Spatial-temporal Super-resolution Land Cover Mapping with a Local Spatial-temporal Dependence Model

Xiaodong Li, Feng Ling, Giles M. Foody, *Fellow, IEEE*, Yong Ge, *Member, IEEE*, Yihang Zhang,

Lihui Wang, Lingfei Shi, Xinyan Li, Yun Du

Abstract—The mixed pixel problem is common in remote sensing. A soft classification can generate land cover class fraction images that illustrate the areal proportions of the various land cover classes within pixels. The spatial distribution of land cover classes within each mixed pixel is, however, not represented. Super-resolution land cover mapping (SRM) is a technique to predict the spatial distribution of land cover classes within the mixed pixel using fraction images as input. Spatial-temporal SRM (STSRM) extends the basic SRM to include a temporal dimension by using a finer-spatial resolution land cover map that pre- or post-dates the image acquisition time as ancillary data. Traditional STSRM methods often use one land cover map as the constraint, but neglect the majority of available land cover maps acquired at different dates and of the same scene in reconstructing a full state trajectory of land cover changes when applying STSRM to time series data. In addition, the STSRM methods define the temporal dependence globally, and neglect the spatial variation of land cover temporal dependence intensity within images. A novel local STSRM (LSTSRM) is proposed in this paper. LSTSRM incorporates more than one available land cover map to constrain the solution, and develops a local temporal dependence model, in which the temporal dependence intensity may vary spatially. The results show that LSTSRM can eliminate speckle-like artifacts and reconstruct the spatial patterns of land cover patches in the resulting maps, and increase the overall accuracy compared with other STSRM methods.

Index Terms—Super-resolution mapping, image series, spatial dependence, temporal dependence

I. INTRODUCTION

LAND cover and its dynamic play a major role in global change. Understanding the distribution and dynamics of land cover is essential to better understand the Earth's processes. Remotely sensed images are the main data source in mapping land cover and monitoring land cover changes at different spatial resolutions. However, land parcels managed at a local scale are often smaller than the resolution of satellite data, in which one pixel often represents composite spectral responses from multiple land cover types. Hard classification methods cannot accurately map the mixed pixels, because they assign a mixed pixel to a single land cover class. Soft classification can generate land cover class fraction images that represent the areal proportions of different land cover classes within a pixel. The output of a soft classification is a number of fraction images equal to the number of land cover classes. However, the spatial distribution of land cover classes within the mixed pixel is still unknown. By dividing the pixel into numerous sub-pixels and assuming the sub-pixels are pure, super-resolution mapping (SRM) can assign the class fractions spatially to sub-pixels [1]. SRM can be viewed as the post processing of soft classification that predicts the spatial distribution of land cover classes at the sub-pixel scale. The fraction images output from a soft classification are inputted to an SRM to produce a land cover map with a finer spatial resolution than the original remotely sensed image.

In general, SRM is an ill-posed problem and the result unavoidable contains uncertainty. In order to decrease the uncertainty, various ancillary data such as panchromatic band image [2], vector boundaries [3] and LIDAR [4] have been used in the SRM models to provide more information. However, these SRM methods require that the acquisition dates of these remotely sensed image and ancillary data should be the same or closer, which restricts the use of these data in SRM. Besides the aforementioned ancillary data, land cover maps with finer spatial resolution than the remotely sensed images obtained at different dates are an alternative ancillary data for SRM. The land cover maps can be generated from remotely sensed images obtained from various platforms. Given that land cover map and remotely sensed images can be acquired at different times and considering that a great number of historical land cover maps which cover almost the entire earth may be available, SRM using these land cover maps is very promising.

X. Li, F. Ling, Y. Zhang, L. Wang, L. Shi, X. Li and Y. Du are with the Key laboratory of Monitoring and Estimate for Environment and Disaster of Hubei province, Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan 430077, PR China, and with the Sino-Africa Joint Research Centre, Chinese Academy of Sciences, Wuhan 430074, China. X. Li (the first author) is also with School of Geography, University of Nottingham, University Park, Nottingham NG7 2RD, U.K.

G. M. Foody is with School of Geography, University of Nottingham, University Park, Nottingham NG7 2RD, U.K.

Y. Ge is with State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences & Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China.

The spatial-temporal super-resolution mapping (STSRM) first proposed by Ling et al. [5] is an approach that incorporates a finer-spatial resolution land cover map that pre- or post-dates the remotely sensed image acquisition time as ancillary data. Hereafter, we refer the input remotely sensed image has a coarse spatial resolution or coarse resolution, and the input and output land cover maps have fine spatial resolution or fine resolution in STSRM. Note that the terms 'coarse' and 'fine' do not mean the absolution spatial resolution of the data, but the relative spatial resolution of the image and the land cover map. STSRM is suited to monitor land cover change [6], and has been applied in many fields such as forest mapping [7], change detection [8], and land cover map updating [9].

Various STSRM models have been proposed and can be generally categorized into two groups. The first group of STSRM models are based on the change detection analysis (CDA_STSRM). In CDA_STSRM models, the remotely sensed image pixels are unmixed to coarse resolution land cover fraction images from soft classification, and the fine resolution land cover map is spatially degraded to coarse resolution land cover fraction images. By comparison of these two kinds of fraction images, if the land cover fraction of a class is unchanged in a coarse resolution pixel, the fine resolution pixels that belong to that class in the coarse resolution pixel are assumed to have unchanged class label. Otherwise, the fine resolution pixels that belong to that class in the coarse resolution pixel are assumed to have changed class label, and the label can be determined using various algorithms such as the pixel-swapping algorithm [5], the Hopfield neural network [10], the maximum a posteriori method [9, 11], the learning based model [12], the interpolation based model [13], the swarm intelligence theory [14], the adaptive cellular automata [15], and the artificial neural network [16]. Since error in fraction images is usually unavoidable, a threshold used to distinguish unchanged and changed fractions in each coarse resolution pixel is necessary to be incorporated in CDA_STSRM, and the results depend greatly on the fraction change detection threshold value used. However, the fraction error is usually spatially variable in different pixels and for different classes. The accurately estimation of the threshold in CDA_STSRM is very difficult [10, 11].

The second group of STSRM models are based on spatial-temporal dependence (STD_STSRM). STD_STSRM assumes that fine resolution pixels are spatially dependent with their spatially closest fine resolution pixels, and are temporally dependent with the corresponding fine resolution pixels in images that pre- or post-date the image under analysis. Xu and Huang [17] proposed the spatial-temporal pixel swapping algorithm (STPSA) model that is applied to bi-temporal data, and Wang et al. [18] extended STPSA to multi-temporal data. Li et al. [7] proposed a Markov Random Field based model to map forest cover from Moderate Resolution Imaging Spectroradiometer (MODIS), and Zhang et al. [19] extended this model by using both the 240 m and 480 m spatial resolution MODIS bands. Generally, STD_STSRM predicts all fine pixel labels based on the spatial-temporal dependence model, and removes the need for a change detection threshold.

Considering this advantage, this paper is focused on STD_STSRM.

All STD_STSRM methods predict the fine resolution land cover map based on the *a-priori* spatial distribution and temporal transition information about land covers. The *a-priori* information in STD_STSRM includes the *a-priori* spatial information that is used to predict the land cover spatial patterns at the fine resolution pixel scale and *a-priori* temporal information that is used to model the temporal transitions between the class labels in the predicted and the input pre- or post-dated land cover maps. The *a-priori* spatial models have been studied in SRM researches including the spatial dependent model [20-23], the direct mapping model [24], the geostatistical model [25, 26], the multi-point simulation based model [27], the learning based model [28, 29], the adaptive model [30], and the linear spatial distribution model [31, 32]. In these models, [20-24] are suitable for predicting spatial patterns of patches that are larger than the coarse resolution pixel, [25-27] are suitable for predicting spatial patterns of patches that are smaller than the coarse resolution pixel, [31, 32] are suitable for linear patch, and [28-30] are suitable for patches with different spatial patterns, respectively. These *a-priori* spatial models used in SRM can be directly applied in STD_STSRM. In contrast, the study on the *a-priori* temporal information in STD_STSRM is very rare. Challenges in STD_STSRM remain, especially if seeking to use a time series of images.

First, in many cases, more than one land cover map of the same scene is available in the area of interest. Since these land cover maps may record the land covers at different dates, incorporating as much fine resolution land cover maps as possible is very useful in accurately mapping land cover trajectories with STSRM. There are three scenarios for the acquisition time of the fine resolution maps and coarse resolution image in STSRM. The first case is that fine resolution maps which pre- and post-date the coarse resolution image are available, the second case is that only a fine resolution map which pre-dates the coarse resolution image is available, and the third case is that only a fine resolution map which post-dates the coarse resolution image is available. For the first case, two available fine resolution maps can be used as the *a-priori* temporal information to constrain STSRM. By contrast, for the second and third cases, only one fine resolution map is available as the *a-priori* temporal information. Existing STSRM methods, including both STD_STSRM and CDA_STSRM models, are focused on the second and third cases in which only one fine resolution map is available [5, 9-18, 33-35], but fail to explore the first case. A thorough study on the three cases should be developed in order that the entire land cover change trajectory from remotely sensed image series can be extracted.

Second, the nature of the temporal dependence is often spatially variable. In STD_STSRM models, the temporal dependence is used to link the fine resolution pixels in different date. All existing STD_STSRM models consider the temporal dependence globally, and the intensity of the temporal dependence is, therefore fixed across the entire image. The
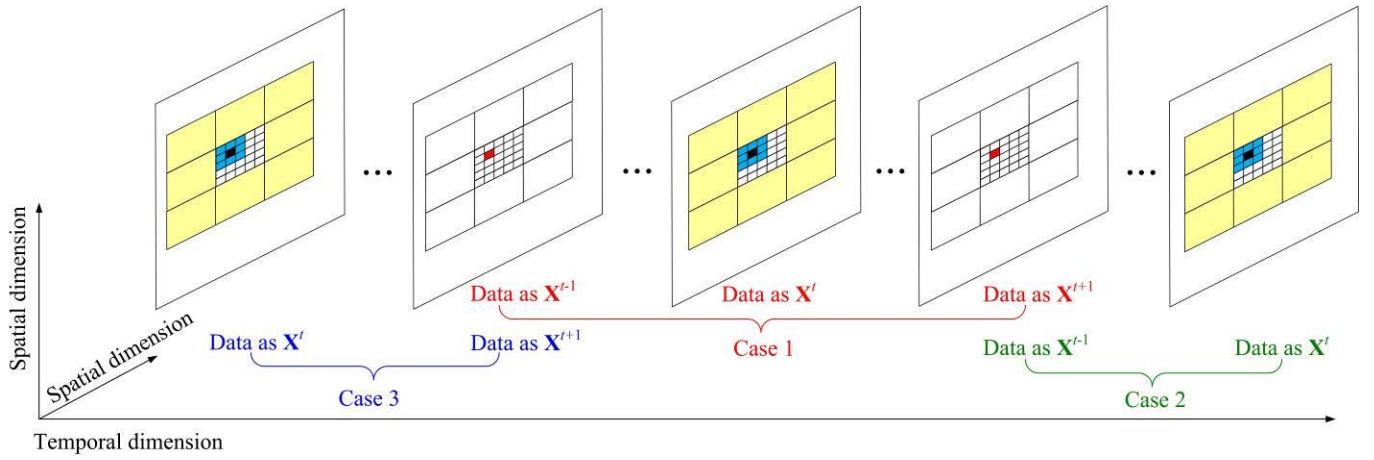
Fig. 1. The spatial and temporal neighborhoods for a fine resolution pixel. Case 1: Both $\mathbf{X}^{t-1}$ and $\mathbf{X}^{t+1}$ are available. Case 2: $\mathbf{X}^{t+1}$ is available. Case 3: $\mathbf{X}^{t+1}$ is available. The fine resolution pixel highlighted in black in $\mathbf{X}^t$ is the target pixel. The fine resolution pixels highlighted in blue and the coarse resolution pixels highlighted in yellow in $\mathbf{X}^t$ are the spatial neighborhood pixels. The fine resolution pixels highlighted in red in $\mathbf{X}^{t-1}$ and $\mathbf{X}^{t+1}$ are the temporal neighborhood pixels.

global temporal dependence model is used for its simplicity but may not be sufficient to model the spatial and temporal variation that may exist [7]. Generally, the land cover temporal dependence is related to the land cover transition probability. For instance, in the case of forest change, the transition probability from forest to nonforest is usually spatially variable depending on many physical and economic factors such as accessibility and land value [36]. Assuming the transition probability from forest to nonforest is invariant in the area covered by the entire image using the global temporal dependence model is not plausible. It is desirable to change the global temporal dependence to the local scale and to accommodate the spatial variation of its intensity.

