* **68**, 2002, 897–903.
954 Foody G. M.G.M., Muslim A. M.A.M. and Atkinson P. M.P.M., Super-resolution mapping of the
955 waterline from remotely sensed data, *Int. J. Remote Sens.* **26**, 2005, 5381–5392,
956 http://dx.doi.org/10.1080/0143116050021329.
957 Friedl M. A.M.A., McIver D. K.D.K., Hodges J. C. F.J.C.F., Zhang X., Muchoney D., Strahler A.
958 H.A.H., et al., Global land cover mapping from MODIS: algorithms and early results, *Remote*
959 *Sens. Environ.* **83**, 2002, 287–302, http://dx.doi.org/10.1016/S0034-4257(02)00078-0.
960 Fu D., Chen B., Wang J., Zhu X. and Hilker T., An improved image fusion approach based on
961 enhanced spatial and temporal the adaptive reflectance fusion model, *Remote Sens.* **5**, 2013,
962 6346–6360, http://dx.doi.org/10.3390/rs5126346.
963 Gao F., Masek J., Schwaller M. and Hall F., On the blending of the Landsat and MODIS surface
964 reflectance: predicting daily Landsat surface reflectance, *IEEE Trans. Geosci. Remote Sens*. **44**,

2006, 2207–2218, http://dx.doi.org/10.1109/TGRS.2006.872081.

Ge Y., Chen Y., Stein A., Li S. and Hu J., Enhanced subpixel mapping with spatial distribution patterns of geographical objects, *IEEE Trans. Geosci. Remote Sens.* **54**, 2016, 2356–2370, http://dx.doi.org/10.1109/TGRS.2015.2499790.

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

Hansen M. C.M.C., Shimabukuro Y. E.Y.E., Potapov P. and Pittman K., Comparing annual MODIS and PRODES forest cover change data for advancing monitoring of Brazilian forest cover, *Remote Sens. Environ.* **112**, 2008, 3784–3793, http://dx.doi.org/10.1016/j.rse.2008.05.012.

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

Hilker T., Wulder M. A.M.A., Coops N. C.N.C., Seitz N., White J. C.J.C., Gao F., et al., Generation of dense time series synthetic Landsat data through data blending with MODIS using a spatial and temporal adaptive reflectance fusion model, *Remote Sens. Environ.* **113**, 2009b, 1988–1999, http://dx.doi.org/10.1016/j.rse.2009.05.011.

Homer C., Dewitz J., Yang L., Jin S., Danielson P., Xian G., et al., Completion of the 2011 National Land Cover Database for the conterminous United States - Rrepresenting a decade of land cover change information, *Photogramm. Eng. Remote Sens.* **81**, 2015, 345–354, http://dx.doi.org/10.14358/PERS.81.5.345.

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

Huang B. and Zhang H., Spatio-temporal reflectance fusion via unmixing: accounting for both phenological and land-cover changes, *Int. J. Remote Sens.* **35**, 2014, 6213–6233, http://dx.doi.org/10.1080/01431161.2014.951097.

Jia K., Liang S., Zhang N., Wei X., Gu X., Zhao X., et al., Land cover classification of finer resolution remote sensing data integrating temporal features from time series coarser resolution data, *ISPRS J. Photogramm. Remote Sens.* **93**, 2014, 49–55, http://dx.doi.org/10.1016/j.isprsjprs.2014.04.004.

Jin S., Yang L., Danielson P., Homer C., Fry J. and Xian G., A comprehensive change detection method for updating the National Land Cover Database to circa 2011, *Remote Sens. Environ.* **132**, 2013, 159–175, http://dx.doi.org/10.1016/j.rse.2013.01.012.

Kasetkasem T., Arora M. K.M.K. and Varshney P. K.P.K., Super-resolution land cover mapping using a Markov random field based approach, *Remote Sens. Environ.* **96**, 2005, 302–314, http://dx.doi.org/10.1016/j.rse.2005.02.006.

Li X., Du Y. and Ling F., Super-resolution mapping of forests with bitemporal different spatial resolution images based on the spatial-temporal Markov random field, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **7**, 2014, 29–39, http://dx.doi.org/10.1109/JSTARS.2013.2264828.

Li L., Chen Y., Xu T., Liu R., Shi K. and Huang C., Super-resolution mapping of wetland inundation from remote sensing imagery based on integration of back-propagation neural network and genetic algorithm, *Remote Sens. Environ.* **164**, 2015, 142–154, http://dx.doi.org/10.1016/j.rse.2015.04.009.

