Updating Landsat-based Forest Cover Maps with MODIS Images Using Multiscale Spectral-Spatial-Temporal Superresolution Mapping

Yihang Zhang [a], Xiaodong Li [a], Feng Ling [a, *], Peter M. Atkinson [b], Yong Ge [c], Lingfei Shi [a, d], 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. Lancaster Environment Centre, Faculty of Science and Technology, Lancaster University, Lancaster LA1 4YQ, UK;

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

d. University of Chinese Academy of Sciences, Beijing 100049, China;

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

Abstract: With the high deforestation rates of global forest covers during the past decades, there is an ever-increasing need to monitor forest covers at both fine spatial and temporal resolutions. Moderate Resolution Imaging Spectroradiometer (MODIS) and Landsat series images have been used commonly for satellite-derived forest cover mapping. However, the spatial resolution of MODIS images and the temporal resolution of Landsat images are too coarse to observe forest cover at both fine spatial and temporal resolutions. In this paper, a novel multiscale spectral-spatial-temporal superresolution mapping (MSSTSRM) approach is proposed to update Landsat-based forest maps by integrating current MODIS images with the previous forest maps generated from Landsat image. Both the 240 m MODIS bands and 480 m MODIS bands were used as inputs of the spectral energy function of the MSSTSRM model. The principle of maximal spatial dependence was used as the spatial energy function to make the updated forest map spatially smooth. The temporal energy function was based on a multiscale spatial-temporal dependence model, and considers the land cover changes between the previous and current time. The novel MSSTSRM model was able to update Landsat-based forest maps more accurately, in terms of both visual and quantitative evaluation, than traditional pixel-based classification and the latest sub-pixel based super-resolution mapping methods. The results demonstrate the great efficiency and potential of MSSTSRM for updating fine temporal resolution Landsat-based forest maps using MODIS images.

Keywords: Forest cover mapping, MODIS, Landsat, Updating, Spectral-spatial-temporal, Super-resolution mapping.

1. Introduction

Forest cover across the world is crucial for the delivery of certain ecosystem services, such as carbon storage, water supplies, biodiversity richness, and climate regulation (Foley et al., 2005). With increasing human disturbance and extreme climate on the world's forest resources, global forest cover is undergoing changing rapidly, especially in the tropical domain, such as Amazon Basin (Malhi et al., 2008), because of the agricultural expansion, rapid urbanization, and fuel-wood collection (Canadell and Raupach, 2008; Foley et al., 2005). This change of forest cover have a significant impact on the climate change, biodiversity loss and carbon cycle (Canadell and Raupach, 2008). As such, to provide greater understanding and improved management of global forest resources, there is an ever-increasing need to update forest cover information at both fine spatial and fine temporal resolutions.

Field survey and photointerpretation techniques are traditional way used to monitor forest resources, but the critical constrain for them is the requirement of intensive human involvement (Gong et al., 1994). Fortunately, remote sensing technique could serve conveniently as a means to provide important satellite images for forest resources monitoring. At present, the most popular used remotely sensed images for forest cover mapping are perhaps the Landsat series and the MODIS images (Friedl et al., 2002; Jin et al., 2013; Xian et al., 2009). The former has a spatial resolution of 30 m, and in most cases, the spatial resolution of Landsat images make them suitable to observe the detailed spatial patterns of forest cover change. As a result, Landsat images have been used to produce various global products related to forest (or tree) cover, such as Hansen's global high-resolution (30 m spatial resolution) forest cover change map of the 21st Century (Hansen et al., 2013), Kim's global Landsat-based forest cover change from 1990 to 2000 (Kim et al., 2014), and Sexton's global 30 m spatial resolution continuous fields of tree cover (Sexton et al., 2013). Generally, single Landsat

satellite has a swath width of nearly 180 km and a minimum repeat frequency of 16 days. For most times, two Landsat satellites, such as Landsat-5 and Landsat-7 during 1999-2011 or Landsat-7 and Landsat-8 during 2013-present, are always available simultaneously, and the combined minimum repeat frequency of Landsat images could be shorten ideally to 8 days. It is, therefore, suggested to apply all available images provided by different Landsat satellites to have a timely monitoring of the forest cover change and forest disturbances (Hansen et al., 2016; Zhu et al., 2012). In real application, the combined minimum 8 days repeat frequency of two Landsat satellites images has a great potential to monitor most of the forest cover change and forest disturbances, but it often be markedly extended by the cloud cover (especially in rainforest areas) or duty cycle limitations (Hilker et al., 2009; Roy et al., 2008). At the local scale, Landsat images have a major obstacle for capturing accurate information about the short term forest cover change occurred through time, such as forest fires and deforestation. At the large scale, cloud free time-continuous Landsat datasets always cannot be collected with a short term interval, for example, the global Landsat image mosaics only available for the circa year of 1975, 1990, 2000, 2005 and 2010.

Compared with Landsat images, MODIS image has more frequent revisit coverage and wider swath scope. The revisit rate of MODIS image would be 1-2 days, and its swath width is more than 2300 km, which are both absolutely superior to those of Landsat images. In this case, MODIS is more able to provide fine temporal resolution cloud-free images to monitor short term forest cover changes and disturbances happened at local scale, and to produce global time-continuous cloud-free MODIS image mosaics for each year since 1999. However, the major obstacle of MODIS images is that the finest spatial resolution is only 240 m, and is always insufficient to observe and monitor forest cover change with complex spatial patterns.

In practice, forest cover patches are always smaller than the footprint of MODIS images and, as such, the coarse spatial resolution MODIS pixels are often mixed (Jin and Sader, 2005; Sexton et al., 2013), especially for spatial heterogeneous areas (Keshava and Mustard, 2002). To reduce the negative effects of the mixed pixel problem in MODIS images, spectral unmixing method that aims to estimate the compositions of mixed pixels are often applied to extract sub-pixel forest cover information (Hansen et al., 2003; Keshava and Mustard, 2002; Lu et al., 2011; Tottrup et al., 2007). However, the sub-pixel spatial location of the forest cover within the mixed pixels is not provided by the spectral unmixing technique.

Super-resolution mapping (SRM) is an approach aiming to produce finer spatial resolution land cover map from original coarse spatial resolution remotely sensed image by predicting sub-pixel spatial locations of different land covers within coarse pixels (Atkinson, 1997; Foody, 1998). Motivated by this, SRM is potentially able to increase the spatial resolution of the forest cover map produced from MODIS images. In general, the key of SRM is the definition of the spatial prior model that is used to guide the labeling of the sub-pixel land cover classes within the mixed pixel. The maximal spatial dependence principle is one of the most popular spatial prior models used for SRM. Algorithms, including pixel swapping (Atkinson, 2005) and sub-pixel/pixel spatial attraction (Ling et al., 2010; Mertens et al., 2006) are typical SRM methods based on the maximal spatial dependence principle.

Besides the spatial prior model, another key problem for SRM is the method used to deal with spectral information included in the original coarse spatial resolution multispectral image. Presently, there are two kinds of approaches can be used for the utilization of spectral information within the observed remotely sensed multispectral image. For the first type, spectral unmixing is performed first for the original multispectral image, and the generated proportional images are then used as the input of

a SRM analysis, and this kind of SRM is called as spatial-based SRM (Atkinson, 2005; Ge, 2013; Ge et al., 2009; Ling et al., 2014a; Ling et al., 2013; Ling et al., 2014b; Mertens et al., 2006; Tatem et al., 2002; Wang et al., 2014; Zhong et al., 2015). It is noteworthy that the spatial-based SRM model also utilizes the spectral information, but not directly, as it is disposed in the separated spectral unmixing model. In contrast, an alternative type of SRM method, spectral-spatial-based SRM, uses directly the original remotely sensed images as input to produce the fine spatial resolution land cover map (Zhang et al., 2015). Various algorithms, such as Markov random field based SRM (Kasetkasem et al., 2005; Li et al., 2012a), linear unmixing model based SRM (Ling et al., 2012) and supervised fuzzy *c*-means based SRM (Li et al., 2012b), all belong to this class of SRM method.

The results of spatial-based SRM rely greatly on the accuracy of the proportional images generated by spectral unmixing (Ling et al., 2012); but unfortunately, spectral unmixing is still an open problem. Compared with spatial-based SRM algorithms, spectral-spatial-based SRM operates directly with the original remotely sensed image and are able to reduce the uncertainty of spectral unmixing to some extent. Moreover, there is always a dilemma for spatial-based SRM to handle the spectral and spatial information simultaneously, and the input proportion images generated by spectral unmixing can only focus on a single spatial scale. Therefore, spatial-based SRM cannot deal directly with multispectral or hyperspectral images that are represented at multiple resolutions. By contrast, spectral-spatial-based SRM can use remotely sensed images with multiple scales as input and take full advantages of the available spectral information.

When a MODIS image is available, the abovementioned SRM algorithms, including spatial-based and spectral-spatial-based SRM, can thus be used readily to produce a fine spatial resolution (e.g. Landsat-based) forest cover map. For this case, the spatial-based and spectral-spatial-based SRM

algorithms are applied with mono-temporal remotely sensed images. Although this “mapping then comparison” approach is instinctive, it does not take full advantage of all available prior information, especially the temporal prior information included in the previous fine spatial resolution forest cover maps. To overcome this shortcoming, an alternative SRM approach was proposed to focus on updating, which means to update a previous fine spatial resolution forest cover map with a coarse spatial resolution remotely sensed image. Given the storage of large amount of Landsat series images during 1972 to now, the previous fine spatial resolution forest cover map can be produced readily from these historical Landsat images.

Ling et al. (2011) first developed a sub-pixel land cover change mapping (SLCCM) rule by comparing the current and previous coarse spatial resolution fraction land cover map, and to incorporate the temporal prior information within the previous fine spatial resolution land cover map into the updated fine spatial resolution land cover map. Based on the framework of SLCCM, some other updated-based spatial-temporal SRM approaches were proposed recently (He et al., 2016; Li et al., 2017; Li et al., 2015; Li et al., 2014b; Wang et al., 2016; Wang et al., 2015; Wu et al., 2017; Xu and Huang, 2014; Zhang et al., 2017). However, the inputs of these updated-based spatial-temporal SRM are coarse spatial resolution proportion images generated by spectral unmixing, and only consider the spatial and temporal prior information. Similar to the spatial-based SRM, the spatial-temporal SRM method also does not take the spectral information of the original coarse spatial resolution remotely sensed image into account, such that the uncertainty of spectral unmixing will greatly impact the corresponding results.