Third, land cover fraction errors caused by the soft classification procedure have to be handled in STD_STSRM. In general, the land cover fraction images that are spectrally unmixed from remotely sensed images are inputted in STD_STSRM. As soft classification continues to be an open problem, fraction images errors are often unavoidable in practice. The errors might lead a decrease in the accuracy for existing STSRM methods which constrain the result land cover map in a way that the class fractions within each coarse resolution pixel should be unchanged between the input coarse resolution fraction images and the output fine resolution land cover maps [10]. STD_STSRM should be developed not to strictly preserve the class fractions from the input coarse resolution class fraction images into the result fine resolution land cover map and act to eliminate fraction errors caused by soft classification.

In this paper, a novel Local STD_STSRM (LSTSRM) model is proposed to generate fine resolution land cover maps from a series of coarse resolution class fraction images and a few fine resolution land cover maps. Unlike traditional STSRM models that consider only one fine resolution land cover map, LSTSRM can use both fine resolution maps that pre- and post-date the coarse resolution image to fully explore information in all available datasets and constrain the STSRM problem. In the proposed model, the temporal dependence intensity may vary from pixel to pixel at the fine resolution

scale. In addition, the proposed model is developed not to strictly preserve the class fractions from the input class fraction images into the result land cover map, in order to eliminate fraction errors caused by the soft classification procedure. The remainder of this paper is organized as follows. Section II introduces the LSTSRM method. Section III examines the performance of LSTSRM using synthetic data experiment, real Sentinel-2 images experiment and real MODIS images experiment. Section IV discusses the influencing factors of the proposed method. Section V concludes this paper.

II. METHOD

A. The LSTSRM framework

LSTSRM inputs coarse resolution class fraction images $\mathbf{F}^t$ at the observation time t ($t=1,2,\cdots,T$), as well as the fine resolution land cover maps $\mathbf{X}^{t-1}$ and/or $\mathbf{X}^{t+1}$ that covers the same geographical region but obtained at times that pre- and/or post-date $\mathbf{F}^t$, as input, and outputs the fine resolution land cover map $\mathbf{X}^t$ at time of t. The coarse resolution fraction images $\mathbf{F}^t$ can be produced from remotely sensed image using various soft classification procedures. The fine resolution maps $\mathbf{X}^{t-1}$ and $\mathbf{X}^{t+1}$ can often be produced from fine resolution remotely sensed image by, for example, classification or manual digitization. Although a series of fine resolution land cover maps may be available and can be incorporated in LSTSRM, only those which are acquired temporally closest to and pre- or post-date $\mathbf{Y}^t$ are selected as $\mathbf{X}^{t-1}$ and/or $\mathbf{X}^{t+1}$. This is because the land cover datasets are temporally more dependent if they are obtained at temporally closer time [6, 18]. $\mathbf{F}^t$ contains $I \times J \times C$ pixels ($I \times J$ is the number of coarse resolution pixels and C is the number of land cover classes). $\mathbf{X}^t$, $\mathbf{X}^{t-1}$ and $\mathbf{X}^{t+1}$ each contains $I \times s \times J \times s$ pixels, where s is the scale factor and each coarse resolution pixel contains $s \times s$ fine resolution pixels. Each pixel in $\mathbf{X}^t$, $\mathbf{X}^{t-1}$ and $\mathbf{X}^{t+1}$ has a land cover class label in C. In LSTSRM, the three scenarios, that is, both $\mathbf{X}^{t-1}$ and $\mathbf{X}^{t+1}$ are available (case 1), only $\mathbf{X}^{t-1}$ is available (case 2), and only $\mathbf{X}^{t+1}$ is available (case 3), are considered (Fig. 1).

Based on Bayesian theory, the optimal $\mathbf{X}^t$ can be expressed as:

TABLE I
LIST OF THE IMPORTANT VARIABLES

Variable name	Definition
$\mathbf{F}^t$	The input coarse resolution class fraction images at time of t;
$\mathbf{X}^t$	The predicted fine resolution land cover map at time of t;
$\mathbf{X}^{t-1}$	The input fine resolution land cover map at time of t-1; $\mathbf{X}^{t-1}$ pre-dates $\mathbf{X}^t$;
$\mathbf{X}^{t+1}$	The input fine resolution land cover map at time of t+1; $\mathbf{X}^{t+1}$ post-dates $\mathbf{X}^t$;
p_{ijk}^t	The kth fine resolution pixel in coarse resolution pixel (i,j) in $\mathbf{X}^t$;
$D_{intra}^S(c(p_{ijk}^t)=c)$	The intra-pixel spatial dependence of fine resolution pixel p_{ijk}^t when it has a label of c ($c=1,\cdots,C$, and C is the total number of classes in the image).
$D_{inter}^S(c(p_{ijk}^t)=c)$	The inter-pixel spatial dependence of fine resolution pixel p_{ijk}^t when it has a label of c ($c=1,\cdots,C$);
$D_G^T(c(p_{ijk}^t)=c)$	The global temporal dependence intensity of fine resolution pixel p_{ijk}^t if p_{ijk}^t belongs to the cth class;
$D_L^T(c(p_{ijk}^t)=c)$	The local adjust factor of fine resolution pixel p_{ijk}^t if p_{ijk}^t belongs to the cth class;
MTSRM	Mono-temporal super-resolution mapping;
STSRM	Spatial-temporal super-resolution mapping;
LSTSRM	Local temporal dependence model based STSRM;
GSTSRM	Global temporal dependence model based STSRM.

$$\tilde{\mathbf{X}}^t = \arg\max\left\{P\left(\mathbf{X}^t\left|\mathbf{F}^t,\mathbf{X}^{t-1},\mathbf{X}^{t+1}\right.\right)\right\} \tag{1}$$

where $P(\mathbf{X}^t|\mathbf{F}^t,\mathbf{X}^{t-1},\mathbf{X}^{t+1})$ is the posterior probability of $\mathbf{X}^t$, given $\mathbf{F}^t$, $\mathbf{X}^{t-1}$ and/or $\mathbf{X}^{t+1}$. The Markov random field can model contextual information by characterizing the local statistical dependence among pixels in terms of conditional prior distribution [37]. The Markov random field can simplify the global model in (1) to a model of the local image properties, and largely reduces the model complexity to make the maximum a posteriori (MAP) model solvable. The optimal $\mathbf{X}^t$, given $\mathbf{F}^t$, $\mathbf{X}^{t-1}$ and/or $\mathbf{X}^{t+1}$, can be formulated by applying the MAP rule, i.e., by solving the maximization problem:

$$\begin{aligned}\tilde{\mathbf{X}}^t &= \arg\max\left\{P\left(\mathbf{X}^t\left|\mathbf{F}^t,\mathbf{X}^{t-1},\mathbf{X}^{t+1}\right.\right)\right\}\\ &= \arg\max\left\{-\frac{1}{Z}\exp\left[-U\left(\mathbf{X}^t\left|\mathbf{F}^t,\mathbf{X}^{t-1},\mathbf{X}^{t+1}\right.\right)\right]\right\}\end{aligned} \tag{2}$$

where $U(\mathbf{X}^t|\mathbf{F}^t,\mathbf{X}^{t-1},\mathbf{X}^{t+1})$ is the posterior energy function of $\mathbf{X}^t$ and Z is a normalizing constant. Based on the Markov random field approach, the searching of the optimal $\mathbf{X}^t$ 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. $U(\mathbf{X}^t|\mathbf{F}^t,\mathbf{X}^{t-1},\mathbf{X}^{t+1})$ is calculated as:

$$U\left(\mathbf{X}^t\left|\mathbf{F}^t,\mathbf{X}^{t-1},\mathbf{X}^{t+1}\right.\right) = U_S(\mathbf{X}^t) + U_T(\mathbf{X}^t\left|\mathbf{X}^{t-1},\mathbf{X}^{t+1}\right.) + U_F(\mathbf{F}^t\left|\mathbf{X}^t\right.) \tag{3}$$

where $U_S(\mathbf{X}^t)$ and $U_T(\mathbf{X}^t|\mathbf{X}^{t-1},\mathbf{X}^{t+1})$ are the spatial and temporal constraint functions, and $U_F(\mathbf{F}^t|\mathbf{X}^t)$ is class fraction constraint function that represents the inconsistency between $\mathbf{F}^t$ and $\mathbf{X}^t$.

B. The Spatial and Temporal Constraint Functions

The LSTSRM spatial and temporal constraint functions are used to incorporate land cover *a-priori* spatial and temporal dependence model according to the spatial-temporal neighborhood system (Fig. 1). Each fine resolution pixel in $\mathbf{X}^t$ highlighted in black is spatially dependent on its spatial neighborhood fine resolution pixels marked in blue and spatial neighborhood coarse resolution pixels highlighted in yellow in $\mathbf{X}^t$, and is temporally dependent on its temporal neighborhood fine resolution pixels highlighted in red in $\mathbf{X}^{t-1}$ and $\mathbf{X}^{t+1}$.

1) Spatial Constraint Function

The LSTSRM spatial constraint function is based on the spatial dependence principle which is the tendency of spatially proximate observations of a given property to be more similar than distant observations. It assumes that a fine resolution pixel and its neighboring fine resolution pixels have high probabilities to be labeled with the same class. At present, there are two methods to describe the spatial dependence of pixels, namely intra-pixel spatial dependence and inter-pixel spatial dependence, which are used to represent the spatial dependence within and between image pixels, respectively [20, 38-40]. Assume p_{ijk}^t is the kth ($k=1,\ldots,s^2$) fine resolution pixel in coarse resolution pixel (i,j) ($i=1,\ldots,I, j=1,\ldots,J$) in $\mathbf{X}^t$. $D_{intra}^S(c(p_{ijk}^t)=c)$ is the intra-pixel spatial dependence of fine resolution pixel p_{ijk}^t when it has a label of c ($c=1,\ldots,C$), and $D_{inter}^S(c(p_{ijk}^t)=c)$ is the inter-pixel spatial dependence of fine resolution pixel p_{ijk}^t when it has a label of c. The spatial energy function $U_S(\mathbf{X}^t)$ can be written as:

$$U_S(\mathbf{X}^t) = \sum_{i=1}^I\sum_{j=1}^J\sum_{k=1}^{s^2}\sum_{c=1}^C(-1)\cdot\left(\alpha_1\cdot D_{intra}^S\left(c\left(p_{ijk}^t\right)=c\right)+\alpha_2\cdot D_{inter}^S\left(c\left(p_{ijk}^t\right)=c\right)\right) \tag{4}$$

where α_1 and α_2 define the weight of the intra- and inter-pixel spatial dependence values. In Eq. (4), -1 is multiplied because the LSTSRM seeks the minimum value as the optimal solution. The determination of intra-pixel spatial dependence and inter-pixel spatial dependence is explained as follows.

1.1) Intra-pixel Spatial Dependence:

The intra-pixel spatial dependence is computed at the sub-pixel/sub-pixel (fine resolution pixel/fine resolution pixel) scale, meaning the spatial dependence of a sub-pixel (highlighted in black in Fig. 1) is determined by its neighborhood same-class sub-pixels (highlighted in blue in Fig. 1). The intra-pixel spatial dependence $D_{intra}^S(c(p_{ijk}^t)=c)$ is determined by the pixel labels of its neighboring eight fine resolution pixels in the neighborhood system:

$$D_{intra}^S\left(c\left(p_{ijk}^t\right)=c\right) = \sum_{l\in N(p_{ijk}^t)}\delta\left(c(p_l^t),c\right)\Big/8 \tag{5}$$

where $N(p_{ijk}^t)$ is the fine resolution spatial neighborhood, and p_l^t is a neighborhood fine resolution pixel in $N(p_{ijk}^t)$. $c(p_{ijk}^t)$ and $c(p_l^t)$ are the land cover class labels for fine resolution pixels p_{ijk}^t and p_l^t. $\delta(c(p_l^t),c)$ equals 1 if $c(p_l^t)$ and c are the same and 0 otherwise.