1009 Li X., Du Y. and Ling F., Sub-pixel-scale land cover map updating by integrating change detection
1010 and sub-pixel mapping, *Photogramm. Eng. Remote Sens.* **81**, 2015, 59–67,
1011 http://dx.doi.org/10.14358/PERS.81.1.59.
1012 Li X., Ling F., Foody G. M.G.M. and Du Y., A superresolution land-cover change detection method
1013 using remotely sensed images with different spatial resolutions, *IEEE Trans. Geosci. Remote*
1014 *Sens.* **54**, 2016, 3822–3841 http://dx.doi.org/10.1109/TGRS.2016.2528583.
1015 Ling F., Li W., Du Y. and Li X., Land cover change mapping at the subpixel scale with different
1016 spatial-resolution remotely sensed imagery, *IEEE Geosci. Remote Sens. Lett*. **8**, 2011, 182–186,
1017 http://dx.doi.org/10.1109/LGRS.2010.2055034.
1018 Ling F., Li X., Xiao F. and Du Y., Superresolution land cover mapping using spatial regularization,
1019 *IEEE Trans. Geosci. Remote Sens.* **52**, 2014, 4424–4439,
1020 http://dx.doi.org/10.1109/TGRS.2013.2281992.
1021 Ling F., Foody G. M.G.M., Zhang Y., Li X., Zhang X., Fang S., et al., Learning-based
1022 superresolution land cover mapping, *IEEE Trans. Geosci. Remote Sens*. **54**, 2016, 3794–3810,
1023 http://dx.doi.org/10.1109/TGRS.2016.2527841.
1024 Liu D. and Cai S., A spatial-temporal modeling approach to reconstructing land-cover change
1025 trajectories from multi-temporal satellite imagery, *Ann. Assoc. Am. Geogr.* **102**, 2012, 1329–
1026 1347, http://dx.doi.org/10.1080/00045608.2011.596357.
1027 Liu D., Kelly M. and Gong P., A spatial-temporal approach to monitoring forest disease spread using
1028 multi-temporal high spatial resolution imagery, *Remote Sens. Environ.* **101**, 2006, 167–180,
1029 http://dx.doi.org/10.1016/j.rse.2005.12.012.
1030 Lu D., Mausel P., Brondizio E. and Moran E., Change detection techniques, *Int. J. Remote Sens.* **25**,

1031 2004, 2365–2407, http://dx.doi.org/10.1080/0143116031000139863.

1032 Lu D., Batistella M., Moran E., Hetrick S., Alves D. and Brondizio E., Fractional forest cover
1033 mapping in the Brazilian Amazon with a combination of MODIS and TM images, *Int. J. Remote*
1034 *Sens.* **32**, 2011, 7131–7149, http://dx.doi.org/10.1080/01431161.2010.519004.

1035 Masek J. G.J.G., Vermote E. F.E.F., Saleous N. E.N.E., Wolfe R., Hall F. G.F.G., Huemmrich K.
1036 F.K.F., et al., A Landsat surface reflectance dataset for North America, 1990–2000, *IEEE Geosci.*
1037 *Remote Sens. Lett.* **3**, 2006, 68–72, http://dx.doi.org/10.1109/LGRS.2005.857030.

1038 van der Meer F. D.F.D. and Jia X., Collinearity and orthogonality of endmembers in linear spectral
1039 unmixing, *Int. J. Appl. Earth Obs. Geoinf.* **18**, 2012, 491–503,
1040 http://dx.doi.org/10.1016/j.jag.2011.10.004.

1041 Richards J. A.J.A. and Jia X., Remote Sensing Digital Imaging Analysis: An Introduction, third ed.,
1042 1999, Springer; Berlin.

1043 Settle J. J.J.J. and Drake N. A.N.A., Linear mixing and the estimation of ground cover proportions,
1044 *Int. J. Remote Sens.* **14**, 1993, 1159–1177, http://dx.doi.org/10.1080/01431169308904402.

1045 Song H. and Huang B., Spatiotemporal satellite image fusion through one-pair image learning, *IEEE*
1046 *Trans. Geosci. Remote Sens.* **51**, 2013, 1883–1896
1047 http://dx.doi.org/10.1109/TGRS.2012.2213095.

1048 Tatem A. J.A.J., Lewis H. G.H.G., Atkinson P. M.P.M. and Nixon M. S.M.S., Increasing the spatial
1049 resolution of agricultural land cover maps using a Hopfield neural network, *Int. J. Geogr. Inf.*
1050 *Sci*. **17**, 2003, 647–672, http://dx.doi.org/10.1080/1365881031000135519.