Given that spectral, spatial and temporal information all have an important role in the SRM; it is of great interest to consider integrating of all of them together in the SRM model. Motivated by this, Li

et al. (2014a) first proposed a spectral-spatial-temporal SRM model, namely the spatial-temporal Markov Random Field SRM method (STMRF), to deal with the spectral, spatial and temporal prior information simultaneously. The STMRF model is organized by solving a maximization problem that is composed of three key terms including spectral, spatial and temporal energy function. The STMRF model has also been applied to produce Landsat-based forest maps by using a coarse spatial resolution MODIS image and a previous Landsat-based forest map as input (Li et al., 2014a). However, it still has great potential to be further improved for updating fine spatial resolution forest maps with MODIS images in real applications. First, the spectral energy function of STMRF is based only on a single spatial scale remotely sensed image. MODIS multispectral imagery has seven bands, where bands 1-2 have a spatial resolution of 240 m and bands 3-7 have a spatial resolution of 480 m. Since STMRF is based on only a single scale, bands 1-2 are often degraded to 480 m, and then combined with bands 3-7 to be used as the input. However, spatial information within 240 m bands 1-2 would be lost during the degrading process. To take full advantage of the MODIS imagery, both the 480 m bands and 240 m bands of MODIS image should be utilized simultaneously. Second, the temporal energy function of STMRF aims only to make the updated fine spatial resolution forest map similar to the previous fine spatial resolution forest map. However, forest cover change certainly may exist between the updated and previous forest maps. In this case, the result of STMRF is sensitive to the forest cover change through time, and the changed pixels in the previous fine spatial resolution land cover map would have a serious negative impact on the updated fine spatial resolution land cover map. To have a better performance, the land cover changes between the updated and previous forest maps should be considered in the STMRF model.

This paper proposes a multiscale spectral-spatial-temporal SRM model (MSSTSRM) which can

be applied to update Landsat-based forest maps from multiscale MODIS images. Compared with the STMRF, the proposed MSSTSRM model includes innovation in terms of the spectral energy function and temporal energy function. The spectral term in the MSSTSRM model is based on a multiscale model, and both the 480 m and 240 m bands of MODIS image are utilized as the inputs of spectral energy function. The temporal term in the MSSTSRM model is based on a multiscale spatial-temporal neighborhood model, in which not only the sub-pixel scale spatial-temporal neighborhood system used to make the updated fine spatial resolution land cover map consistent with the previous land cover map, but also the coarse-pixel scale spatial-temporal neighborhood system used to incorporate the land cover change information is taken into account. As a result, the proposed MSSTSRM is able to take full advantage of the spectral, spatial and temporal prior information within the multiscale MODIS images and the previous Landsat-based forest maps.

2. Study area and methods

2.1 Study area and dataset

From the early 21st century to now, the rainforests in Amazon basin are undergoing continuous reducing due to the severe deforestation and integrated farming (Almeida et al., 2009; Morton et al., 2005). Monitoring the forest cover change in Amazon basin is, therefore, becoming an increasing important research body. It is noteworthy that the Mato Grosso State of Brazilian Amazon Basin is one of the worst-hit places that suffered forest logging, and in this research, two study sites in this state were chosen.

For the first study site, it was designed to validate the performance of MSSTSRM approach on forest map updating with strong forest cover change and long-time interval. The Landsat Enhanced Thematic Mapper Plus (ETM+) image (Path: 226, Row: 069) acquired on 23 July 2001 was used to

produce the previous subarea Landsat-based forest map, and the Landsat Operational Land Imager (OLI) image (Path: 226, Row: 069) acquired on 23 July 2015 was used to produce the current reference subarea Landsat-based forest map. The subarea forest maps covering the first study site have a spatial size of 800 × 800 pixels (covering 24 km × 24 km area), and the central geographical coordination is 12°34′6″S and 55°25′54″W . The MODIS/Terra Surface Reflectance Daily L2G Global composites of 240 m MOD09GQ and 480 m MOD09GA images (the Terra MODIS tile: h12v10) acquired on 22 July 2015 were used to generate the subarea coarse spatial resolution images for the proposed MSSTSRM approach.

To validate the performance of the proposed MSSTSRM approach on forest map updating with weak forest cover change but short-time interval in real application, the second study site located at the geographical coordination of 12°36′30″S and 55°50′39″W was chosen. The Landsat ETM+ image (Path: 226, Row: 069) acquired on 01 August 2010 was used to produce the previous subarea Landsat-based forest map, which has a spatial size of 2400 × 2400 pixels and covers 72 km × 72 km area. The MOD09GQ and MOD09GA images (the Terra MODIS tile: h12v10) acquired on 17 July 2010 were used to generate the subarea coarse spatial resolution images for the proposed MSSTSRM approach, and the Landsat ETM+ image (Path: 226, Row: 069) acquired on the same day (17 August 2010) was used to produce the corresponding reference subarea Landsat-based forest map. Moreover, during 01 July 2010 and 17 July 2010, another three time-series MOD09GQ and MOD09GA images acquired on 06, 10 and 15 August 2010 were used as the coarse spatial resolution images for MSSTSRM to update the time-series Landsat-based forest maps.

All of the abovementioned Landsat ETM+ and OLI images were obtained from the USGS Earth Explorer (http://earthexplorer.usgs.gov), while the MODIS composites of MOD09GQ and MOD09GA

were obtained from the NASA's Earth Observing System Data and Information System (EOSDIS, http://reverb.echo.nasa.gov/reverb). The original MODIS images and Landsat images have different Geographic Reference Systems, and there are two ways for them to achieve the same Geographic Reference Systems. The first widely used way is to reproject the MODIS images into the same Geographic Reference System of the original Landsat images, and the second is to reproject the Landsat images into the Geographic Reference System of MOD09GQ and MOD09GA images. The second way is based on the Landsat pixel, and the reprojecting error is also based on the Landsat pixel, which would be lower than the reprojecting error based on the MODIS pixel. Therefore, in order to make the MODIS images avoiding the reprojecting error based on MODIS pixel, all of the original Landsat images used in this paper were reprojected into the same Geographic Reference System of the original MOD09GQ and MOD09GA products.

2.2 Methods

2.2.1 MSSTSRM model objective function

Suppose that the original multiscale MODIS images are Y^{480} and Y^{240} at current time T_c, where current time T_c refers to the updating date. Y^{480} represents the 480 m MODIS images and Y^{240} represents the 240 m MODIS images. Each band of Y^{480} contains N pixels, and then each band of Y^{240} has $4\times N$ pixels, because each pixel in Y^{480} contains 2×2 pixels in Y^{240}. The previous fine spatial resolution land cover map is defined as X^{T_p} at previous time T_p with spatial resolution R and k land cover classes. The output of MSSTSRM is a fine spatial resolution land cover map X^{T_c} at T_c with spatial resolution R and k land cover classes. The scale factor between X^{T_c} and Y^{480} is defined as z, and each pixel in Y^{480} is divided into z^2 fine pixels in X^{T_c}. The scale factor between X^{T_c} and Y^{240} is defined as $z/2$, and each pixel in Y^{240} is divided

into $z^2/4$ fine pixels in X^{T_c}.

A general spectral-spatial-temporal SRM framework (MSSTSRM) is proposed here to find the solution of the fine spatial resolution land cover map X^{T_c}. In this framework, spectral, spatial and temporal goal functions are integrated to form a united optimization function E that is expressed as

$$\text{Min } E = E_{spectral} + \lambda_S \cdot E_{spatial} + \lambda_T \cdot E_{temporal}. \quad (1)$$

where $E_{spectral}$ is the spectral energy function and aims to make the fine spatial resolution land cover map X^{T_c} constrained to the original MODIS multiscale images Y^{480} and Y^{240}, respectively. $E_{spatial}$ is the spatial energy function, and is used to incorporate the spatial smoothness information into X^{T_c}. $E_{temporal}$ is the temporal energy function, which aims to incorporate the temporal information within the previous X^{T_p} into X^{T_c}. λ_S and λ_T are two trade-off parameters used to balance the contributions of the three terms. The class label of one fine pixel in updated X^{T_c} is determined by the combination of minimum values of the three different terms. The flowchart of the proposed MSSTSRM modeling process is shown in figure 1, and more details about it are presented in the following sections.

2.2.2 Spectral energy function

The spectral energy function $E_{spectral}$ is used here to describe the matching degree between the observed image pixel spectra of the MODIS multiscale images and the synthetic image pixel spectra calculated from the fine spatial resolution land cover map X^{T_c}. In this paper, the least square error method was utilized (Ling et al., 2012) to minimize the spectral value difference between the synthetic pixels and the observed pixels at both scales of 480 m MODIS image Y^{480} and 240 m MODIS image Y^{240}, respectively. Therefore, the spectral energy function $E_{spectral}$ is described as an integrated multi-objective spectral energy function of MODIS images Y^{480} and Y^{240}, and is formulated as

$$E_{spectral} = E_{spectral}^{480} + E_{spectral}^{240}, \quad (2)$$

where $E_{spectral}^{480}$ is the spectral energy function for 480 m MODIS image Y^{480}, and $E_{spectral}^{240}$ is that of 240 m MODIS image Y^{240}.

Let $y^{480}(i)$ be a column observed spectral pixel vector located at the ith pixel in MOD09GA image Y^{480}, and $y^{240}(j)$ be a column observed spectral pixel vector located at jth pixel in MOD09GQ image Y^{240}. For the 480 m Y^{480}, we assume that M^{480} is the signature matrix denoted by $\left[m_1^{480}, m_2^{480}, \cdots, m_k^{480}\right]$, where m_i^{480} is a column vector represented by the signature of the ith land cover class, and it is indeed the endmember values for the ith land cover class and collected by using the Pixel Purity Index algorithm (Chang and Plaza, 2006). $P_i^{480} = [p_1^{480}, p_2^{480}, \cdots, p_k^{480}]^T$ is a $k \times 1$ land cover proportion column vector associated with $y^{480}(i)$, where $p_1^{480}, p_2^{480}, \cdots, p_k^{480}$ are the proportion values for the different land cover classes in $y^{480}(i)$. Assuming that the spectra is a linear mixture of the target signature, and then the synthetic spectra $S^{480}(i)$ for MODIS image $y^{480}(i)$ can be represented as

$$S^{480}(i) = M^{480} \times P_i^{480} + e_i^{480}, \tag{3}$$

where e_i^{480} is the estimation errors for pixels $y^{480}(i)$ in the linear spectral mixture model. For the 240 m MODIS image y^{240}, let $M^{240} = \left[m_1^{240}, m_2^{240}, \cdots, m_k^{240}\right]$ be the signature matrix and m_i^{240} be a column vector represented by the signature of the ith land cover class, and $P_j^{240} = [p_1^{240}, p_2^{240}, \cdots, p_k^{240}]^T$ be a $k \times 1$ land cover proportion column vector associated with $y^{240}(j)$. Similar to that of synthetic spectra $S^{480}(i)$ shown in equation (3), the synthetic spectra $S^{240}(j)$ for MODIS image can be represented as

$$S^{240}(j) = M^{240} \times P_j^{240} + e_j^{240}, \tag{4}$$

where e_j^{240} is the estimation errors for pixels $y^{240}(j)$ in the linear spectral mixture model.