1.2) Inter-pixel Spatial Dependence:

The inter-pixel spatial dependence is calculated at the sub-pixel/pixel scale, meaning that the inter-pixel spatial
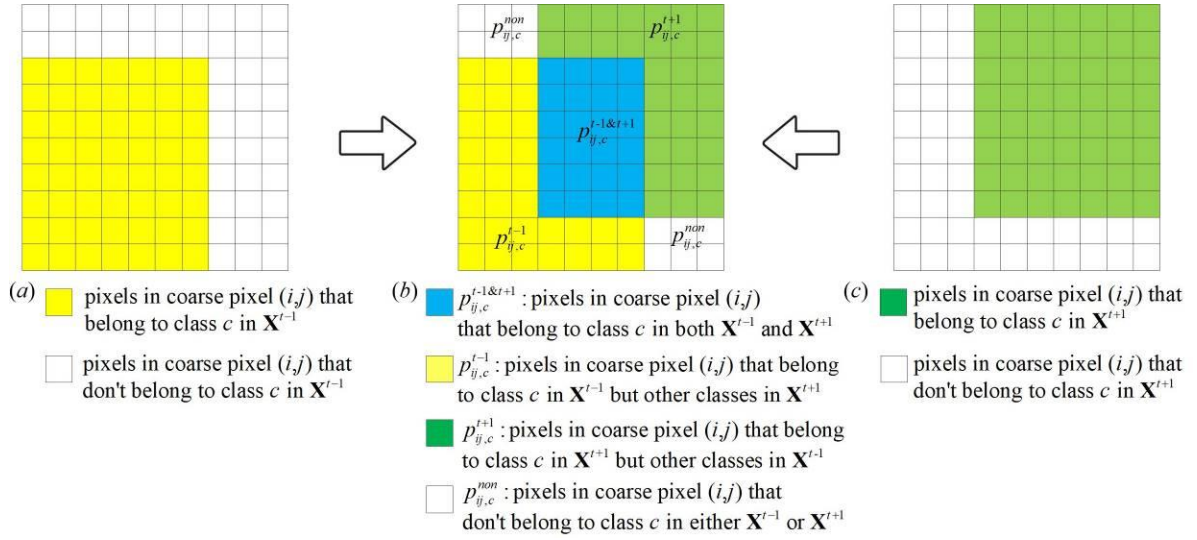
Fig. 2. An illustration of building the local adjust factor for the temporal dependence according to the fine resolution pixels labels of the cth class in pixel (i,j). (a) One coarse resolution pixel in $\mathbf{X}^{t-1}$. (b) One coarse resolution pixel in $\mathbf{X}^{t}$. (c) One coarse resolution pixel in $\mathbf{X}^{t+1}$. Each coarse resolution pixel contains 10×10 fine resolution pixels.

dependence of a sub-pixel (highlighted in black in Fig. 1) is determined by the same-class fine resolution pixels in the neighboring coarse resolution pixel (highlighted in yellow in Fig. 1) [39]. Spatial interpolation algorithms such as inverse distance weighted function or Kriging can be used to represent the relationship between sub-pixel and pixels [39, 41]. By spatially interpolating the neighborhood coarse pixel class fractions of each class to the sub-pixel scale, the inter-pixel spatial dependence $D_{inter}^{S}(c(p_{ijk}^{t})=c)$ is defined as:

$$D_{inter}^{S}\left(c\left(p_{ijk}^{t}\right)=c\right)=f_{c}(p_{ijk}^{t}) \quad (6)$$

where $f_{c}(p_{ijk}^{t})$ is the spatially interpolated fractions of the cth class at sub-pixel p_{ijk}^{t}. The value of $f_{c}(p_{ijk}^{t})$ is related to the cth class fractions in the neighborhood coarse pixels, the distance between p_{ijk}^{t} and the neighborhood coarse pixels, and the spatially interpolation method being used [42].

2) The Spatial and Temporal Constraint Functions

The LSTSRM fine pixel temporal dependence intensity is defined according to $D_{G}^{T}(c(p_{ijk}^{t})=c)$ and $D_{L}^{T}(c(p_{ijk}^{t})=c)$, where $D_{G}^{T}(c(p_{ijk}^{t})=c)$ is the global temporal dependence intensity of fine resolution pixel p_{ijk}^{t} if p_{ijk}^{t} belongs to the cth class, and $D_{L}^{T}(c(p_{ijk}^{t})=c)$ is the local adjust factor of fine resolution pixel p_{ijk}^{t} if p_{ijk}^{t} belongs to the cth class:

$$U_{T}(\mathbf{X}^{t}\left|\mathbf{X}^{t-1},\mathbf{X}^{t+1}\right)$$
$$=\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{s^{2}}\sum_{c=1}^{C}(-1)\cdot\beta\cdot D_{G}^{T}\left(c\left(p_{ijk}^{t}\right)=c\right)\cdot D_{L}^{T}\left(c\left(p_{ijk}^{t}\right)=c\right)\cdot (7)$$

where β is the temporal constraint function weight parameter. The global temporal dependence intensity $D_{G}^{T}(c(p_{ijk}^{t})=c)$ is assigned to 1 if the kth fine resolution pixel in coarse resolution pixel (i,j) belongs to the cth class in $\mathbf{X}^{t-1}$ or $\mathbf{X}^{t+1}$, and is assigned to 0 otherwise. Therefore, the fine resolution pixels in $\mathbf{X}^{t}$ and $\mathbf{X}^{t-1}$ (or $\mathbf{X}^{t+1}$) are temporally dependent if they have the same class label, and are temporally independent if they have different class labels. The local adjust factor $D_{L}^{T}(c(p_{ijk}^{t})=c)$ depends not only on the fine resolution class labels in $\mathbf{X}^{t-1}$

and/or $\mathbf{X}^{t+1}$, but also on the coarse resolution class fractions at times of $t-1$, t and $t+1$. The calculation of the local adjust factor $D_{L}^{T}(c(p_{ijk}^{t})=c)$ according to different input data are explained as follows.

2.1) Both $\mathbf{X}^{t-1}$ and $\mathbf{X}^{t+1}$ are Available:

Before the calculation of the local adjust factor $D_{L}^{T}(c(p_{ijk}^{t})=c)$, the fine resolution pixels in coarse pixel (i,j) in $\mathbf{X}^{t}$ are grouped into different sets according to the spatial distribution of fine resolutions of the cth class in $\mathbf{X}^{t-1}$ and $\mathbf{X}^{t+1}$, assuming different set of fine pixels may have different temporal dependence and different local adjust factors. An example on grouping different fine resolution pixel sets is shown in Fig. 2. Fig. 2 (a) and (c) shows fine pixels that belong or not belong to c^{th} class in the coarse pixel (i,j) in $\mathbf{X}^{t-1}$ and $\mathbf{X}^{t+1}$. By comparing Fig. 2 (a) and (c), four sets of fine resolution pixels in coarse resolution pixel (i,j), which are $p_{ij,c}^{t-1\&t+1}$, $p_{ij,c}^{t-1}$, $p_{ij,c}^{t+1}$ and $p_{ij,c}^{non}$, are defined in Fig. 2(b). The detailed definitions are given in Fig. 2.

Let $f(p_{ij,c}^{t})$ be the fractions of the cth class in coarse resolution pixel (i,j) in $\mathbf{F}^{t}$ that is unmixed using soft classification. Let $f(p_{ij,c}^{t-1\&t+1})$, $f(p_{ij,c}^{t-1})$ and $f(p_{ij,c}^{t+1})$ be the fractions of the cth class in coarse resolution pixel (i,j), which are calculated by dividing the number of pixels in $p_{ij,c}^{t-1\&t+1}$, $p_{ij,c}^{t-1}$ and $p_{ij,c}^{t+1}$ in coarse resolution pixel (i,j) by s^{2} (s is the scale factor). The local adjust factor is quantified by comparing $f(p_{ij,c}^{t})$ with $f(p_{ij,c}^{t-1\&t+1})$, $f(p_{ij,c}^{t-1})$ and $f(p_{ij,c}^{t+1})$. In LSTSRM, a simple rule is used in which a fine pixel that belongs to the cth class is more probably to be in the set $p_{ij,c}^{t-1\&t+1}$ than in the sets $p_{ij,c}^{t-1}$ and $p_{ij,c}^{t+1}$. In addition, a fine resolution pixel is not likely to belong to the cth class if this pixel does not belong to the cth class in $\mathbf{X}^{t-1}$ and $\mathbf{X}^{t+1}$, and the local adjust factor for fine resolution pixels in $p_{ij,c}^{non}$ is set to 0. Based on this rule, the local adjust factor according to the following cases is calculated.

If $f(p_{ij,c}^{t})\geqslant f(p_{ij,c}^{t-1\&t+1})+f(p_{ij,c}^{t-1})+f(p_{ij,c}^{t+1})$, then the fine resolution pixels in the sets $p_{ij,c}^{t-1\&t+1}$, $p_{ij,c}^{t-1}$, $p_{ij,c}^{t+1}$ are all temporally dependent. The fine resolution pixels in $p_{ij,c}^{t-1\&t+1}$, $p_{ij,c}^{t-1}$ and $p_{ij,c}^{t+1}$ should all belong to the cth class in $\mathbf{X}^{t}$, and the corresponding local adjust

factor equals to 1, which is the maximal local adjust factor value.

If $f(p_{ij,c}^t)$ is lower than $f(p_{ij,c}^{t-1\&t+1})+f(p_{ij,c}^{t-1})+f(p_{ij,c}^{t+1})$ but higher than $f(p_{ij,c}^{t-1\&t+1})$, the fine resolution pixels in $p_{ij,c}^{t-1\&t+1}$ should belong to the cth class and the corresponding local adjust factor remains to be 1, whereas the fine resolution pixels in $p_{ij,c}^{t-1}$ and $p_{ij,c}^{t+1}$ are not definitely to belong to the cth class and the corresponding local adjust factor is lower than 1. More specifically, the probability of fine resolution pixels in $p_{ij,c}^{t-1}$ and $p_{ij,c}^{t+1}$ belonging to the cth class is proportional to the difference between $f(p_{ij,c}^t)$ and $f(p_{ij,c}^{t-1\&t+1})$ in Eqs (8-9). In addition, the probability of fine resolution pixels in $p_{ij,c}^{t-1}$ and $p_{ij,c}^{t+1}$ belonging to the cth class decreases with the time interval between $\mathbf{X}^t$ and $\mathbf{X}^{t-1}$ or between $\mathbf{X}^t$ and $\mathbf{X}^{t+1}$, assuming the temporal dependence decreases with the time interval between images. Let $\Delta t(\mathbf{X}^{t-1}, \mathbf{X}^t)$ and $\Delta t(\mathbf{X}^t, \mathbf{X}^{t+1})$ be the time interval from the acquisition time from $\mathbf{X}^{t-1}$ to $\mathbf{X}^t$ and from the acquisition time from $\mathbf{X}^t$ to $\mathbf{X}^{t+1}$, respectively. The local adjust factor for the sets $p_{ij,c}^{t-1}$ and $p_{ij,c}^{t+1}$ are calculated as

$$D_L^T\left(c\left(p_{ijk}^t\right)=c\right)=\frac{f\left(p_{ij,c}^t\right)-f\left(p_{ij,c}^{t-1\&t+1}\right)}{f\left(p_{ij,c}^{t-1}\right)+f\left(p_{ij,c}^{t+1}\right)}\cdot\left(1-\frac{\Delta t(\mathbf{X}^{t-1},\mathbf{X}^t)}{\Delta t(\mathbf{X}^{t-1},\mathbf{X}^t)+\Delta t(\mathbf{X}^t,\mathbf{X}^{t+1})}\right)$$

(8)

if p_{ijk}^t belongs to $p_{ij,c}^{t-1}$ and as

$$D_L^T\left(c\left(p_{ijk}^t\right)=c\right)=\frac{f\left(p_{ij,c}^t\right)-f\left(p_{ij,c}^{t-1\&t+1}\right)}{f\left(p_{ij,c}^{t-1}\right)+f\left(p_{ij,c}^{t+1}\right)}\cdot\left(1-\frac{\Delta t(\mathbf{X}^t,\mathbf{X}^{t+1})}{\Delta t(\mathbf{X}^{t-1},\mathbf{X}^t)+\Delta t(\mathbf{X}^t,\mathbf{X}^{t+1})}\right)$$

(9)

if p_{ijk}^t belongs to $p_{ij,c}^{t+1}$.

If $f(p_{ij,c}^t)$ is lower than $f(p_{ij,c}^{t-1\&t+1})$, the fine resolution pixels in the set $p_{ij,c}^{t-1\&t+1}$ are temporally dependent with the corresponding fine pixels in $\mathbf{X}^{t-1}$ and $\mathbf{X}^{t+1}$, and the probability that fine resolution pixels in $p_{ij,c}^{t-1\&t+1}$ belongs to the cth class is proportional to the value of $f(p_{ij,c}^t)$:

$$D_L^T\left(c\left(p_{ijk}^t\right)=c\right)=\frac{f\left(p_{ij,c}^t\right)}{f\left(p_{ij,c}^{t-1\&t+1}\right)}.$$

(10)

In contrast, the fine resolution pixels in the sets $p_{ij,c}^{t-1}$ and $p_{ij,c}^{t+1}$ are temporally independent, and the corresponding local adjust factor equals to 0.