1051 Tolpekin V. A.V.A. and Stein A., Quantification of the effects of land-cover-class spectral
1052 separability on the accuracy of Markov-random-field-based superresolution mapping, *IEEE*

Trans. Geosci. Remote Sens. **47**, 2009, 3283–3297, http://dx.doi.org/10.1109/TGRS.2009.2019126.

Walker J. J.J.J., de Beurs K. M.K.M., Wynne R. H.R.H. and Gao F., Evaluation of Landsat and MODIS data fusion products for analysis of dryland forest phenology, *Remote Sens. Environ.* **117**, 2012, 381–393, http://dx.doi.org/10.1016/j.rse.2011.10.014.

Wang Q., Shi W., Atkinson P. M.P.M. and Li Z., Land cover change detection at subpixel resolution with a Hopfield neural network, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **8**, 2015, 1339–1352, http://dx.doi.org/10.1109/JSTARS.2014.2355832.

Wu M., Niu Z., Wang C., Wu C. and Wang L., Use of MODIS and Landsat time series data to generate high-resolution temporal synthetic Landsat data using a spatial and temporal reflectance fusion model, *J. Appl. Remote. Sens.* 2012, 6, http://dx.doi.org/10.1117/1.JRS.6.063507.

Xian G., Homer C. and Fry J., Updating the 2001 National Land Cover Database land cover classification to 2006 by using Landsat imagery change detection methods, *Remote Sens. Environ.* **113**, 2009, 1133–1147, http://dx.doi.org/10.1016/j.rse.2009.02.004.

Xu Y. and Huang B., A spatio–temporal pixel-swapping algorithm for subpixel land cover mapping, *IEEE Geosci. Remote Sens. Lett.* **11**, 2014, 474–478, http://dx.doi.org/10.1109/LGRS.2013.2268153.

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

Zhu X., Helmer E. H.E.H., Gao F., Liu D., Chen J. and Lefsky M. A.M.A., A flexible spatiotemporal

1075 method for fusing satellite images with different resolutions, *Remote Sens. Environ.* **172**, 2016,
1076 165–177, http://dx.doi.org/10.1016/j.rse.2015.11.016.
1077 Zhukov B., Oertel D., Lanzl F. and Reinhackel G., Unmixing-based multisensor multiresolution
1078 image fusion, *IEEE Trans. Geosci. Remote Sens.* **37**, 1999, 1212–1226,
1079 http://dx.doi.org/10.1109/36.763276.
1080 Zurita-Milla R., Kaiser G., Clevers J., Schneider W. and Schaepman M. E.M.E., Downscaling time
1081 series of MERIS full resolution data to monitor vegetation seasonal dynamics, *Remote Sens.*
1082 *Environ.* **113**, 2009, 1874–1885, http://dx.doi.org/10.1016/j.rse.2009.04.011.

LIST OF FIGURE CAPTIONS

Fig. 1 Production of a series of fine spatial and temporal resolution land cover maps from a series of coarse spatial resolution remotely sensed images and a few fine spatial resolution land cover maps in STIMFM.

Fig. 2 Flowchart of STIMFM.

Fig. 3 Input and result maps for the entire study area in the Landsat–NLCD experiment.

Fig. 4 Input and result maps for the zoomed area in the Landsat–NLCD experiment.

Fig. 5 Landsat, ESTARFM and FSDAF images in the zoomed area for the Landsat–NLCD experiment.

Fig. 6 MODIS, Landsat images, and reference maps in the MODIS–Landsat experiment from 2002/06/08 to 2002/09/12.

Fig. 7 Input, reference, and result images and maps for the zoomed area at different years for the MODIS–Landsat experiment. The MODIS image used in different methods was the eight-day composite data from 2002/07/04 to 2002/07/11.

Fig. 8 30 m spatial extent of land cover change with MODIS repetition rates derived from STIMFM. The colors represent the date when pixels begin to change. "unchan or prev chan" marked as white color means unchanged or previously changed before 2002/06/08.

Fig. 9 Forest and nonforest areas extracted using STIMFM in the MODIS–Landsat experiment.

Fig. 10 Google Earth images, land cover maps and change maps in the Landsat–GEI experiment.

Fig. 11 Landsat image, Google Earth image, reference and result maps in the Landsat–GEI experiment. The Landsat image used in the different methods was acquired on 2014/10/06.

Fig. 12 5 m spatial extent of land cover change with Landsat repetition rates derived from STIMFM. The colors represent the date when pixels begin to change. "unchan or prev chan" marked as white color

1106 means unchanged or previously changed before 2012/04/26.

1107 **Fig. 13** Areas of different classes extracted using STIMFM in the Landsat–GEI experiment.

1108