Since $y^{480}(i)$ and $y^{240}(j)$ are the observed spectral information within the original MODIS

images, and the synthetic spectrums $S^{480}(i)$ and $S^{240}(j)$ should thus be consistent to them. In this section, the least square error estimator is applied to describe the spectral signature difference between the observed spectral pixel values $y^{480}(i)$ and $y^{240}(j)$ and the synthetic image pixel spectra $S^{480}(i)$ and $S^{240}(j)$ calculated from the updated fine spatial resolution land cover map X^{T_c}. The spectral energy functions $E_{spectral}^{480}$ and $E_{spectral}^{240}$ based on the least square error estimator aim to achieve a minimum difference between observed spectrum and synthetic spectrum, and they can thus be expressed as

$$E_{spectral}^{480}(k(a_l^i)) = (y^{480}(i) - S^{480}(i))^T (y^{480}(i) - S^{480}(i)), \quad (5)$$

$$E_{spectral}^{240}(k(a_l^j)) = (y^{240}(j) - S^{240}(j))^T (y^{240}(j) - S^{240}(j)), \quad (6)$$

where $k(a_l^i)$ is the land cover class label for the lth sub-pixel a_l^i in coarse pixel $y^{480}(i)$, and $k(a_l^j)$ is the land cover class label for lth sub-pixel a_l^j in coarse pixel $y^{240}(j)$.

Therefore, according to equations (2), (5) and (6), the total spectral energy function $E_{spectral}$ for the updated fine spatial resolution land cover map X^{T_c} is written as

$$E_{spectral} = \frac{1}{2}\sum_i \sum_j \left[(y^{480}(i) - S^{480}(i))^T (y^{480}(i) - S^{480}(i)) + (y^{240}(j) - S^{240}(j))^T (y^{240}(j) - S^{240}(j)) \right]. \quad (7)$$

2.2.3 Spatial energy function

The spatial energy function $E_{spatial}$ is based on the current predicting fine spatial resolution land cover map X^{T_c}. It is always based on a certain spatial prior model, and aims to describe the spatial distribution of fine pixels X^{T_c}. Among various spatial prior models, the maximum spatial dependence model is the most widely used, due to its efficiency and simplicity. The maximum spatial dependence model is based on the principle that the class label of the target fine pixel is similar to the class labels of the spatially neighboring fine pixels (Atkinson, 2005), and then $E_{spatial}$ can be formulated as

$$E_{spatial} = \sum_{l=1}^{N \times z^2} \sum_{v \in N_v} \delta(k(l), k(v)) \cdot D(l, v), \quad (8)$$

$$\delta(k(l),k(v))=\begin{cases}1 \ \ if \ \ k(l)=k(v)\\ 0 \ if \ \ k(l)\neq k(v)\end{cases}. \tag{9}$$

The target class label $k(l)$ at time T_c is spatially dependent on the neighboring sub-pixel class labels $k(v)$ in the corresponding neighboring system N_v at current time T_c. $D(l,v)$ is the weight of a neighboring fine pixel v which spatially contributes to the central fine pixel l within the sub-pixel neighborhood system N_v. In general, $D(l,v)$ is calculated as

$$D(l,v)=e^{(-d(v,l)/\varphi)}, \tag{10}$$

where $D(v,l)$ is the geometric distance between the central fine pixel l and its neighboring fine pixel v, and φ is the nonlinear parameter of the distance decay model.

2.2.4 Temporal energy function

The temporal energy function $E_{temporal}$ is used here to incorporate the spatial pattern information within the previous fine spatial resolution land cover map X^{T_p} into the fine spatial resolution land cover map X^{T_c}. To link X^{T_p} and X^{T_c}, a spatial-temporal dependence model is applied (Ling et al., 2014c). Based on the spatial-temporal neighbor system as shown in figure 2(a), the spatial-temporal dependence model assumes that the class label of a fine pixel l in X^{T_c} is determined by the class labels of the neighboring fine pixel v_p centered at fine pixel l_p in X^{T_p}, where fine pixel l_p shares the same location as that of the target fine pixel l. Based on the sub-pixel spatial-temporal dependence model, $E_{temporal}$ is then formulated as

$$E_{temporal}=\sum_{l=1}^{N\times z^2}\sum_{v_p\in N_{v_p}^T}\delta(k(l),k(v_p))\cdot D(l,v_p), \tag{11}$$

$$\delta(k(l),k(v_p))=\begin{cases}1 \ \ if \ \ k(l)=k(v_p)\\ 0 \ if \ \ k(l)\neq k(v_p)\end{cases}, \tag{12}$$

where $k(l)$ is the class label of the target fine pixel at T_c, $k(v_p)$ is the class label of the temporal neighboring pixels at T_p, and $N_{v_p}^T$ is the fine pixel temporal neighboring system shown in figure 2(a).

$D(l, v_p)$ is the weight of a temporal neighboring fine pixel v_p spatially contributing to the central fine pixel l within the fine pixel temporal neighborhood system $N_{v_p}^T$. In general, $D(l, v_p)$ is calculated as

$$D(l, v_p) = e^{(-d(v_p, l)/\varphi)}, \tag{13}$$

where $d(v_p, l)$ is the geometric location distance between the central fine pixel l and its temporal neighboring fine pixel v_p, and φ is the nonlinear parameter of the distance decay model.

The spatial-temporal dependence model shown in equation (11) considers only that the target fine pixel at T_c is dependent on the temporal neighboring fine pixels at T_p, and the class label of the target fine pixel should be similar to the class labels of the temporal neighboring fine pixels. In practice, however, land cover change may occur between time T_p and T_c; in this case, the class label of the target fine pixel at T_c would not be similar to the temporal neighboring fine pixels at T_p. Therefore, the temporal land cover change information should also be considered in the spatial-temporal dependence model shown in equation (11).

To integrate the temporal land cover change information between X^{T_c} and X^{T_p}, a novel multiscale spatial-temporal dependence model is proposed in this section. The multiscale spatial-temporal dependence model assumes that the class label of the target fine pixel l is determined by the integrating contributions of temporal neighboring fine pixel v_p shown in figure 2(a) and temporal neighboring coarse pixels V_p and V shown in figure 2(b). Specifically, the multiscale spatial-temporal dependence model is built at both the sub- and coarse-pixel scales. The sub-pixel scale is based on the traditional spatial-temporal neighbor system shown in figure 2(a), and is used to make X^{T_c} similar to X^{T_p}. On the other hand, the coarse-pixel scale is based on the coarse-pixel spatial-temporal neighbor system shown in figure 2(b), and is used to incorporate the land cover change

information between X^{T_c} and X^{T_p}. Therefore, by applying the multiscale spatial-temporal dependence model, the temporal energy function $E_{temporal}$ is expressed as

$$E_{temporal} = \sum_{l=1}^{N\times z^2} \sum_{v_p \in N_{v_p}^T} \delta(k(l), k(v_p)) \cdot D(l, v_p) \cdot \tau(N_V, N_{V_p}^T), \quad (14)$$

where $\tau(N_V, N_{V_p}^T)$ is the temporal land cover change indicator value, and it is calculated based on the coarse pixel neighboring systems N_V and $N_{V_p}^T$ as shown in figure 2(b). Specially, N_V is the coarse pixel neighboring system of fine pixel l and $N_{V_p}^T$ is the coarse pixel neighboring system of fine pixel l_p. N_V and N_{V_p} have the same window size of w. $\tau(N_V, N_{V_p}^T)$ is described as

$$\tau(N_V, N_{V_p}^T) = e^{\frac{-6}{w\times w}\times\sqrt{\sum_{V,V_p=1}^{w\times w}\left|p(V,k(l))-p(V_p,k(l))\right|^2}}, \quad (15)$$

where $p(V, k(l))$ and $p(V_p, k(l))$ are the proportion values of land cover class $k(l)$ for the coarse pixels V in N_V and V_p in N_{V_p}, respectively. Specially, the coarse pixel V_p and V are based on the proportion images calculated from the previous fine spatial resolution land cover map X^{T_p} and current predicted fine spatial resolution land cover map X^{T_c} by using spatially averaging filter.

2.2.5 Model optimization and accuracy assessment

The Iterative Conditional Model (ICM) (Besag, 1986) is used here to minimize the global energy function shown in (1) iteratively, and the implementation process is formulated as:

1) Extract the endmember spectra for different land cover classes, and obtain the coarse spatial resolution proportion images by using soft classification, such as linear spectral unmixing model. The current fine spatial resolution land cover map is then initiated in terms of the generated proportion images.

2) Update the sub-pixel class labels of the entire image in terms of equation (1) by integrating the spectral, spatial and temporal energy functions shown in equations (2), (8) and (11), respectively,

and the class label of the sub-pixel that contributes to the minimal energy is accepted as the class label.

3) The iteration is terminated when less than 0.1% of the total number of sub-pixels is changed after two consecutive iterations; otherwise, return to Step (2).

4) Obtain the final fine spatial resolution forest map by labeling the sub-pixels which do not belong to forest as non-forest labels in the resultant fine spatial resolution land cover map.

Four popular methods, including support vector machines (SVM) classification based on the MODIS 480 m image (SVM480), SVM based on the MODIS 240 m image (SVM240), Markov random field based SRM (MRFSRM) (Kasetkasem et al., 2005) and spatial-temporal Markov Random Field based SRM (STMRF) (Li et al., 2014a), were used as the comparison methods to validate the proposed MSSTSRM. For the assessment indices, the Kappa value, overall accuracy, producer's accuracy, user's accuracy, omission error and commission error were used in the experiments to assess the accuracy of the proposed MSSTSRM against other algorithms. The omission error and commission error were used to assess the categorical-level accuracy, whereas the overall accuracy and Kappa value were used to assess the map-level accuracy. All the accuracy values were measured at the sub-pixel scale which is identical to the spatial resolution of the reference map.

3. Experimental results

3.1 Experiment implementation

In the first study site, the subarea Landsat ETM+ and OLI images shown in figure 3(a)-(b) were used to produce the previous and current reference fine spatial resolution forest maps shown in figure 4(a)-(b) by applying the support vector machines (SVM) classifier, due to its superiority in remote sensing image classification (Mountrakis et al., 2011). In the second study site, since there are stripes in

the Landsat ETM+ SLC-off images shown in figure 7(a)-(b), general remote sensing image classification method, such as SVM, cannot be used to produce the forest map. Therefore, the previous and current reference fine spatial resolution forest maps covering the second study site (see figure 8(a)-(b)) were generated by digitizing manually the Landsat ETM+ image shown in figure 7(a)-(b) with the help of Google Earth Map. Moreover, SVM classifier was also used to produce the 240 m and 480 m coarse spatial resolution forest maps (see figure 4(d)-(e) and figure 8(d)-(e)) for both of the two study sites from the MOD09GQ and MOD09GA images (see figure 3(c)-(d) and figure 7(c)-(d)). During the SVM classification processing, training samples of three land cover classes, including forest, grass, bare land, were randomly collected from the Landsat and MODIS images by artificial visual interpretation with the help of Google Earth Map, and the grass and bare land were combined into the non-forest in the final resultant forest map.