2.2) Only $\mathbf{X}^{t-1}$ is Available:

Only the number of fine resolution pixels of $p_{ij,c}^{t-1}$ in $\mathbf{X}^{t-1}$ is considered. If $f(p_{ij,c}^t)>f(p_{ij,c}^{t-1})$, the local adjust factor equals to 1 if p_{ijk}^t belongs to $p_{ij,c}^{t-1}$ and 0 otherwise. If $f(p_{ij,c}^t)\leqslant f(p_{ij,c}^{t-1})$, the local adjust factor is calculated as:

$$D_L^T\left(c\left(p_{ijk}^t\right)=c\right)=\frac{f\left(p_{ij,c}^t\right)}{f\left(p_{ij,c}^{t-1}\right)}$$

(11)

if p_{ijk}^t belongs to $p_{ij,c}^{t-1}$ and 0 otherwise.

2.3) Only $\mathbf{X}^{t+1}$ is Available:

Only the number of fine resolution pixels of $p_{ij,c}^{t+1}$ in $\mathbf{X}^{t+1}$ is considered. If $f(p_{ij,c}^t)>f(p_{ij,c}^{t+1})$, the local adjust factor equals to 1 if p_{ijk}^t belongs to $p_{ij,c}^{t+1}$ and 0 otherwise. If $f(p_{ij,c}^t)\leqslant f(p_{ij,c}^{t+1})$, the local adjust factor is calculated as:

$$D_L^T\left(c\left(p_{ijk}^t\right)=c\right)=\frac{f\left(p_{ij,c}^t\right)}{f\left(p_{ij,c}^{t+1}\right)}$$

(12)

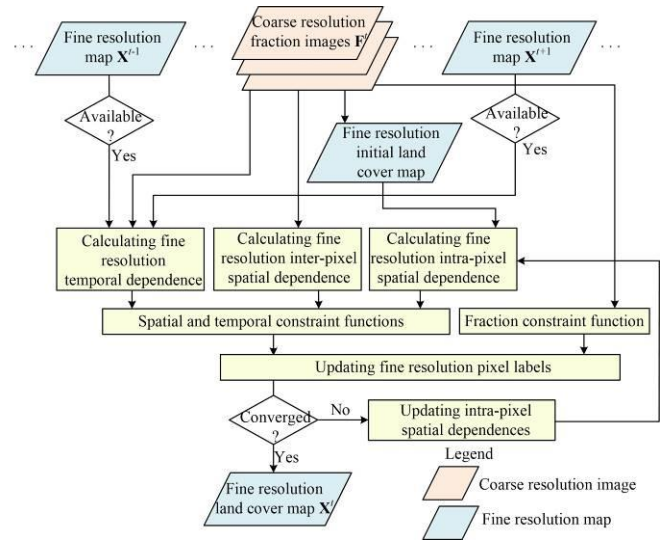if p_{ijk}^t belongs to $p_{ij,c}^{t+1}$ and 0 otherwise.

Fig. 3. The flowchart of LSTSRM.

Although only $\mathbf{X}^{t-1}$ or $\mathbf{X}^{t+1}$ is considered in cases (2) and (3), the LSTSRM temporal dependence model is different from those in [7, 11, 17, 18, 43], because the local information is considered in LSTSRM but not in the previous studies.

C. Fraction Constraint Function

The land cover fraction constraint function represents the difference between the class fractions in the input fraction images $\mathbf{F}^t$ and the final fine resolution map $\mathbf{X}^t$:

$$U_F(\mathbf{F}^t|\mathbf{X}^t)=\sum_{i=1}^I\sum_{j=1}^J\left\|f_{ij,\mathbf{F}}^t-f_{ij,\mathbf{X}}^t\right\|_2$$

(13)

where $f_{ij,\mathbf{F}}^t$ is a $C\times 1$ vector of different class fraction values in the coarse resolution pixel (i,j) in $\mathbf{F}^t$. $f_{ij,\mathbf{X}}^t$ is a $C\times 1$ vector of different class fraction values in the coarse resolution pixel (i,j) in $\mathbf{X}^t$ calculated by dividing the number of fine resolution pixels of different classes in coarse resolution pixel (i,j) by s^2 in $\mathbf{X}^t$. $\|\cdot\|_2$ indicates the L2 norm.

D. Fine Resolution Map Initialization and Updating

The flowchart of LSTSRM is shown in Fig. 3. An initial fine resolution land cover map is used as input to LSTSRM at the outset. The initial map is produced according to the land cover class fraction images. The fine resolution pixels are allocated class labels randomly in a manner that maintains the class proportional information conveyed by fraction values [44]. The class labels in the initial fine resolution land cover map are then updated iteratively. The iterative conditional mode, a simple gradient-based optimization algorithm, was applied for updating the fine resolution pixel class labels.

III. EXPERIMENT AND RESULTS

The proposed LSTSRM model was assessed in three experiments. The first used the National Land Cover Database (NLCD) of U.S.A. [45], the second used Sentinel-2 and Google Earth images, and the third used MODIS and Landsat images. In each experiment, in order to explore the influence of input fine resolution map on the proposed method, LSTSRM using

two fine resolution maps (i.e., LSTSRM$^{t-1\&t+1}$, the superscripts '$t-1$' and '$t+1$' indicate the fine resolution land cover maps that pre- and post-dates the prediction map) was compared with LSTSRM using only one fine resolution map (i.e., LSTSRM^{t-1} and LSTSRM^{t+1}). In addition, the proposed method using two fine resolution maps but using global land cover temporal dependence model (i.e., GSTSRM$^{t-1\&t+1}$) was also compared. In GSTSRM$^{t-1\&t+1}$, the local adjust factor, which may vary for different fine pixels and for different classes in LSTSRM, is set to 1 for all fine resolution pixels and for different classes in the entire image in Eq. (7).

Several popular SRM algorithms were used for comparison including the pixel swapping algorithm based SRM (PSA) [20], the Kriging interpolation based SRM (KI) [41, 46, 47], the Hopfield neural network based SRM (HNN) [25], the spatial-temporal pixel swapping algorithm (STPSA) [17], the subpixel land cover change mapping algorithm (SLCCM) [5], and the SRM based on spatial–temporal dependence from a former map (SRM_STD) [34]. Among different methods, PSA, KI and HNN are applied to mono-temporal coarse resolution land cover fraction images, which are referred as mono-temporal SRM (MTSRM) methods, and STPSA, SLCCM and SRM_STD are applied to coarse resolution land cover fraction images and a fine resolution land cover map, which are referred as spatial-temporal SRM (STSRM) methods. The LSTSRM weight parameters in all experiments were set through trial and error.

A. Simulated NLCD Experiment

1) Data Preparation

The 30 m resolution NLCD maps were adopted in this experiment. 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. The study area is located in Charlotte (33°7'00"N and 81°3'00"W), U.S.A. The NLCD maps acquired in 2001, 2006 and 2011, each contains 800×800 pixels in size, were used as the fine resolution land cover maps (Fig. 4(a-c)). The original NLCD maps contain sixteen classes according to the NLCD classification system modified from the Anderson Land Cover Classification System [45]. The original sixteen classes were reclassified to eight classes, namely water, developed, barren, forest, shrubland, herbaceous, planted/cultivated, and wetlands in this experiment.

This experiment was intended to predict the NLCD 2006 map, using coarse resolution fractions images in 2006 and fine resolution NLCD maps in 2001 and 2011.The coarse resolution fraction images in 2006 were simulated based on NLCD 2006. The coarse resolution fraction image of each class in the year 2006 was produced by dividing the number of fine resolution pixels that belong to that class in each coarse resolution pixel in NLCD 2006 according to the scale factor s, which was set s=10. This approach can produce error-free fraction images compared with those produced by soft classification [20, 24].

The input of different methods was set. For all methods, the coarse resolution class fraction images in the year 2006 were inputted. For SRM methods, STPSA, SLCCM and SRM_STD
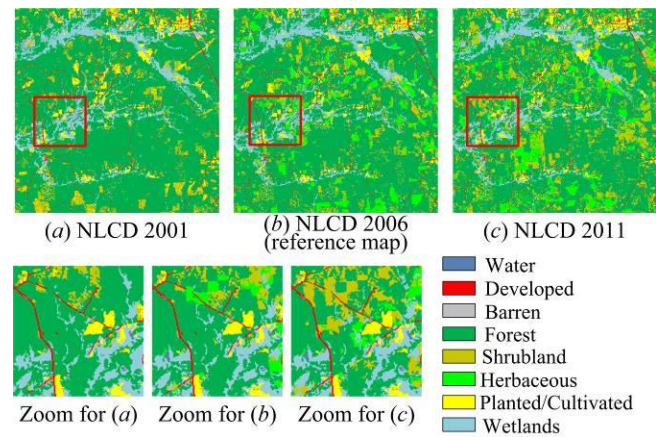

Fig. 4. The reference maps and change maps of the NLCD data.

and LSTSRM using fraction images in the year 2006 and NLCD 2001 map (i.e., STPSA^{t-1}, SLCCM^{t-1}, SRM_STD^{t-1}, and LSTSRM^{t-1}) and using fraction images in the year 2006 and NLCD 2011 map (i.e., STPSA^{t+1}, SLCCM^{t+1}, SRM_STD^{t+1} and LSTSRM^{t+1}) were tested. LSTSRM$^{t-1\&t+1}$ and GSTSRM$^{t-1\&t+1}$ using both the 2001 and 2011 NLCD maps were compared. The accuracies of different methods were assessed using the NLCD 2006 (Fig. 4 (b)).

2) Results

The resulting maps in the zoomed area from different methods are shown in Fig. 5. The KI map contained unsmoothed boundaries because it discarded the intra-pixel spatial dependence which can help to generate locally smoothed boundaries in the result (Fig. 5(b)). The PSA map failed to reconstruct the holistic land cover spatial patterns because it discarded the land cover inter-pixel spatial dependence (Fig. 5(c)). The PSA map contained many land cover patches with small size represented as speckle-like artifacts. In contrast, the HNN map eliminated the speckle-like artifacts (Fig. 5(d)). This difference lays in the fact that KI and PSA must preserve class fractions in the resulting map, and the land covers with small area proportion or fraction in a coarse resolution pixel would be aggregated to small land cover patches represented as speckle-like artifacts. In contrast, HNN does not strictly preserve the class fractions from the input class fraction images into the result land cover map, and would eliminate the speckle-like artifacts due to spatial smoothing effect [20, 25]. However, since KI, PSA and HNN only considers the land cover spatial information but neglect land cover temporal information in the land cover map that pre- and/or post-dates the prediction date, the spatial details were not represented in the result maps.

The STSRM methods, including STPSA, SLCCM, SRM_STD and LSTSRM, preserved the spatial details of land cover classes in Fig. 5 (Fig. 5(e-n)). All these maps were very similar to the reference NLCD 2006 map. The linear developed class was connected in these maps, and the shapes of objects such as herbaceous, planted/cultivated, and water were reconstructed.

TABLE II
THE OVERALL ACCURACIES (%) OF DIFFERENT METHODS IN THE NLCD EXPERIMENT.

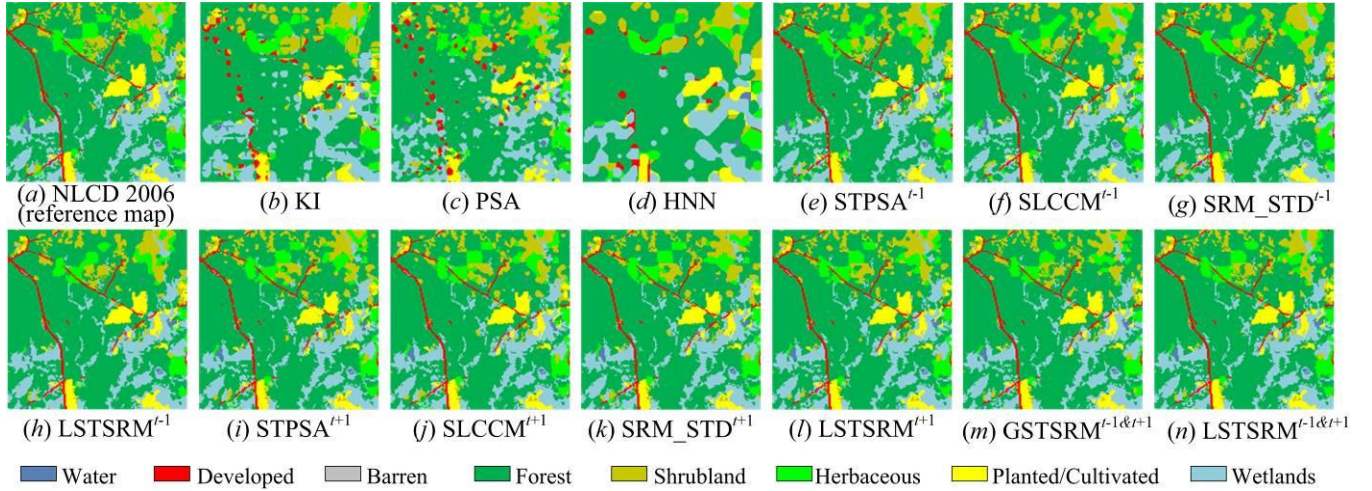| KI	PSA	HNN	$STPSA^{t-1}$	$STPSA^{t+1}$	$SLCCM^{t-1}$	$SLCCM^{t+1}$	SRM_STD^{t-1}	SRM_STD^{t+}	$LSTSRM^{t-1}$	$LSTSRM^{t+1}$	$GSTSRM^{t-1\&t+1}$	$LSTSRM^{t-1\&t+1}$
78.06	74.59	80.11	89.11	91.34	90.14	92.31	91.09	93.08	91.61	93.41	93.37	**94.39**