For the SRM methods of MRFSRM, STMRF and the proposed MSSTSRM, endmembers of different land covers should be collected first. It is noted that the non-forest is indeed composed of many different land cover classes with varied spectral features, and the intra-class spectral variance of the non-forest class is high. Therefore, by using the Pixel Purity Index algorithm (Chang and Plaza, 2006), three land cover classes, including forest, grass, bare land were used, and the corresponding endmember spectral features were chosen for the two study sites. Since the spatial resolutions of MOD09GQ and MOD09GA are different, endmembers were then collected individually from them. Since MRFSRM and STMRF are based on the 480 m MOD09GA image, endmembers of the three land cover classes were collected from the original 480 m MOD09GA image with seven bands. For the proposed MSSTSRM approach, endmembers of the three land cover classes were collected individually from the 240 m MOD09GQ and 480 m MOD09GA images and used as the inputs of

different classifications.

3.2 Experiment results and analysis

3.2.1 The first study site

For the first study site, the updated forest maps of four classification methods are shown in figure 4, and the corresponding forest error maps generated by comparing with the reference forest map (see figure 4(b)) are shown in Figure 5. Specially, figure 4(c) shows the differences between the previous fine spatial resolution forest map at T_p (see figure 4(a)) and the current reference fine spatial resolution forest map at T_p (see figure 4(b)). Indeed, figure 4(a) is the forest cover change map between figure 4(a) and figure 4(b), and the change ratio is 21.94%.

For the result of SVM based on MOD09GA (see figure 4(d)), the boundaries were mapped as jagged squares; meanwhile, numerous spatial details, especially the small-sized forest features, were lost and more forest pixels in the reference forest map were incorrectly mapped as non-forest pixels shown in the forest error map of figure 5(a) (blue pixels). This is because the classification rule of SVM is undertaken at the pixel scale, and the spatial resolution of MOD09GA is too coarse to provide spatially smooth boundaries and represent the spatial detail of forest cover. Compared with the result of SVM480, SVM based on the MOD09GQ produced a forest map (see figure 4(e)) with more spatial detail and the jagged boundaries are smaller than that of SVM480; however, many non-forest pixels in the reference forest map are incorrectly mapped as forest pixels (the red pixels in figure 5(b)) in the result of SVM240. The reason for this is that the MOD09GA image has only two bands, which are insufficient to distinguish forest pixels from the other land covers. For the result of MRFSRM, as shown in figure 4(f), jagged boundaries as shown in the results of SVM480 and SVM240 are spatially smooth and greater spatial detail is incorporated, and this is due to the sub-pixel based classification

rule of SRM. Compared with figure 5(a), fewer forest pixels were incorrectly mapped as non-forest (blue pixels), but more non-forest pixels were incorrectly mapped as forest (red pixels) in the forest error map of MRFSRM (see figure 5(c)).

SVM and MRFSRM cannot utilize the spatial-temporal information of the previous fine spatial land cover map in the model processing, and the results are always limited in spatial detail. Compared with MRFSRM, a temporal energy function is added in the model processing of STMRF, and the result of STMRF (see figure 4(g)) contains more spatial detail, especially the small-sized linear forest features. However, more non-forest pixels in the reference forest map were mapped as forest pixels (red pixels) as shown in figure 5(d). This occurs because that STMRF is a global model and only considers the sub-pixel temporal link between the previous and current fine spatial resolution forest maps, and then cannot deal with the local forest cover change of temporal information as shown in figure 4(c). To take advantage of the multiscale MODIS images, the proposed MSSTSRM is based on a multiscale model, and both the MOD09GA and MOD09GQ were used as the inputs. The result of MSSTSRM, as shown in figure 4(h), has more spatially smooth boundaries, and greater spatial detail is exploited. Moreover, more of the forest pixels are correctly mapped than those of the SVM, MRFSRM and STMRF methods, as shown in the forest error maps (see figure 5(e)). The result of MSSTSRM is more similar to the reference forest maps shown in figure 4(b). This is because MSSTSRM not only benefits from the abundant spectral information in MOD09GA, but also the useful spatial detail in MOD09GQ; on the other hand, the useful spatial-temporal information within the previous fine spatial resolution forest map is also incorporated efficiently into the result of MSSTSRM. Compared with the result of STMRF, the forest cover change temporal information was reduced greatly in the result of MSSTSRM, suggesting that the proposed approach is more able to balance the temporal and spatial information

within the previous fine spatial resolution forest map than STMRF.

The statistical accuracies are listed in table 1. SVM480 results have the smallest Kappa, OA, PA and UA values, and the RMSE values are the largest, because the spatial resolution of MOD09GA is too coarse. Compared with the results of SVM480, the results of SVM240 has higher Kappa, OA, PA and UA values, and this is because MOD09GQ has a finer spatial resolution than that of MOD09GA. Compared with SVM based on the MOD09GA image, the MRFSRM result is based on the MOD09GA image and has larger Kappa, OA, PA and UA values, and smaller RMSE; however, the improvement is not significant, which are lower than those of SVM240. By incorporating the temporal information into the STMRF model processing, the accuracy for the results of STMRF based on MOD09GA is better than those of MRFSRM, but it is still weaker than that of SVM240. The reason behind this is that STMRF is not able to balance the temporal change information between the previous and current fine spatial resolution forest maps. By contrast, the updated forest map produced by the proposed MSSTSRM has the largest Kappa and OA, PA and UA values, and lowest RMSE. Table 1 also lists the computing time of different classification methods. MSSTSRM is developed from the STMRF, but it has more complex spectral and temporal energy functions and requires more time to be completed, especially when the study area is large.

The trade-off parameter λ_T plays an important role in the performance of the proposed MSSTSRM. As shown in figure 6 and table 2, when λ_T is 0 the result is limited in spatial detail, such as missing the narrow linear forest features, and the accuracy value is the lowest. This is because there is no temporal information incorporated into the result when λ_T is 0. With an increase of λ_T, more spatial detail is exploited, and some of the boundaries become spatially smooth and the corresponding accuracy values also have an increase, which shows the advantages of integrating the temporal

information. However, with a continued increase of λ_T, the impact of the temporal information within the previous fine spatial resolution forest map will become stronger, and the temporal change information as shown in figure 4(c) potentially has a negative contribution to the result. The temporal information within the previous forest map should provide a positive influence on the final forest map at the current time with a suitable value of λ_T; however, many changed forest features between the previous and current time provide a negative influence on the final forest map if the value of λ_T is set too large. On the other hand, trade-off parameter λ_S plays an important role to provide spatial smoothness prior information for the proposed MSSTSRM. Figure 7 shows the MSSTSRM results with different values of λ_S, and table 3 lists the accuracy values of the corresponding results. When λ_S is 0, the spatial energy function would not provide any contribution to the solution of the final result, many isolated pixels and patches appear around the boundaries and the accuracy values are the lowest. With the increase of λ_S, the isolated patches were gradually eliminated, and achieved fine statistical accuracies. However, when λ_S is 80, some forest features are over-smoothed, and the corresponding accuracy become decreasing. Therefore, to obtain an optimal result with the proposed MSSTSRM approach, the trade-off involved in parameters λ_T and λ_S should be considered comprehensively.

3.2.2 The second study site

For the second study site, the original Landsat false color images and mutilscale MODIS images are shown in figure 8, and the resultant forest maps and corresponding forest error maps produced by different classification methods are shown in figure 9. In this study site, the time gap between the previous and current forest maps is 16 days, and the forest cover changes shown in figure 9(c) are much smaller than that shown in the first study site. For clearer visual comparison between the different results, five zoomed subareas of the produced forest maps and forest error maps are also shown in

figure 9. Similar trends as shown in the first study site can also be observed in this study site. SVM results based on MOD09GA and MOD09GQ have jagged boundaries, and numerous spatial details are missing. MRFSRM result has spatially smooth boundaries, but it still lacks amount of spatial detail information about forest cover. Since STMRF cannot deal with the forest cover changes through time, many non-forest pixels were incorrectly mapped as forest (red pixels). Compared with the other classification results, MSSTSRM produced a fine spatial resolution forest map with more spatially smooth boundaries and greater spatial detail, and the number of mis-classifications (red and blue pixels in figure 9(m)) within the forest error map was greatly reduced. The result of MSSTSRM is more similar to the reference forest map as shown in figure 9(a); and for the statistical accuracies listed in table 4, MSSTSRM not only has the largest Kappa, OA, PA and UA values, but also the smallest RMSE, which demonstrates the efficiency of the proposed approach in real application.

As shown in table 5, during 01 August and 17 August 2010, there are two scenes of Landsat ETM+ images (no Landsat TM images), but there are five cloud-free MOD09GA and MOD09GQ images covering the second study sites, and the false color Landsat ETM+ and MOD09GA images are presented in figure 10(a)-(e). Only based on the forest maps extracted from the Landsat ETM+ images shown in figure 10(a) and figure 10(j), the gradual forest cover change information between 01 August and 17 August 2010 was lost. By contrast, the MOD09GA images acquired on 6, 10 and 15 August 2010 can supplement the missing forest cover information, but its spatial resolution is too coarse to observe the forest cover spatial detail. The proposed MSSTSRM approach was, therefore, applied to update the 30 m forest maps shown in figure 10(g)-(i) from the MOD09GA and MOD09GQ images acquired on 6, 10 and 15 August 2010. Compared with the MOD09GA images as shown in figure 10(b)-(d), the corresponding updated 30 m forest maps present much spatial detail about the forest

covers, especially the gradual fine spatial resolution forest cover change detail as clearly shown in the zoomed areas.

4. Discussion

4.1 Repeat frequency

Single Landsat satellite image has a minimal repeat frequency of 16 days, and the combined two Landsat satellites images would have an ideal repeat frequency of 8 days, but it is still beyond the requirement of some special forest cover mapping situations, which was demonstrated in the experiment for the second study site. Moreover, in real situation, because of the duty cycle gap of different Landsat satellites, such as Landsat-5 and Landsat-7 or Landsat-8 and Landsat-7, the synthetical repeat frequency can be hardly shorten from 16 days to 8 days. For the proposed MSSTSRM approach, the repeat frequency of the updated Landsat-based forest map is basically relied on the MODIS images, and is, therefore, having a great potential to produce ideally daily Landsat-based forest map in real applications.

It is noted worthy that Sentinel-2A satellite lunched at 23 June 2015, and it has a repeat frequency of 10 days. Moreover, with the recently lunched (07 March 2017) Sentinel-2B satellite, the minimal synthetical repeat frequency of the Sentinel-2A and Sentinel-2B satellites would be shorten to 5 days. Motivated by this, the Sentinel-2 images can also be used together with the Landsat-7 and Landsat-8 images to have an expected repeat frequency monitoring of the forest cover. However, there is no Sentinel-2A and Sentinel-2B dataset before 2015, and thus historic forest cover monitoring cannot benefit from it. Therefore, there is an urgent need for the proposed MSSTSRM to update Landsat-based forest maps with high repeat frequency before 2015. On the other hand, it is of great interest to further improve MSSTSRM to fuse Landsat and Sentinel-2 images, and have a finer spatial and temporal

resolution monitoring of the forest cover.

4.2 Error sources

When applying the MSSTSRM model, there are two major error sources that could impose a considerably negative effect on the updated Landsat-based forest map. The first is the error caused by endmember selection. Generally, endmember selection is a common but challenging problem in the field of supervised classification of remotely sensed images (Bateson and Curtiss, 1996; Dennison and Roberts, 2003). Given that the collection of endmember spectra is a time-consuming task and is always empirical, if the endmembers were selected too less, forest cover features would have high risk to be confused with other land covers; otherwise, it would be difficult to fully recognize the forest cover features. MSSTSRM is indeed a supervised classification method, and the error of endmember selection would affect directly the performance of MSSTSRM. The second is the estimation of parameter values of the two trade-off parameters λ_S and λ_T. From the above experiment, it can be found that λ_S and λ_T have an important impact on the performance of the proposed MSSTSRM approach. In future research, a more automatic method will be explored for estimating the trade-off parameters λ_S and λ_T which balance the contributions of the spatial and temporal terms in MSSTSRM.

5. Conclusion

To achieve fine spatial and temporal resolution forest maps, in this paper, a novel multiscale spectral-spatial-temporal SRM (MSSTSRM) method was proposed to update Landsat-based forest map from current MODIS images and previous Landsat forest maps. The proposed MSSTSRM method involves solution of a multiobject optimization problem, composed of three key terms: spectral, spatial and temporal energy functions. Considering that MODIS images have two different spatial resolution

bands, the spectral energy function in MSSTSRM is based on a multiscale model, and both of the 240 m bands and 480 bands of the MODIS image are used as inputs. For the spatial energy function, MSSTSRM is based on the maximal spatial dependence model to provide spatially smooth information. To deal with the temporal change information between the previous and current fine spatial resolution forest maps, the temporal energy function in MSSTSRM not only considers the sub-pixel spatial-temporal neighbor system, but also the land cover change information within the coarse-pixel spatial-temporal neighbor system. Experiments demonstrated that the proposed MSSTSRM approach is able to produce updated Landsat-based forest maps with the greatest visual and quantitative accuracy compared with three benchmark methods, and can produce time-serious Landsat-based forest maps with finer repeat frequency than that of Landsat image, which was demonstrated in the results shown in figure 9.

Acknowledgment

This work was partly supported by the Youth Innovation Promotion Association CAS (Grant No. 2017384), in part by the Natural Science Foundation of China (Grant No. 61671425) and the State Key Laboratory of Resources and Environmental Informational System of China.

References

Almeida, R., Shimabukuro, Y.E., Rosenqvist, A., Sanchez, G.A., 2009. Using dual-polarized ALOS PALSAR data for detecting new fronts of deforestation in the Brazilian Amazonia. Int. J. Remote Sens. 30, 3735-3743.

Atkinson, P.M., 1997. Mapping sub-pixel boundaries from remotely sensed images. Innovations in GIS 4, 166-180.

Atkinson, P.M., 2005. Sub-pixel target mapping from soft-classified, remotely sensed imagery. Photogramm. Eng. Remote Sens. 71, 839-846.

Bateson, A., Curtiss, B., 1996. A method for manual endmember selection and spectral unmixing. Remote Sens. Environ. 55, 229-243.

Besag, J., 1986. On the Statistical Analysis of Dirty Pictures. Journal of the Royal Statistical Society. Series B (Methodological) 48, 259-302.

Canadell, J.G., Raupach, M.R., 2008. Managing forests for climate change mitigation. Science 320, 1456-1457.

Chang, C.I., Plaza, A., 2006. A fast iterative algorithm for implementation of pixel purity index. IEEE Geosci. Remote Sens. Lett. 3, 63-67.

Dennison, P.E., Roberts, D.A., 2003. Endmember selection for multiple endmember spectral mixture analysis using endmember average RMSE. Remote Sens. Environ. 87, 123-135.

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

Foody, G.M., 1998. Sharpening fuzzy classification output to refine the representation of sub-pixel land cover distribution. Int. J. Remote Sens. 19, 2593-2599.

Friedl, M.A., McIver, D.K., Hodges, J.C.F., Zhang, X.Y., Muchoney, D., Strahler, A.H., Woodcock, C.E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F., Schaaf, C., 2002. Global land cover mapping from MODIS: algorithms and early results. Remote Sens. Environ. 83, 287-302.

Ge, Y., 2013. Sub-pixel land-cover mapping with improved fraction images upon multiple-point simulation. Int. J. Appl. Earth Obs. Geoinf. 22, 115-126.

Ge, Y., Li, S.P., Lakhan, V.C., 2009. Development and Testing of a Subpixel Mapping Algorithm. IEEE Trans. Geosci. Remote Sensing 47, 2155-2164.

Gong, P., Miller, J.R., Spanner, M., 1994. Forest Canopy Closure from Classification and Spectral Unmixing of Scene Components - Multisensor Evaluation of an Open Canopy. IEEE Trans. Geosci. Remote Sensing 32, 1067-1080.

Hansen, M.C., DeFries, R.S., Townshend, J.R.G., Carroll, M., Dimiceli, C., Sohlberg, R.A., 2003. Global Percent Tree Cover at a Spatial Resolution of 500 Meters: First Results of the MODIS Vegetation Continuous Fields Algorithm. Earth Interact 7.

Hansen, M.C., Krylov, A., Tyukavina, A., Potapov, P.V., Turubanova, S., Zutta, B., Ifo, S., Margono, B., Stolle, F., Moore, R., 2016. Humid tropical forest disturbance alerts using Landsat data. Environ Res Lett 11.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change.

Science 342, 850-853.

He, D., Zhong, Y.F., Feng, R.Y., Zhang, L.P., 2016. Spatial-Temporal Sub-Pixel Mapping Based on Swarm Intelligence Theory. Remote Sens-Basel 8.

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

Jin, S.M., Sader, S.A., 2005. Comparison of time series tasseled cap wetness and the normalized difference moisture index in detecting forest disturbances. Remote Sens. Environ 94, 364-372.

Jin, S.M., Yang, L.M., Danielson, P., Homer, C., Fry, J., Xian, G., 2013. A comprehensive change detection method for updating the National Land Cover Database to circa 2011. Remote Sens. Environ 132, 159-175.

Kasetkasem, T., Arora, M.K., Varshney, P.K., 2005. Super-resolution land cover mapping using a Markov random field based approach. Remote Sens. Environ. 96, 302-314.

Keshava, N., Mustard, J.F., 2002. Spectral unmixing. IEEE Signal Process. Mag. 19, 44-57.

Kim, D.H., Sexton, J.O., Noojipady, P., Huang, C.Q., Anand, A., Channan, S., Feng, M., Townshend, J.R., 2014. Global, Landsat-based forest-cover change from 1990 to 2000. Remote Sens. Environ 155, 178-193.

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

Li, X.D., Du, Y., Ling, F., 2012a. Spatially adaptive smoothing parameter selection for Markov random field based sub-pixel mapping of remotely sensed images. Int. J. Remote Sens. 33, 7886-7901.

Li, X.D., Du, Y., Ling, F., 2014a. 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 Observ. Remote Sens. 7, 29-39.

Li, X.D., Du, Y., Ling, F., 2015. Sub-pixel-scale Land Cover Map Updating by Integrating Change Detection and Sub-Pixel Mapping. Photogramm. Eng. Remote Sens. 81, 59-67.

Li, X.D., Ling, F., Du, Y., 2012b. Super-resolution mapping based on the supervised fuzzy c-means approach. Remote Sens. Lett 3, 501-510.

Li, X.D., Ling, F., Du, Y., Feng, Q., Zhang, Y.H., 2014b. A spatial-temporal Hopfield neural network approach for super-resolution land cover mapping with multi-temporal different resolution remotely sensed images. ISPRS-J. Photogramm. Remote Sens 93, 76-87.

Ling, F., Du, Y., Li, X.D., Zhang, Y., Xiao, F., Fang, S., Li, W., 2014a. Superresolution Land Cover Mapping With Multiscale Information by Fusing Local Smoothness Prior and Downscaled Coarse Fractions. IEEE Trans. Geosci. Remote Sens 52, 5677-5692.

Ling, F., Du, Y., Li, X.D., Li, W.B., Xiao, F., Zhang, Y.H., 2013. Interpolation-based super-resolution land cover mapping. Remote Sens. Lett 4, 629-638.

Ling, F., Du, Y., Xiao, F., Li, X.D., 2012. Subpixel Land Cover Mapping by Integrating Spectral and Spatial Information of Remotely Sensed Imagery. IEEE Geosci. Remote Sens. Lett 9, 408-412.

Ling, F., Du, Y., Xiao, F., Xue, H.P., Wu, S.J., 2010. Super-resolution land-cover mapping using multiple sub-pixel shifted remotely sensed images. Int. J. Remote Sens. 31, 5023-5040.

Ling, F., Li, W.B., Du, Y., Li, X.D., 2011. Land Cover Change Mapping at the Subpixel Scale With Different Spatial-Resolution Remotely Sensed Imagery. IEEE Geosci. Remote Sens. Lett. 8, 182-186.

Ling, F., Li, X., Xiao, F., Du, Y., 2014b. Superresolution Land Cover Mapping Using Spatial Regularization. IEEE Trans. Geosci. Remote Sens 52, 4424-4439.

Ling, F., Li, X.D., Du, Y., Xiao, F., 2014c. Super-Resolution Land Cover Mapping with Spatial-Temporal Dependence by Integrating a Former Fine Resolution Map. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 7, 1816-1825.

Lu, D.S., Batistella, M., Moran, E., Hetrick, S., Alves, D., Brondizio, E., 2011. Fractional forest cover mapping in the Brazilian Amazon with a combination of MODIS and TM images. Int. J. Remote Sens. 32, 7131-7149.

Malhi, Y., Roberts, J.T., Betts, R.A., Killeen, T.J., Li, W.H., Nobre, C.A., 2008. Climate change, deforestation, and the fate of the Amazon. Science 319, 169-172.

Mertens, K.C., De Baets, B., Verbeke, L.P.C., De Wulf, R.R., 2006. A sub-pixel mapping algorithm based on sub-pixel/pixel spatial attraction models. Int. J. Remote Sens 27, 3293-3310.

Morton, D.C., DeFries, R.S., Shimabukuro, Y.E., Anderson, L.O., Espirito-Santo, F.D.B., Hansen, M., Carroll, M., 2005. Rapid assessment of annual deforestaion in the Brazilian Amazon using MODIS data. Earth Interact 9.

Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: A review. ISPRS-J. Photogramm. Remote Sens. 66, 247-259.

Roy, D.P., Ju, J., Lewis, P., Schaaf, C., Gao, F., Hansen, M., Lindquist, E., 2008. Multi-temporal MODIS-Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data. Remote Sens. Environ. 112, 3112-3130.

Sexton, J.O., Song, X.P., Feng, M., Noojipady, P., Anand, A., Huang, C.Q., Kim, D.H., Collins, K.M., Channan, S., DiMiceli, C., Townshend, J.R., 2013. Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. Int J Digit Earth 6, 427-448.

Tatem, A.J., Lewis, H.G., Atkinson, P.M., Nixon, M.S., 2002. Super-resolution land cover pattern prediction using a Hopfield neural network. Remote Sens. Environ. 79, 1-14.

Tottrup, C., Rasmussen, M.S., Eklundh, L., Jonsson, P., 2007. Mapping fractional forest cover across the highlands of mainland Southeast Asia using MODIS data and regression tree modelling. Int. J. Remote Sens 28, 23-46.

Wang, Q.M, Shi, W., Atkinson, P.M., 2014. Sub-pixel mapping of remote sensing images based on radial basis function interpolation. ISPRS-J. Photogramm. Remote Sens. 92, 1-15.

Wang, Q.M., Shi, W.Z., Atkinson, P.M., 2016. Spatiotemporal Subpixel Mapping of Time-Series Images. IEEE Trans. Geosci. Remote Sensing 54, 5397-5411.

Wang, Q.M., Shi, W.Z., Atkinson, P.M., Li, Z.B., 2015. Land Cover Change Detection at Subpixel Resolution With a Hopfield Neural Network. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 8, 1339-1352.

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

Xian, G., Homer, C., Fry, J., 2009. Updating the 2001 National Land Cover Database land cover classification to 2006 by using Landsat imagery change detection methods. Remote Sensing of Environment 113, 1133-1147.

Xu, Y., Huang, B., 2014. A Spatio-Temporal Pixel-Swapping Algorithm for Subpixel Land Cover Mapping. IEEE Geosci. Remote Sens. Lett 11, 474-478.

Zhang, Y.H., Atkinson, P.M., Li, X., Ling, F., Wang, Q., Du, Y., 2017. Learning-Based

Spatial-Temporal Superresolution Mapping of Forest Cover With MODIS Images. IEEE Trans. Geosci. Remote Sensing 55, 600-614.

Zhang, Y.H., Du, Y., Ling, F., Wang, X., Li, X.D., 2015. Spectral-spatial based sub-pixel mapping of remotely sensed imagery with multi-scale spatial dependence. Int. J. Remote Sens. 36, 2831-2850.

Zhong, Y.F., Wu, Y., Xu, X., Zhang, L., 2015. An Adaptive Subpixel Mapping Method Based on MAP Model and Class Determination Strategy for Hyperspectral Remote Sensing Imagery. IEEE Trans. Geosci. Remote Sensing 53, 1411-1426.

Zhu, Z., Woodcock, C.E., Olofsson, P., 2012. Continuous monitoring of forest disturbance using all available Landsat imagery. Remote Sens. Environ. 122, 75-91.

Table 1. Kappa, overall accuracy (OA), RMSE, Producer's accuracy (PA) and User's accuracy (UA) of the resultant forest maps generated by the different methods for the first study site.

				PA		UA		
	Kappa	OA	RMSE	forest	non-forest	forest	non-forest	Time/s
SVM480	0.7290	86.77%	0.2789	88.26%	85.77%	80.63%	91.59%	0.3
SVM240	0.8259	91.63%	0.1528	89.72%	92.91%	89.47%	93.09%	0.3
MRFSRM	0.7576	88.35%	0.2083	85.47%	90.29%	85.52%	90.25%	501.4
STMRF	0.7740	88.99%	0.2059	90.48%	87.99%	83.49%	93.23%	624.8
MSSTSRM	0.8741	93.868%	0.1272	96.31%	92.23%	89.27%	97.39%	1146.8

Notes: SVM480 means SVM based on 480 m MOD09GA image, and SVM240 means SVM based on MOD09GQ image.

Table 2. Kappa, overall accuracy (OA), RMSE, omission error and commission error of the resultant forest maps generated by the proposed MSSTSRM with different values of λ_T for the first study site.

λ_T	Kappa	OA	RMSE	Omission Errors	Commission Errors
0	0.8490	92.72%	0.1329	8.20%	9.76%
50	0.8641	93.41%	0.1278	6.59%	10.15%
100	0.8703	93.69%	0.1279	4.34%	10.62%
150	0.8692	93.63%	0.1334	3.81%	11.15%
400	0.8448	92.38%	0.1745	2.99%	14.15%

Table 3. Kappa, overall accuracy (OA), RMSE, omission error and commission error of the resultant forest maps generated by the proposed MSSTSRM with different values of λ_S for the first study site.

λ_S	Kappa	OA	RMSE	Omission Errors	Commission Errors
0	0.8529	92.82%	0.1444	4.30%	12.42%
20	0.8644	93.40%	0.1335	4.29%	11.26%
40	0.8709	93.71%	0.1292	3.86%	10.62%
60	0.8741	93.89%	0.1272	3.69%	10.93%
80	0.8713	93.73%	0.1316	3.75%	10.97%

Table 4. Kappa, overall accuracy (OA), RMSE, Producer's accuracy (PA) and User's accuracy (UA) of the resultant forest maps generated by the different methods for the second study site.

				PA		UA		
	Kappa	OA	RMSE	forest	non-forest	forest	non-forest	Time/s
SVM480	0.7648	91.12%	0.2307	96.25%	77.18%	91.98%	88.33%	2.7
SVM240	0.8096	92.94%	0.1689	98.51%	77.81%	92.35%	95.04%	2.7
MRFSRM	0.7785	91.50%	0.1881	95.55%	80.47%	93.01%	86.93%	2906.5
STMRF	0.8283	93.45%	0.1652	97.28%	83.05%	93.98%	91.82%	3489.8
MSSTSRM	0.9359	97.51%	0.0970	99.11%	93.17%	97.53%	97.47%	6566.6

Notes: SVM480 means SVM based on 480 m MOD09GA image, and SVM240 means SVM based on MOD09GQ image.

Table 5. Available Landsat TM, Landsat EMT+, MOD09GA, MOD09GQ images for the second study site between 01 August and 17 August 2010.

Time	Landsat TM (P226R69)	Landsat ETM+ (P226R69)	MOD09GA (h12v10)	MOD09GQ (h12v10)
1-Aug-2010	×	✓	✓	✓
6-Aug-2010	×	×	✓	✓
10-Aug-2010	×	×	✓	✓
15-Aug-2010	×	×	✓	✓
17-Aug-2010	×	✓	✓	✓

Notes: × means unavailable, ✓ means available.

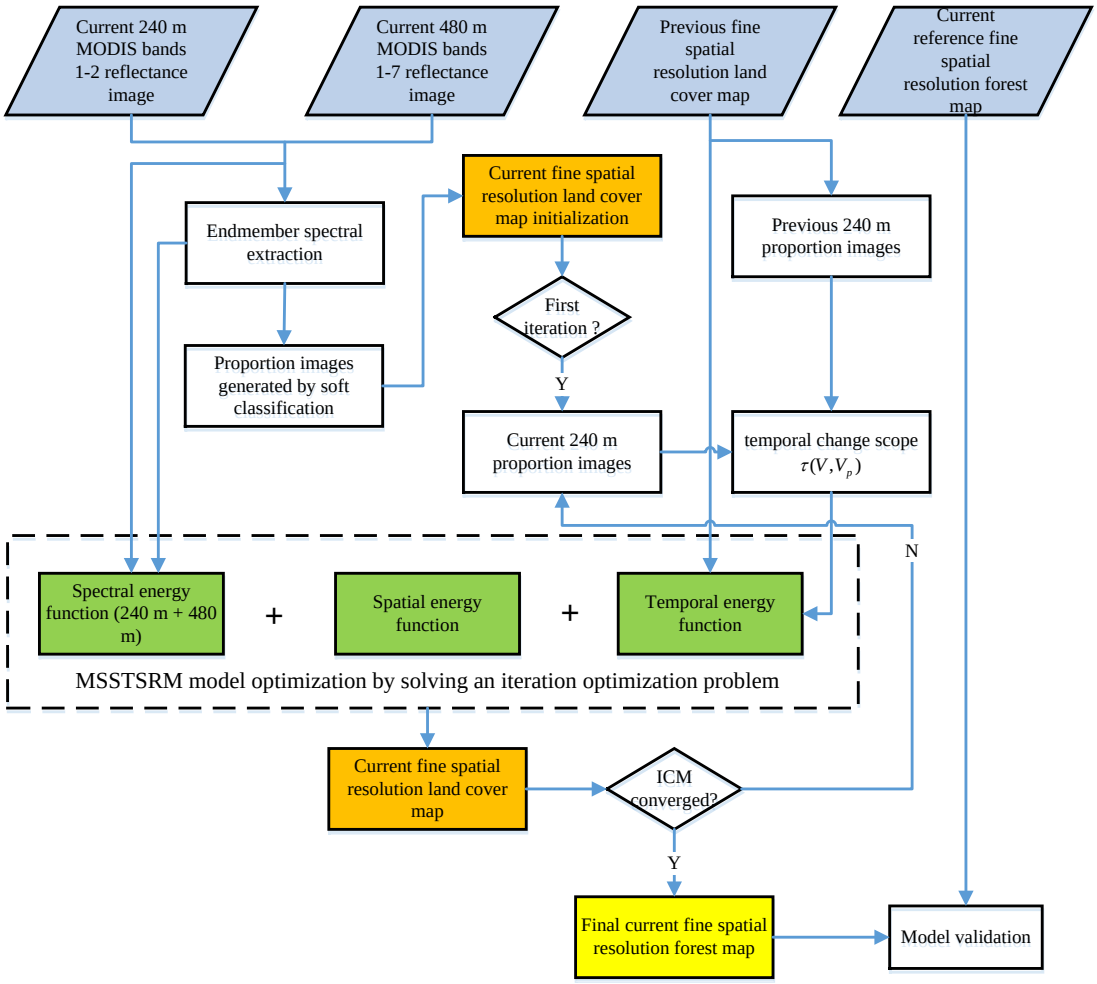

Figure 1. Flowchart of the proposed MSSTSRM approach.