Fig. 5. The result maps in the zoomed area from different methods. The superscripts '$^{t-1}$' and '$^{t+1}$' indicate NLCD 2001 and 2011 maps used in the methods.
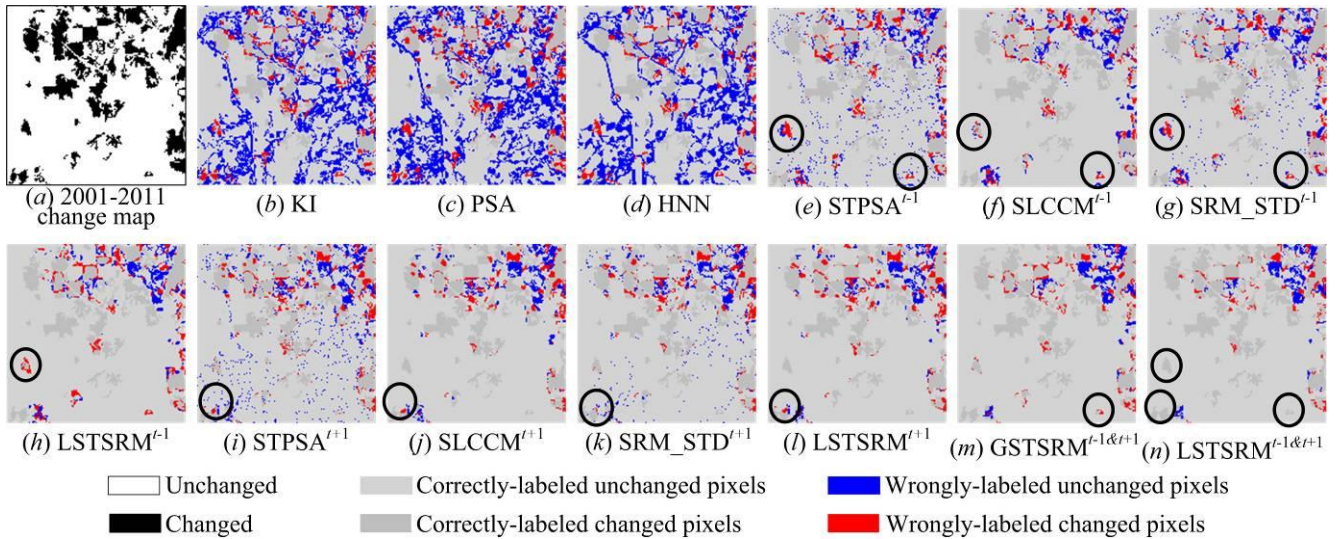

Fig. 6. The error maps in the zoomed area from different methods. The superscripts '$^{t-1}$' and '$^{t+1}$' indicate NLCD 2001 and 2011 maps used in the methods.

Fig. 6 shows the error maps from different methods. The MTSRM algorithms of KI, PSA and HNN generated more error pixels (highlighted in red and blue in Fig. 6(b-d)) compared with those STSRM methods in Fig. 6(e-n). In addition, the MTSRM generated more wrongly-labeled unchanged pixels highlighted in blue than wrongly-labeled changed pixels highlighted in red in Fig. 6(b-d)), whereas STSRM eliminated most wrongly-labeled unchanged pixels highlighted in blue in Fig. 6(e-n), showing that incorporating temporal dependence in SRM can reduce the commission error especially for unchanged pixels. Among all the result maps, the $LSTSRM^{t-1\&t+1}$ contained the least wrongly-labeled fine pixels

in Fig. 6(n), and the wrongly-labeled fine pixels highlighted in the circles in other STSRM maps (Fig. 6(e-l)) were eliminated in the $LSTSRM^{t-1\&t+1}$ map. This result not only shows that the proposed $LSTSRM^{t-1\&t+1}$ increased the accuracy than the existing STSRM algorithms of STPSA, SLCCM and SRM_STD, but also shows that the proposed STSRM using two fine maps is superior to that using only one fine map, and the proposed STSRM using local temporal dependence model is superior to that using global temporal dependence model.

The overall accuracies of different methods are shown in Table II. The overall accuracies of MTSRM methods were lower than 81%, whereas those of STSRM methods were

(a) Land cover map at 2016-2-20

(b) Land cover map at 2016-9-26

(c) Land cover map at 2017-12-20

(d) Sentinel-2 image at 2016-9-7

(e) Change map from 2016-2-20 to 2016-9-26

(f) Change map from 2016-9-26 to 2016-12-20

Zoom for (a)   Zoom for (b)   Zoom for (c)

■ Water
■ Vegetation
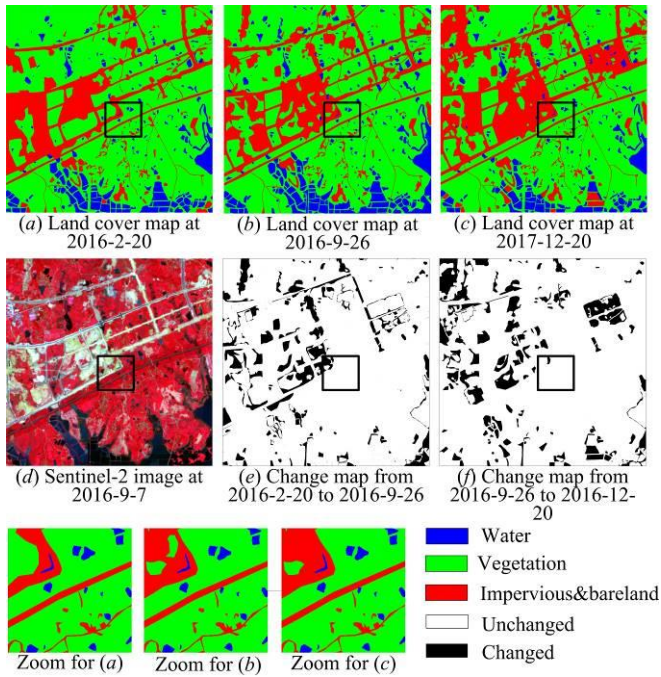■ Impervious&bareland
□ Unchanged
■ Changed

Fig.7. The reference fine resolution maps, Sentinel-2 images and change maps. The Sentinel-2 image uses NIR-red-green as RGB.

higher than 89%. The accuracies of STPSA^{t-1}, SLCCM^{t-1} and SRM_STD^{t-1} were similar, and the accuracies of STPSA^{t+1}, SLCCM^{t+1} and SRM_STD^{t+1} were similar, showing that the input fine resolution map plays a key role for STPSA, SLCCM and SRM_STD. In addition, the overall accuracy for LSTSRM^{t-1} was higher than those of STPSA^{t-1}, SLCCM^{t-1} and SRM_STD^{t-1}, and the overall accuracy for LSTSRM^{t+1} was higher than those of STPSA^{t+1}, SLCCM^{t+1} and SRM_STD^{t+1}. It shows that LSTSRM improves the accuracy compared with STPSA, SLCCM and SRM_STD when only one fine map is used in STSRM. LSTSRM$^{t-1\&t+1}$ generated the highest overall accuracy among all methods, showing the advantage of the proposed method.

B. Sentinel-2 Experiment

1) Data Preparation

The LSTSRM was tested using real Sentinel-2 remotely sensed images in this experiment. Sentinel-2 was launched by the European Space Agency in 2015, and can provide global acquisitions of fine resolution multi-spectral images with a fine revisit frequency. The Sentinel-2 image is useful in land cover mapping due to its appealing properties (10 days at the equator with one satellite, and 5 days with 2 satellites which result in 2-3 days at mid-latitudes) and the free access. In this experiment, Sentinel-2 image was utilized to map land covers in an urban area located in Wuhan (30°27′30″N and 114°32′30″E), Hubei province, China. The Sentinel-2 image acquired on September 7 2016 with four 10 m spatial resolution Sentinel-2 bands (blue, green, red and infrared bands) was used to generate land cover map in the study area (Fig. 7(d)). A Google Earth image acquired on September 26 2016 was digitized to a 2 m spatial resolution land cover map for accuracy assessment (Fig. 7(b)). Two fine resolution Google Earth

images that acquired on February 20 2016 and December 20 2017, respectively, were digitized to 2 m spatial resolution land cover maps as the STSRM input (Fig. 7(a),(c)). The study area covers 400×400 Sentinel-2 pixels, which correspond to 2000×2000 fine resolution pixels in the input and reference maps, with a scale factor $s = 5$. There are three land cover types, which are water, vegetation and impervious/bareland, contained in the input and reference land cover maps.

The proposed LSTSRM using both fine maps and using local temporal dependence model, i.e., LSTSRM$^{t-1\&t+1}$, was compared with the same methods used in the first experiments, including MTSRM of KI, PSA and HNN, as well as the STSRM methods of STPSA^{t-1}, STPSA^{t+1}, SLCCM^{t-1}, SLCCM^{t+1}, SRM_STD^{t-1}, SRM_STD^{t+1}, LSTSRM^{t-1}, LSTSRM^{t+1} and GSTSRM$^{t-1\&t+1}$ (the superscripts 't-1' and 't+1' indicate the Google Earth 2016 and 2017 maps, respectively). For the MTSRM and STSRM methods, the multiple endmember spectral mixture analysis was applied to generate land cover class fraction images [48].

2) Results

Fig. 8 shows the result maps and zoomed images from different methods. Different to the NLCD experiment which used error-free coarse spatial resolution land cover fraction images, this experiment used fraction images that were unmixed from the Sentinel-2 image which inevitably contained errors in Fig. 8. The zoomed image for KI (b) and PSA (c) contained many speckle-like artifacts which were resulted from soft classification error. For instance, if a coarse pixel does not contain pixels of water class and the unmixed water fraction is 12% in this coarse pixel, then a total number of $5^2 \times 12\%$ =3 (5 is the scale factor) fine pixels are labeled as water class within this coarse pixel, which may be represented as speckle-like artifacts since these methods must preserve class fractions in the resulting map. KI and PSA preserved class fractions in the resulting map, resulting in speckle-like artifacts due to soft classification errors in the class fraction images. HNN eliminated the speckle-like artifacts because it had the spatial smoothing effect and did not strictly preserve the class fractions from the input class fraction images into the result land cover map. However, the impervious&bareland patch highlighted by the ellipse in the zoomed area in Fig. 8(d) was wrongly labeled as water, and the detailed spatial pattern of the linear impervious&bareland patch highlighted by the circle was not reconstructed.