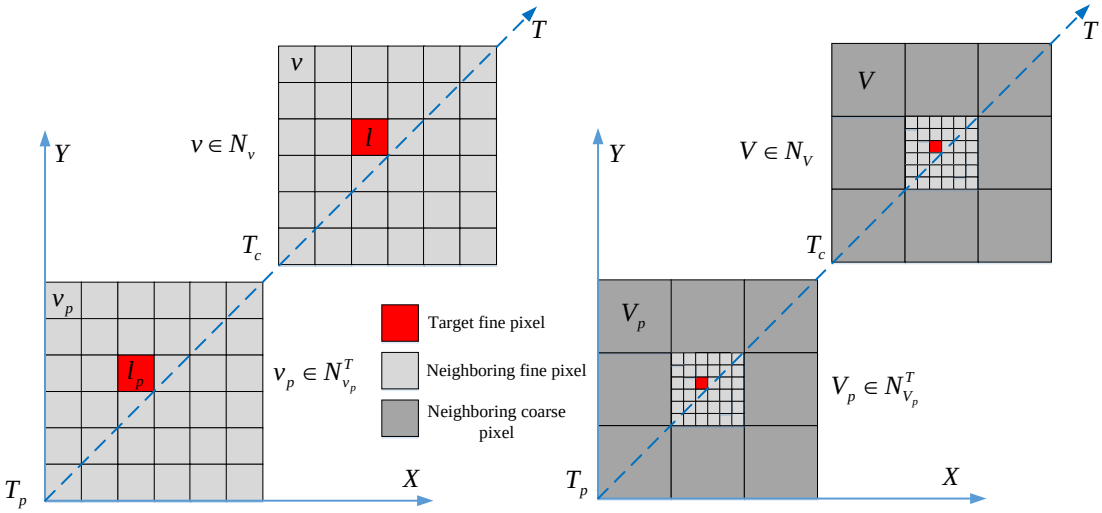

Figure 2. Multiscale spatial-temporal neighbor system of the proposed MSSTSRM approach. (a) Sub-pixel neighbor system with window size 6; (b) Coarse-pixel neighbor system with window size 3. Axes X and Y are spatial coordinates, and T is the temporal index.

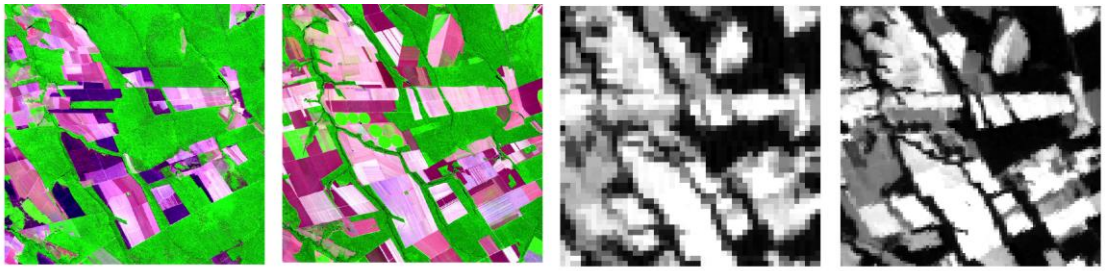

Figure 3. Original Landsat subarea images and multiscale MODIS band images for the first study site. (a) Previous false color Landsat-7 ETM+ image (RGB: 5-4-3); (b) Current reference false color Landsat-8 OLI image (RGB:6-5-4); (c) Current MOD09GA band 7 image; (d) Current MOD09GQ band 1 image.

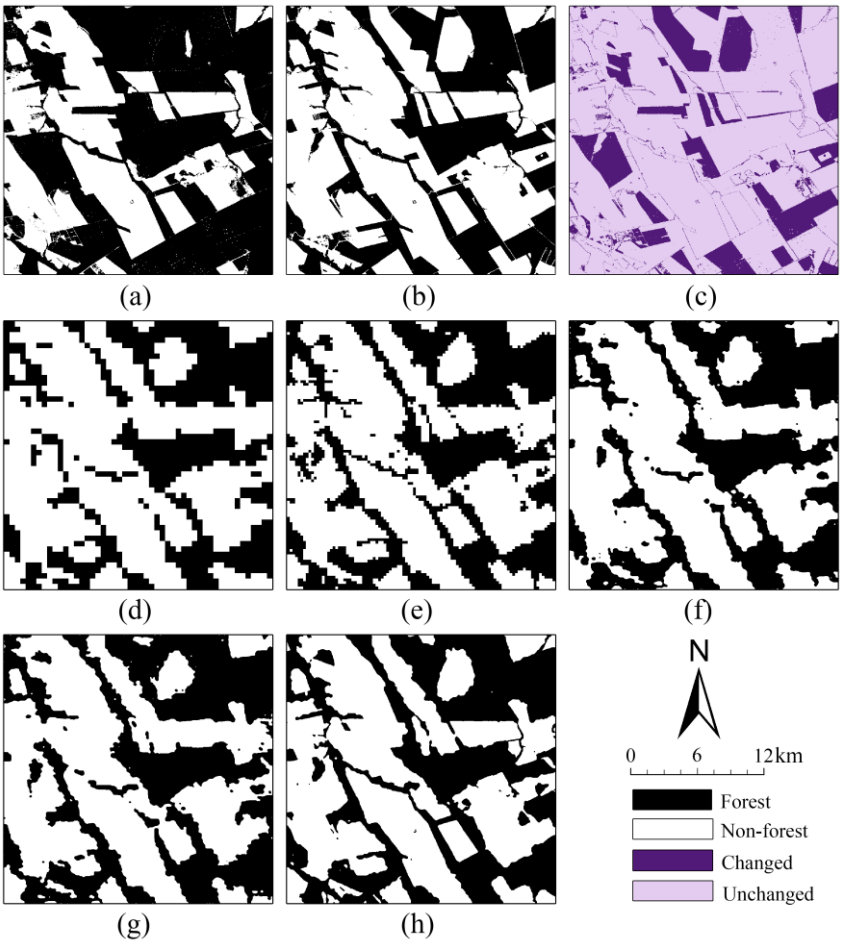

Figure 4. Forest maps obtained by different classification methods for the first study site. (a) Previous fine spatial resolution forest map; (b) Current reference fine spatial resolution forest; (c) Forest cover change map between (a) and (b); (d) Forest map generated by SVM with MOD09GA image; (e) Forest map generated by SVM with MOD09GQ image; (f) Forest map generated by MRFSRM with MOD09GA image; (g) Forest map generated by STMRF with MOD09GA image; (h) Forest map generated by MSSTSRM with both MOD09GA and MOD09GQ images.

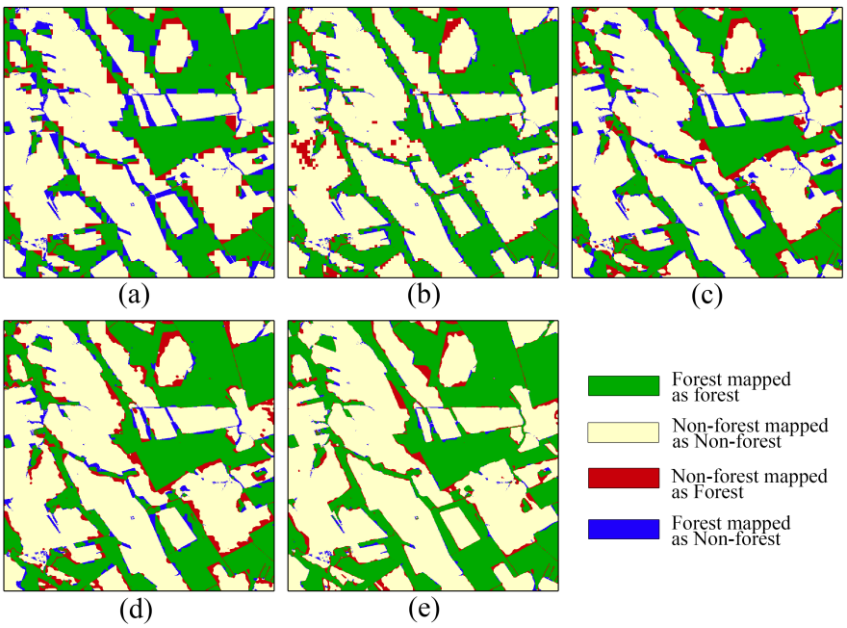

Figure 5. Forest error maps for the results of different methods by comparing with the reference forest map of the first study site. (a) Forest error map of the SVM480 result; (b) Forest error map of the SVM240 result; (c) Forest error map of the MRFSRM result; (d) Forest error map of the STMSRF result; (e) Forest error map of the MSSTSRM result.

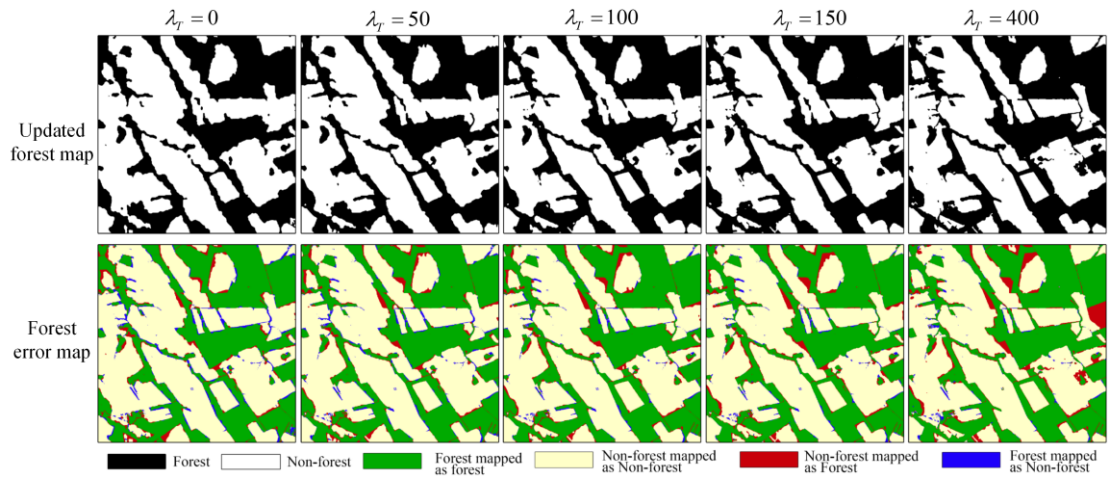

Figure 6. Updated forest maps and corresponding forest error maps generated by the proposed MSSTSRM approach with different values of trade-off parameter λ_T values (where λ_S is 80 for all the cases).