Among the STSRM results in Fig. 8, the STPSA, SLCCM and SRM_STD (Fig. 8(e-g),(i-k)) contained a large number of speckle-like artifacts due to soft classification error, whereas LSTSRM (Fig. 8(h),(l),(n)) and GSTSRM (Fig. 8(m)) eliminated these errors due to spatial smoothing effect. Similar to HNN, LSTSRM and GSTSRM eliminated the speckle-like artifacts because they have the spatial smoothing effect and do not to strictly preserve the class fractions from the input class fraction images in the result land cover map. The LSTSRM and GSTSRM maps in Fig. 8 (h) and (l-n) were much similar to the reference map than the HNN map in Fig. 8(d). This is because LSTSRM and GSTSRM incorporated land cover temporal information from the input land cover map whereas HNN did not. As soft classification error is usually unavoidable in real applications, LSTSRM and GSTSRM would be more suitable

TABLE III
THE ACCURACIES (%) OF DIFFERENT METHODS IN THE SENTINEL-2 IMAGE EXPERIMENT.

|		KI	PSA	HNN	STPSA^{t-1}	STPSA^{t+1}	SLCCM^{t-1}	SLCCM^{t+1}	SRM_STD^{t-1}	SRM_STD^{t+1}	LSTSRM^{t-1}	LSTSRM^{t+1}	GSTSRM$^{t-1\&t+1}$	LSTSRM$^{t-1\&t+1}$
Overall accuracy		76.49	75.66	84.77	76.68	76.78	76.67	76.77	76.64	76.72	93.34	93.64	93.92	**95.24**
Producer's accuracy	Water	88.78	87.36	94.25	89.32	89.29	89.29	89.24	89.29	89.28	92.73	**94.53**	85.60	91.58
	Vegetation	76.88	76.21	86.11	77.02	77.10	77.01	77.09	77.03	77.10	95.59	94.82	95.90	**96.50**
	Impervious /bareland	69.81	68.79	76.65	70.00	70.21	69.99	70.21	69.78	69.93	87.25	89.91	92.08	**93.35**
User's accuracy	Water	43.29	42.57	59.48	43.52	43.50	43.51	43.48	43.51	43.50	90.68	85.05	**94.41**	94.32
	Vegetation	90.28	89.59	92.65	90.53	90.63	90.53	90.62	90.55	90.63	94.53	95.90	95.15	**96.38**
	Impervious /bareland	74.05	72.85	82.06	74.13	74.35	74.12	74.36	73.90	74.06	91.08	91.61	90.26	**92.45**
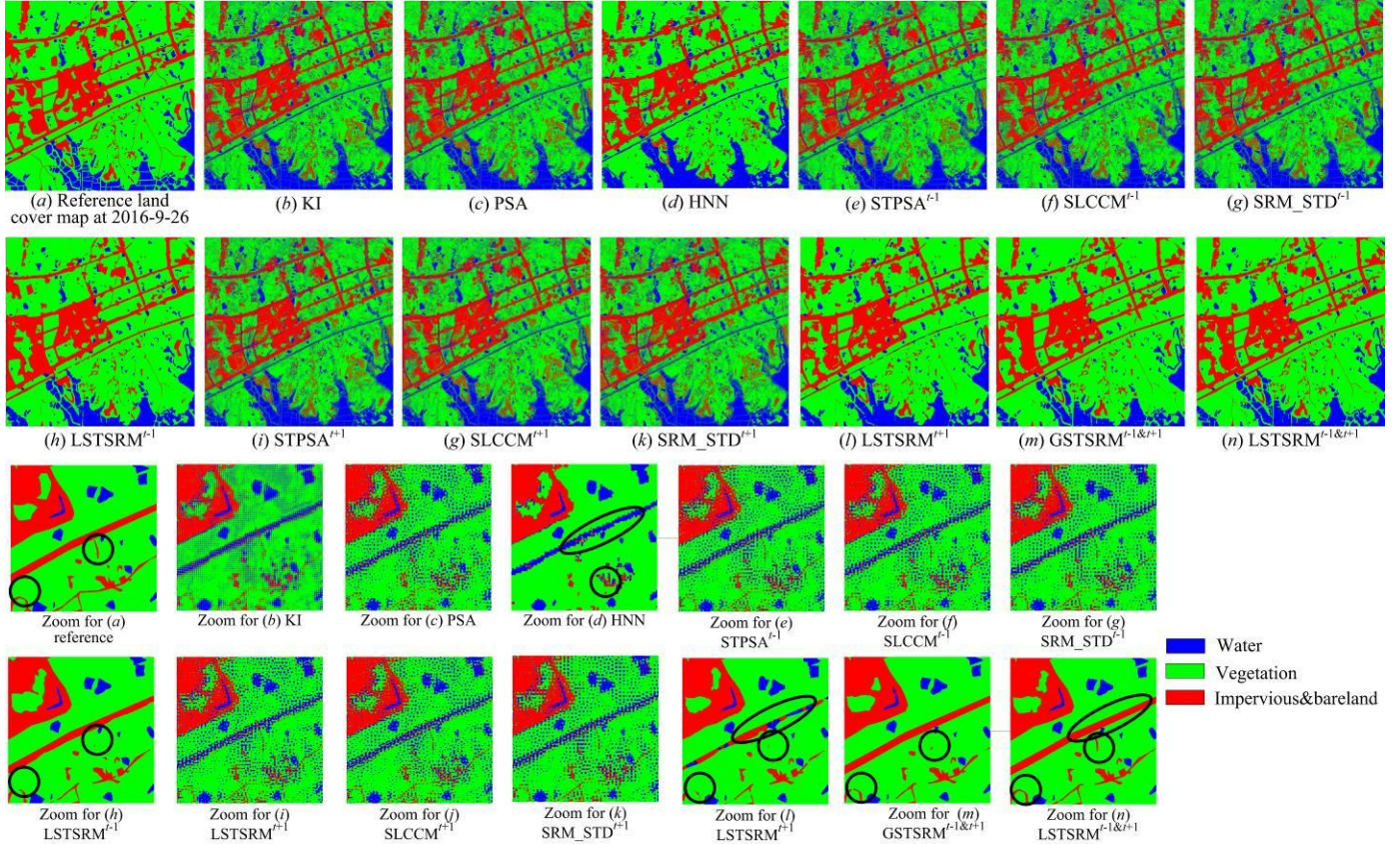
Fig. 8. The result maps in the zoomed area from different methods in the Sentinel-2 image experiment. The superscripts '$^{t-1}$' and '$^{t+1}$' indicate Google Earth 2016 and 2017 maps used in the methods.

for land cover mapping of image series compared with STPSA, SLCCM and SRM_STD in practice.

In the zoomed images, LSTSRM^{t-1}, LSTSRM^{t+1} and GSTSRM$^{t-1\&t+1}$ failed to reconstruct the linear impervious&bareland patch highlighted by the circles in Fig. 8(h),(l) and (m). LSTSRM^{t+1} erroneously labeled a part of impervious&bareland as water highlighted by the ellipse in the zoomed image for Fig. 8(l). LSTSRM^{t-1} and GSTSRM$^{t-1\&t+1}$ erroneously labeled a part of impervious&bareland as vegetation highlighted by the circle in the zoomed image for Fig. 8 (h), (m). LSTSRM$^{t-1\&t+1}$ correctly labeled the impervious&bareland patch highlighted by the ellipse and reconstructed most parts of the linear impervious&bareland patch highlighted by the circles in the zoomed image for (n), and was similar to the reference image.

The overall accuracy, producer's and user's accuracies of different methods are presented in table III. The overall accuracies for KI, PSA, STPSA, SLCCM and SRM_STD, which strictly preserve the class fractions from the input class fraction images in the result land cover map, were lower than 80%, showing that the soft classification error strongly affects these MTSRM and STSRM methods. HNN had an overall accuracy of about 85%, and LSTSRM and GSTSRM increased the overall accuracy to higher than 93%. Among LSTSRM and GSTSRM, for the water class, LSTSRM^{t+1} has the highest producer's accuracy but the lowest user's accuracy served as a high commission error of water. This is shown in the zoomed image of (l) in which some impervious&bareland pixels were wrongly labeled as water highlighted by the ellipse in LSTSRM^{t+1}. In addition, among LSTSRM and GSTSRM and

for the water class, GSTSRM$^{t-1\&t+1}$ has the highest user's accuracy, but the lowest producer's accuracy served as a high omission error of water. For the vegetation to impervious&bareland class which have a large degree of land cover change in Fig. 7(e-f), LSTSRM$^{t-1\&t+1}$ has the highest producer's and user's accuracies served as the lowest omission and commission errors for these two classes. LSTSRM$^{t-1\&t+1}$ has the highest overall accuracy, showing the advantages of the proposed method.

C. MODIS Experiment

1) Data Preparation

The LSTSRM was tested using real MODIS image in this experiment. The study area is located near Sorriso (12º33'00"S and 55º42'00"W) in Mato Grosso State, Brazil. This area was mostly covered by tropical forests but has suffered from deforestation in recent years [8]. Three Landsat TM images (path 226, row 069) acquired on July 12 2002, July 23 2003 and June 23 2004 were downloaded from the USGS website. 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. The three Landsat images were classified at a 30 m spatial resolution (Fig. 9 (a-c)). Two land cover classes, forest and nonforest, were considered in this experiment. The endmembers of each class were manually selected from each Landsat image, and the maximum likelihood classifier was applied to generate the fine resolution forest/nonforest maps each year.

A 8-day surface reflectance MODIS product (MOD09A1) datasets comprising seven spectral bands (620 nm - 2055 nm) with a spatial resolution of 463 m acquired in July 2003 was used (Fig. 9(d)). The MODIS image was re-projected to the UTM coordinate system and resampled to a spatial resolution of 450 m using the nearest neighbor interpolation which may not over-smooth the resized image. The study area covers 300×300 MODIS pixels, which correspond to 4500×4500 Landsat pixels, with a scale factor s=15.

The same MTSRM and STSRM methods that used in the NLCD and Sentinel-2 experiments were used in this experiment. The multiple endmember spectral mixture analysis was applied to the MODIS image to generate coarse resolution land cover fraction images. STSRM incorporated the 2002 and/or 2004 land cover maps in Fig. 9(a), (c) as ancillary data. The accuracies of different methods were assessed using the 30 m resolution 2003 land cover maps (Fig. 9(b)).

2) Results

The zoomed areas of the result maps from different methods were shown in Fig. 10. Similar to the Sentinel-2 experiments, the KI, PSA STPSA, SLCCM and SRM_STD maps contained speckle-like artifacts due to soft classification errors in the class fraction images in Fig. 10. The HNN maps eliminated speckle-like artifacts due to spatial smoothing effect. However, HNN generated disconnected forest patches highlighted by the circle in Fig. 10(d), and the shape of the forest patch was smoothed and dissimilar to that in the reference map. By contrast, the LSTSRM and GSTSRM maps were similar to the reference map than those generated from other methods. The shape of the forest patch was mostly reconstructed by LSTSRM and GSTSRM. Both LSTSRM^{t-1} and LSTSRM^{t+1} generated

(a) 2002 land cover map   (b) 2003 land cover map   (c) 2004 land cover map

(d) 2003 MODIS   (e) 2002-2003 change map   (f) 2003-2004 change map

Zoom for (a)   Zoom for (b)   Zoom for (c)
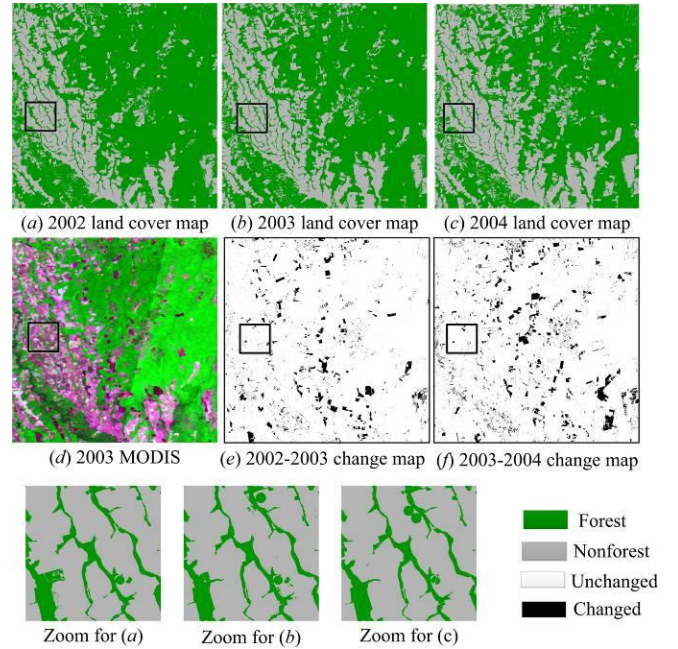
Forest
Nonforest
Unchanged
Changed

Fig.9. The reference fine resolution maps, MODIS images and change maps in the MODIS image experiment. The MODIS images use SWIR-NIR-red as RGB.

disconnected forest patch that were highlighted by the circle in Fig. 10 (h) and (l), whereas GSTSRM$^{t-1\&t+1}$ and LSTSRM$^{t-1\&t+1}$ generated more connected forest patches in Fig. 10 (m-n), showing incorporating two fine maps that pre- and post-dates the predicting time usually increase the accuracy than those using only one fine image. In addition, the LSTSRM$^{t-1\&t+1}$ was more similar to the reference image than GSTSRM$^{t-1\&t+1}$ such as those highlighted by the ellipse and circle.

The accuracies of different methods are shown in table IV. The overall accuracies of KI, PSA, STPSA, SLCCM and SRM_STD were lower than 90%, whereas that of HNN, LSTSRM and GSTSRM were higher than 90%. It shows that the soft classification error has a strong effect on these MTSRM and STSRM methods. The overall accuracies of LSTSRM and GSTSRM were all higher than 95%. Among LSTSRM and GSTSRM, for the forest class, GSTSRM$^{t-1\&t+1}$ has the highest producer's accuracy and the second lowest user's accuracy served as commission errors of forest. Among LSTSRM and GSTSRM and for the nonforest class, LSTSRM^{t+1} has the highest producer's accuracy but the lowest user's accuracy served as the highest commission error of nonforest. For instance, the forest patch was erroneously labeled as nonforest highlighted by the ellipse in Fig. 10(l), resulting in a higher commission error of nonforest for LSTSRM^{t+1}. In contrast, the small forest patch which was erroneously labeled as nonforest highlighted by the ellipse in LSTSRM^{t+1} was correctly predicted in LSTSRM$^{t-1\&t+1}$ in Fig. 10(n). LSTSRM$^{t-1\&t+1}$ has relatively high producer's and user's accuracies for both forest and nonforest, and the highest overall accuracy among all methods.

IV. DISCUSSION

In this section, the similarity and difference between STSRM and two popular image fusion methods, which are
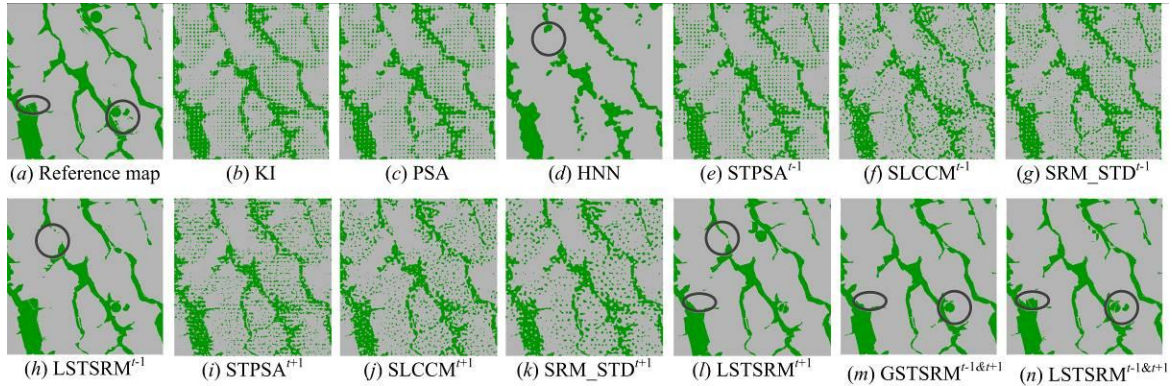
Fig. 10. The result maps from different methods in the zoomed areas in the MODIS image experiment. The superscripts "t-1" and "t+1" indicate 2002 and 2004 fine resolution maps used in the methods.

TABLE IV
THE OVERALL ACCURACIES (%) OF DIFFERENT METHODS IN THE MODIS IMAGE EXPERIMENT.

		KI	PSA	HNN	STPSA^{t-1}	STPSA^{t+1}	SLCCM^{t-1}	SLCCM^{t+1}	SRM_STD^{t-1}	SRM_STD^{t+1}	LSTSRM^{t-1}	LSTSRM^{t+1}	GSTSRM$^{t-1\&t+1}$	LSTSRM$^{t-1\&t+1}$
Overall accuracy		86.58	86.51	92.07	88.10	87.89	87.90	87.92	88.11	87.97	96.63	95.78	96.74	**97.43**
Producer's accuracy	forest	86.10	86.04	92.23	87.22	87.06	87.07	87.08	87.23	87.12	96.98	94.00	**97.05**	96.82
	nonforest	87.60	87.49	91.74	89.94	89.61	89.62	89.66	89.96	89.74	95.91	**99.49**	96.09	98.72
User's accuracy	forest	93.54	93.48	95.88	94.76	94.58	94.59	94.61	94.77	94.65	98.02	**99.74**	98.10	99.37
	nonforest	75.14	75.04	85.00	77.15	76.86	76.88	76.90	77.16	76.98	93.84	88.84	**93.98**	93.71

spatial-temporal image fusion and hyper-spectral image super-resolution, are discussed. Then the influencing factor of the changed and unchanged pixels to the proposed LSTSRM is discussed.

A. Comparison of STSRM, Spatial-temporal Image Fusion (STIF), and Hyper-spectral Image (HSI) Super-resolution

With the development of remote sensing society, a huge number of remotely sensors have been launched recently. The optical remotely sensed images usually have a tradeoff between the spatial, temporal, and spectral resolutions, due to technical limitations factors and the orbit of the platforms. Various methods are proposed to fuse images of the same scene, using complementary information provided. LSTSRM was compared theoretically with STIF and HSI super-resolution.

1) Comparison of STSRM and STIF

STIF is an approach that generates a fine resolution image for the date represented by a coarse resolution image by integrating the spatial and temporal information in a pair of fine and coarse resolution images of the same region acquired at other dates [49-51]. Both STSRM and STIF aim to overcome the limitation caused by the tradeoff between spatial and temporal resolutions of optical remotely sensed images. The main difference lays in that STIF predicts fine spatial-temporal resolution reflectance images or indices such as Normalized Difference Vegetation Index (NDVI) time-series [52] which can be used in applications such as the monitoring of vegetation seasonal change [53] and in the assessment of vegetation status [54]. In contrast, STSRM predicts fine spatial-temporal resolution land cover maps which can be used in applications such as land cover change analysis. STIF is more appropriate in

the analysis based on image reflectance whereas STSRM is more appropriate in the analysis based on land cover types.

2) Comparison of STSRM and HSI Super-resolution

HSI super-resolution is an approach that fuses coarse spatial resolution HSI with fine spatial resolution multispectral images or panchromatic images in order to obtain super-resolution (spatial and spectral) hyperspectral images [55-57]. HSI super-resolution aims to overcome the limitations caused by the tradeoff between spatial and spectral resolutions of optical remotely sensed images, using the complementary characteristics in the inference of images with fine spatial-spectral resolutions. There are two main differences between HSI super-resolution and STSRM. First, HSI super-resolution usually requires the input coarse spatial resolution HSI and fine spatial resolution multi-spectral images to be acquired at the same or close date so that land cover does not change between the acquisition dates of these images. In contrast, for STSRM, the input coarse spatial resolution class fraction images and the fine spatial resolution land cover map are derived from remotely sensed images that are acquired at different dates. Second, HSI super-resolution is used to predict fine spatial-temporal resolution images whereas STSRM directly outputs fine spatial-temporal resolution land cover maps. If the aim is to extract land cover information, a procedure of land cover classification is still need to be applied to image outputted from HSI super-resolution.

B. Influence of Changed and Unchanged Pixels on LSTSRM

The influence of the percentage of changed and unchanged pixels in the land cover maps on LSTSRM is explored. Take the NLCD experiment for example, table V showed the percentage of changed pixels for each class as well as the producer's and

TABLE V

THE PERCENTAGE OF CHANGED PIXELS IN THE INPUT FINE RESOLUTION MAPS FOR EACH CLASS (%), AS WELL AS THE PRODUCER'S AND USER'S ACCURACIES (%) IN THE NLCD EXPERIMENT.

Class	Years 2001-2006			Years 2006-2011			Years 2001-2011				
	Changed percent of fine pixels	LSTSRM^{t-1}		Changed percent of fine pixels	LSTSRM^{t+1}		Changed percent of fine pixels	GSTSRM$^{t-1\&t+1}$		LSTSRM$^{t-1\&t+1}$	
		Producer's	User's		Producer's	User's		Producer's	User's	Producer's	User's
Water	10.19	95.51	98.02	5.62	97.59	98.86	9.15	97.19	98.3	98.88	98.72
Developed	0.00	100.00	99.91	0.00	100.00	99.89	0.00	99.98	100.00	100.00	99.93
Barren	100.00	35.63	46.97	80.46	63.22	70.51	62.07	58.62	94.44	62.07	65.85
Forest	5.87	96.58	93.78	17.41	97.14	95.39	22.97	97.92	94.38	97.8	95.94
Shrubland	46.26	64.69	73.68	35.54	73.05	78.17	33.94	73.61	82.96	77.81	83.62
Herbaceous	99.88	65.25	77.83	88.56	70.79	78.6	55.53	61.95	80.45	71.25	79.44
Planted/Cultivated	4.95	96.55	95.48	6.03	96.92	97.78	4.6	97.79	98.27	98.44	98.82
Wetlands	2.52	96.97	95.49	0.23	99.8	99.16	2.75	99.66	98.74	99.94	99.61

user's accuracies. During years 2001-2006, for classes with a low (<=10%) percentage of changed pixels such as water, developed, forest, planted/cultivated and wetlands class, the producer's and user's accuracies were higher than 95% for LSTSRM^{t-1}. For classes with a high (>40%) percentage of changed pixels during 2001-2006 such as barren, shrubland and herbaceous classes, the producer's and user's accuracies were lower than 75% for LSTSRM^{t-1}. Similarly, for GSTSRM$^{t-1\&t+1}$ and LSTSRM$^{t-1\&t+1}$, for classes with a low (<=5%) percentage of changed pixels such as developed and planted/cultivated during 2001-2011, the producer's and user's accuracies were higher than 97%, and for classes with a high (>50%) percentage of changed pixels such as barren and herbaceous classes, the producer's and user's accuracies were usually lower than 80%. This shows that the proposed method is more competent in predicting pixels with unchanged labels. Other land cover temporal models could be developed to deal with the complicated land cover change scenarios.

V. CONCLUSION

In this paper, a novel local adaptive dependence based spatial-temporal super-resolution mapping model was proposed. Unlike traditional STSRM models using only one fine resolution land cover map as ancillary data, the proposed LSTSRM model considers the fine resolution maps pre- and/or post-dates the coarse resolution cases, and develops the local temporal dependence model, in which the dependence intensity may vary from fine resolution pixel to fine resolution pixel. LSTSRM does not to strictly preserve the class fractions from the input class fraction images into the result land cover map, and can eliminate fraction errors caused by the soft classification procedure to some extent.

The LSTSRM performance was validated using NLCD data, real Sentinel-2 imagery and real MODIS imagery by comparing with several popular SRM algorithms. Results showed that LSTSRM resulting maps eliminated most speckle-like artifacts. Moreover, the LSTSRM resulting maps maintained the connectivity for the linear shaped patches, and were closer to the reference maps than other methods. The proposed LSTSRM generated the highest overall accuracies in all the experiments. In addition, the proposed method using two fine resolution maps and using local temporal dependence model improved the accuracy by comparing with that using two fine resolution maps but using global temporal dependence model, and by comparing with that using only one fine resolution map and using local temporal dependence model. The producer's and user's accuracies were higher for unchanged classes than for changed classes for different methods in the NLCD experiment. Research focusing on using more fine resolution land cover maps and improving the local land cover transitions for changed land cover classes in STSRM should be studied in the future.

REFERENCES

[1] P. M. Atkinson, "Mapping sub-pixel boundaries from remotely sensed images," *Innovations in GIS* 4, pp. 166-180: Z. Kemp (Ed.), 1997.
[2] X. Li, F. Ling, Y. Du, and Y. Zhang, "Spatially adaptive superresolution land cover mapping with multispectral and panchromatic images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2810-2823, May 2014.
[3] P. Aplin and P. M. Atkinson, "Sub-pixel land cover mapping for per-field classification," *Int. J. Remote Sens.*, vol. 22, no. 14, pp. 2853-2858, Sep. 2001.
[4] M. Q. Nguyen, P. M. Atkinson, and H. G. Lewis, "Superresolution mapping using a Hopfield neural network with LIDAR data," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 3, pp. 366-370, Jul. 2005.
[5] F. Ling, W. Li, Y. Du, and X. Li, "Land cover change mapping at the subpixel scale with different spatial-resolution remotely sensed imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 1, pp. 182-186, Jan. 2011.
[6] X. Li, F. Ling, G. M. Foody, Y. Ge, Y. Zhang, and Y. Du, "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.*, vol. 196, pp. 293-311, Jul. 2017.
[7] X. Li, Y. Du, and F. Ling, "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.*, vol. 7, no. 1, pp. 29-39, Jan. 2014.
[8] M. C. Hansen, Y. E. Shimabukuro, P. Potapov, and K. Pittman, "Comparing annual MODIS and PRODES forest cover change data for advancing monitoring of Brazilian forest cover," *Remote Sens. Environ.*, vol. 112, no. 10, pp. 3784-3793, Oct. 2008.
[9] X. Li, Y. Du, and F. Ling, "Sub-pixel-scale land cover map updating by integrating change detection and sub-pixel mapping," *Photogramm. Eng. Remote Sens.*, vol. 81, no. 1, pp. 59-67, Jan. 2015.
[10] X. Li, F. Ling, Y. Du, Q. Feng, and Y. Zhang, "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.*, vol. 93, pp. 76-87, Jul. 2014.
[11] X. Li, F. Ling, G. M. Foody, and Y. Du, "A superresolution land-cover change detection method using remotely sensed images with different

spatial resolutions," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 3822-3841, Jul. 2016.

[12] Y. Zhang, P. M. Atkinson, X. Li, F. Ling, Q. Wang, and Y. Du, "Learning-based spatial-temporal superresolution mapping of forest cover with MODIS images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 600-614, Jan. 2017.

[13] Q. Wang, P. M. Atkinson, and W. Shi, "Fast subpixel mapping algorithms for subpixel resolution change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 1692-1706, Apr. 2015.