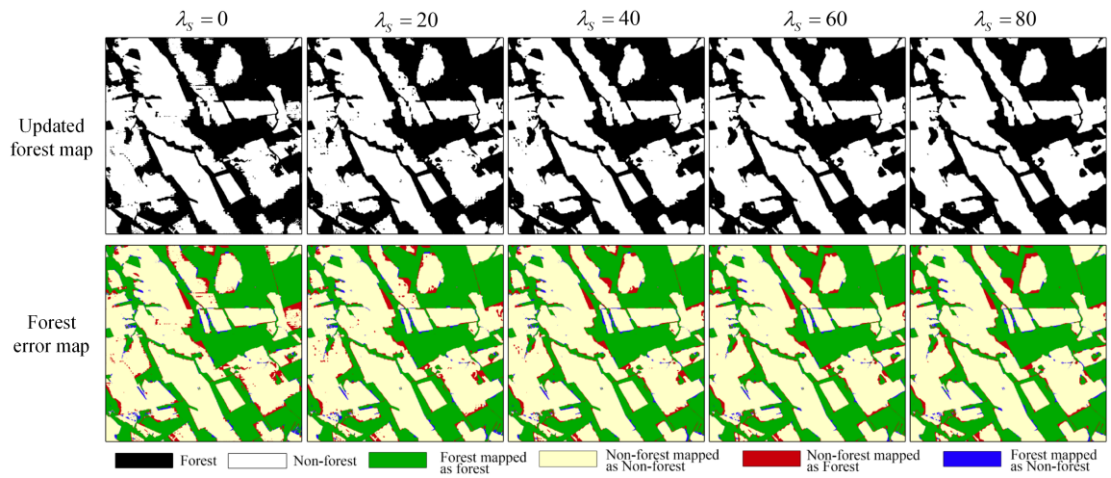

Figure 7. Updated forest maps and corresponding forest error maps generated by the proposed MSSTSRM approach with different values of trade-off parameter λ_S values (where λ_T is 140 for all the cases).

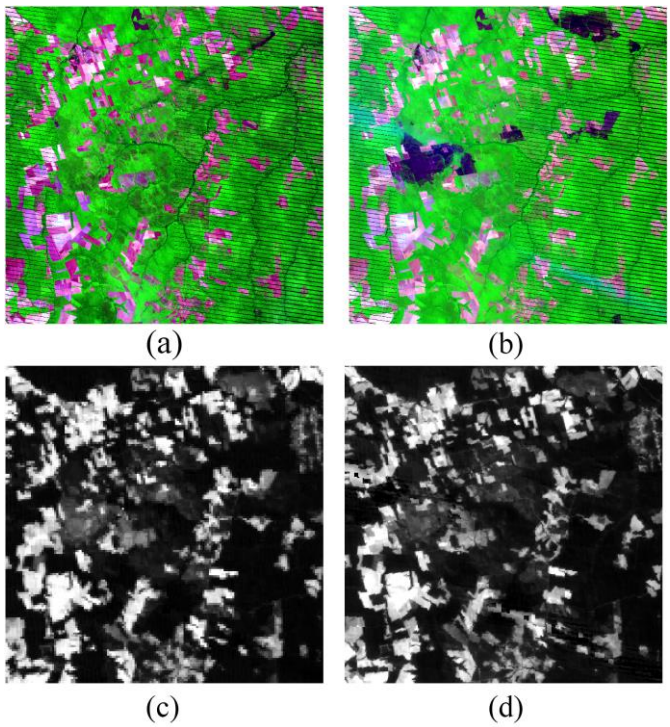

Figure 8. Original subarea Landsat images and multiscale MODIS band images for the second study site. (a) Previous false color Landsat-7 ETM+ image (RGB: 5-4-3); (b) Current false color Landsat-7 ETM+ image (RGB:5-4-3); (c) Current MOD09GA band 7 image; (d) Current MOD09GQ band 1 image.

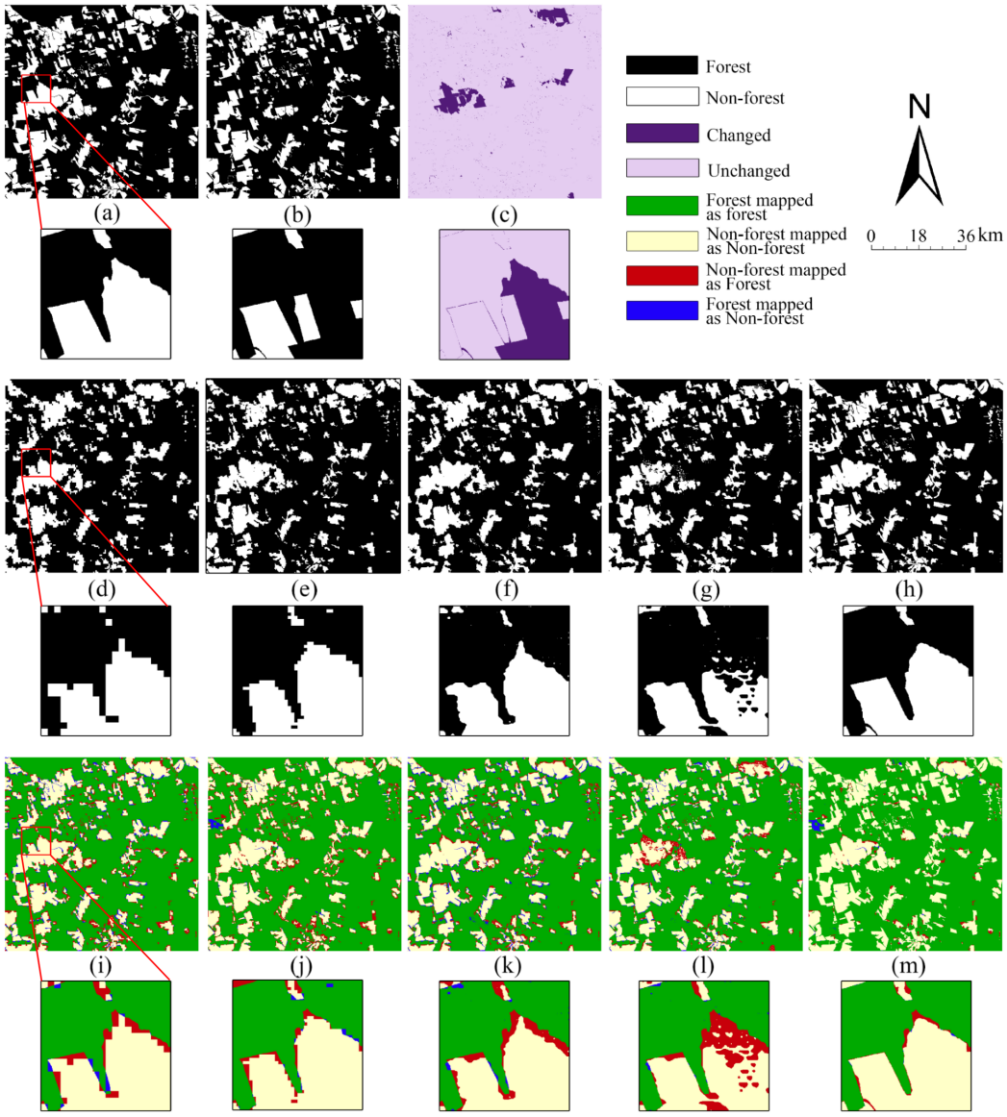

Figure 9. Forest maps and corresponding forest error maps obtained by different classification methods for the second study site. (a) Previous fine spatial resolution forest map; (b) Current reference fine spatial resolution forest map; (c) Forest cover change map between (a) and (b); (d) and (i) Forest map and forest error map generated by SVM with MOD09GA image; (e) and (j) Forest map and forest error map generated by SVM with MOD09GQ image; (f) and (k) Forest map and forest error map generated by MRFSRM with MOD09GA image; (g) and (l) Forest map and forest error map generated by STMRF with MOD09GA image; (h) and (m) Forest map and forest error map generated by MSSTSRM with both MOD09GA and MOD09GQ images.

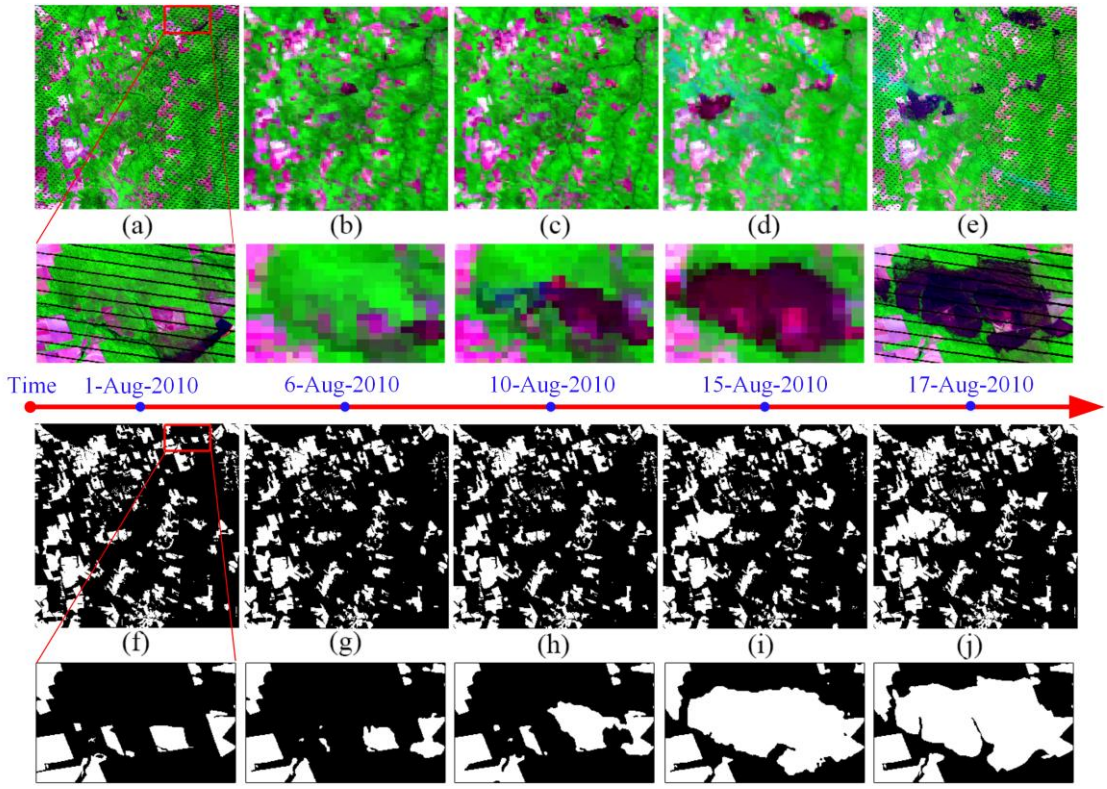

Figure 10. Time-series Landsat and MODIS multispectral images and forest maps. (a) Landsat-7 ETM+ false color image acquired on 01 August 2010 (RGB: 5-4-3); (b)-(d) MOD09GA false color images acquired on 06, 10 and 15 August 2010 respectively (RGB: 7-2-1); (e) Landsat-7 ETM+ false color image acquired on 17 August 2010 (RGB: 5-4-3); (f) and (j) 30 m forest maps extracted from Landsat-7 ETM+ images acquired on 01 and 17 August 2010; (g)-(i) 30 m forest maps updated from MOD09GA and MOD09GQ images acquired on 06, 10 and 15 August 2010 by using the proposed MSSTSRM.