[14] D. He, Y. Zhong, R. Feng, and L. Zhang., "Spatial-temporal sub-pixel mapping based on swarm intelligence theory," *Remote Sens.*, vol. 8, no. 11, Nov. 2016.

[15] K. Wu, W. Yi, R. Niu, and L. Wei., "Subpixel land cover change mapping with multitemporal remote-sensed images at different resolution," *J. Appl. Remote Sens.*, vol. 9, Feb. 2015.

[16] W. Ke, D. Qian, W. Yi, and Y. Yang, "Supervised sub-pixel mapping for change detection from remotely sensed images with different resolutions," *Remote Sens.*, vol. 9, no. 3, pp. 284, Mar. 2017.

[17] Y. Xu and B. Huang, "A spatio-temporal pixel-swapping algorithm for subpixel land cover mapping," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 2, pp. 474-478, Feb. 2014.

[18] Q. Wang, W. Shi, and P. M. Atkinson, "Spatial-temporal sub-pixel mapping of time-series images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 9, pp. 5397-5411, Sep. 2016.

[19] Y. Zhang, X. Li, F. Ling, P. M. Atkinson, Y. Ge, L. Shi, and Y. Du, "Updating Landsat-based forest cover maps with MODIS images using multiscale spectral-spatial-temporal superresolution mapping," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 63, pp. 129-142, Dec. 2017.

[20] P. M. Atkinson, "Sub-pixel target mapping from soft-classified, remotely sensed imagery," *Photogramm. Eng. Remote Sens.*, vol. 71, no. 7, pp. 839-846, Jul. 2005.

[21] P. M. Atkinson, "Issues of uncertainty in super-resolution mapping and their implications for the design of an inter-comparison study," *Int. J. Remote Sens.*, vol. 30, no. 20, pp. 5293-5308, Sep. 2009.

[22] G. M. Foody and H. T. X. Doan, "Variability in soft classification prediction and its implications for sub-pixel scale change detection and super resolution mapping," *Photogramm. Eng. Remote Sens.*, vol. 73, no. 8, pp. 923-933, Aug. 2007.

[23] G. M. Foody, A. M. Muslim, and P. M. Atkinson, "Super-resolution mapping of the waterline from remotely sensed data," *Int. J. Remote Sens.*, vol. 26, no. 24, pp. 5381-5392, Dec. 2005.

[24] Y. Ge, S. Li, and V. C. Lakhan, "Development and testing of a subpixel mapping algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2155-2164, Jul. 2009.

[25] A. J. Tatem, H. G. Lewis, P. M. Atkinson, and M. S. Nixon., "Super-resolution target identification from remotely sensed images using a Hopfield neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 4, pp. 781-796, Apr. 2001.

[26] Q. Wang, P. M. Atkinson, and W. Shi, "Indicator cokriging-based subpixel mapping without prior spatial structure information," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 309-323, Jan. 2015.

[27] Y. Ge, "Sub-pixel land-cover mapping with improved fraction images upon multiple-point simulation," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 22, pp. 115-126, Jun. 2013.

[28] Y. Zhang, Y. Du, F. Ling, S. Fang, and X. Li, "Example-based super-resolution land cover mapping using support vector regression," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 7, no. 4, pp. 1271-1283, Apr. 2014.

[29] F. Ling, G. M. Foody, Y. Zhang, X. Li, X. Zhang, S. Fang, W. Li, and Y. Du, "Learning-based superresolution land cover mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 3794 - 3810, Jul. 2016.

[30] X. Xu, Y. Zhong, and L. Zhang, "Adaptive subpixel mapping based on a multiagent system for remote-sensing imagery.," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 787-804, Feb. 2014.

[31] M. W. Thornton, P. M. Atkinson, and D. A. Holland, "A linearised pixel-swapping method for mapping rural linear land cover features from fine spatial resolution remotely sensed imagery," *Comput. Geosci.*, vol. 33, no. 10, pp. 1261-1272, Oct. 2007.

[32] B. Ai, X. Liu, G. Hu, and X. Li, "Improved sub-pixel mapping method coupling spatial dependence with directivity and connectivity," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 7, no. 12, pp. 4887-4896, Dec. 2014.

[33] F. Ling, G. M. Foody, X. Li, Y. Zhang, and Y. Du, "Assessing a temporal change strategy for sub-pixel land cover change mapping from

[34] F. Ling, X. Li, Y. Du, and F. Xiao, "Super-resolution land cover mapping with spatial-temporal dependence by integrating a former fine resolution map," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 7, no. 5, pp. 1816-1825, May 2014.

[35] Q. Wang, W. Shi, P. M. Atkinson, and Z. Li, "Land cover change detection at subpixel resolution with a Hopfield neural network," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 8, no. 3, pp. 1339-1352, Mar. 2015.

[36] D. Liu, K. Song, J. R. G. Townshend, and P. Gong, "Using local transition probability models in Markov random fields for forest change detection," *Remote Sens. Environ.*, vol. 112, no. 5, pp. 2222-2231, May 2008.

[37] D. Liu and S. Cai, "A Spatial-temporal modeling approach to reconstructing land-cover change trajectories from multi-temporal satellite imagery," *Ann. Assoc. Am. Geogr.,* vol. 102, no. 6, pp. 1329-1347, 2012.

[38] K. C. Mertens, B. De Baets, L. P. C. Verbeke, and R. R. de Wulf, "A sub-pixel mapping algorithm based on sub-pixel/pixel spatial attraction models," *Int. J. Remote Sens.*, vol. 27, no. 15, pp. 3293-3310, Aug. 2006.

[39] F. Ling, X. Li, Y. Du, and F. Xiao, "Sub-pixel mapping of remotely sensed imagery with hybrid intra- and inter-pixel dependence," *Int. J. Remote Sens.*, vol. 34, no. 1, pp. 341-357, Mar. 2013.

[40] Y. Chen, Y. Ge, Q. Wang, and Y. Jiang, "A subpixel mapping algorithm combining pixel-level and subpixel-level spatial dependences with binary integer programming," *Remote Sens. Lett.*, vol. 5, no. 10, pp. 902-911, Oct. 2014.

[41] Y. Makido and A. Shortridge, "Weighting function alternatives for a subpixel allocation model," *Photogramm. Eng. Remote Sens.*, vol. 73, no. 11, pp. 1233-1240, Nov. 2007.

[42] A. Boucher, P. C. Kyriakidis, and C. Cronkite-Ratcliff, "Geostatistical solutions for super-resolution land cover mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 1, pp. 272-283, Jan. 2008.

[43] S. Cai, D. Liu, D. Sulla-Menashe, and M. A. Friedl, "Enhancing MODIS land cover product with a spatial-temporal modeling algorithm," *Remote Sens. Environ.*, vol. 147, pp. 243-255, May. 2014.

[44] T. Kasetkasem, M. K. Arora, and P. K. Varshney, "Super-resolution land cover mapping using a Markov random field based approach," *Remote Sens. Environ.*, vol. 96, no. 3-4, pp. 302-314, Jun. 2005.

[45] C. Homer, J. Dewitz, L. Yang, S. Jin, P. Danielson, G. Xian, J. Coulston, N. Herold, J. Wickham, and K. Megown, "Completion of the 2011 national land cover database for the conterminous United States - representing a decade of land cover change information," *Photogramm. Eng. Remote Sens.*, vol. 81, no. 5, pp. 345-354, May 2015.

[46] J. Verhoeye and R. De Wulf, "Land cover mapping at sub-pixel scales using linear optimization techniques," *Remote Sens. Environ.*, vol. 79, no. 1, pp. 96-104, Jan. 2002.

[47] H. Jin, G. Mountrakis, and P. Li, "A super-resolution mapping method using local indicator variograms," *Int. J. Remote Sens.*, vol. 33, no. 24, pp. 7747-7773, Jul. 2012.

[48] D. A. Roberts, M. Gardner, R. Church, S. Ustin, G. Scheer, and R. O. Green, "Mapping chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models," *Remote Sens. Environ.*, vol. 65, no. 3, pp. 267-279, Sep. 1998.

[49] F. Gao, J. Masek, M. Schwaller, and F. Hall, "On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2207-2218, Aug. 2006.

[50] X. Zhu, J. Chen, F. Gao, X. Chen, and J. G.Masek, "An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions," *Remote Sens. Environ.*, vol. 114, no. 11, pp. 2610-2623, Nov. 2010.

[51] R. Zurita-Milla, J. G. P. W. Clevers, and M. E. Schdepman, "Unmixing-based Landsat TM and MERIS FR data fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 453-457, Jul. 2008.

[52] X. Chen, M. Liu, X. Zhu, J. Chen, Y. Zhong, and X. Cao, ""Blend-then-Index" or "Index-then-Blend": A theoretical analysis for generating high-resolution NDVI time series by STARFM," *Photogramm. Eng. Remote Sens.*, vol. 84, no. 2, pp. 66-74, Feb. 2018.

[53] R. Zurita-Milla, G. Kaiser, J. Clevers, W. Schneider, and M. E. Schaepman, "Downscaling time series of MERIS full resolution data to monitor vegetation seasonal dynamics," *Remote Sens. Environ.*, vol. 113, no. 9, pp. 1874-1885, Sep. 2009.

[54] R. Zurita-Milla, J. Clevers, J. A. E. Van Gijsel, and M. E. Schaepman, "Using MERIS fused images for land-cover mapping and vegetation status assessment in heterogeneous landscapes," *Int. J. Remote Sens.*, vol. 32, no. 4, pp. 973-991, Mar. 2011.

[55] M. Simoes, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3373-3388, Jun. 2015.

[56] M. T. Eismann and R. C. Hardie, "Hyperspectral resolution enhancement using high-resolution multispectral imagery with arbitrary response functions," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 455-465, Mar. 2005.

[57] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3658-3668, Jul. 2015.

Xiaodong Li received his B.S. in Geographic Information System from the China University of Geosciences, Wuhan, China in 2006. He received his M.S. and Ph.D. in Physical Geography from the Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan, China in 2009 and 2012, respectively. He is currently an associate professor at the Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan, China. He has been supported by the Youth Innovation Promotion Association CAS and the Hubei Province Natural Science Fund for Distinguished Young Scholars. His research interest is super-resolution mapping of remotely sensed imagery and time series analysis.

Feng Ling received the B.S. degree in geology and the M.S. degree in geophysical prospecting and information technology from China University of Geosciences, Wuhan, China in 1999 and 2002, respectively, and the Ph.D. degree in system analysis and integration from Huazhong University of Science and Technology, Wuhan, China in 2006.

He is currently a professor with Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan, China. His research interests include super-resolution mapping and spatial-temporal fusion of remotely sensed imagery. He has published more than 50 refereed journal articles. He received the 2016 Boeing Award for Best Scientific Paper in Image Analysis and Interpretation from the American Society for Photogrammetry and Remote Sensing (ASPRS) as the corresponding author. He has been supported by several talent programs, including the Hubei Province Natural Science Fund for Distinguished Young Scholars (2013) and the Distinguished Young Scientist Grant of the Chinese Academy of Sciences (2014).

Giles M. Foody (M'01–SM'10–F'13) earned the B.Sc. and Ph.D. degrees from the University of Sheffield, U.K., in 1983 and 1986, respectively. He is currently Professor of Geographical Information Science at the School of Geography at the University of Nottingham, U.K. His main research interests focus on the interface between remote sensing, ecology, and informatics. Topics of particular interest relate to image classification for land cover mapping and monitoring applications, addressing issues at scales ranging from the sub-pixel to global. His publication list includes 9 books and more than 200 refereed journal articles. Prof. Foody serves as founding editor-in-chief of *Remote Sensing Letters*. He holds additional editorial roles on ten other journals including *Landscape Ecology*, *International Journal of Remote Sensing*, *Remote Sensing of Environment*, *Geocarto International*, and the *International Journal of Applied Earth Observation and Geoinformation*. He has served on numerous committees and panels, notably the U.K.'s two most recent research assessment exercises: RAE2008 and REF2014. He was chair of the European Union COST Action on 'Mapping and the Citizen Sensor' (TD1202) which involves researchers from 30 countries.

Yong Ge (M'14) received the Ph.D. degree in cartography and geographical information system from the Chinese Academy of Sciences (CAS), Beijing, China, in 2001. She is a professor with the State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, CAS. Her research activity focuses on spatial data analysis and data quality assessment. She has directed research in more than ten national projects. She is the author or coauthor of over 110 scientific papers published in refereed journals, one book, and six chapters in books; she holds six granted patents in improving the accuracy of information extraction from remotely sensed imagery.

Dr. Ge has been involved in the organization of several international conferences and workshops. She is a steering committee member of International Spatial Accuracy Research Association (ISARA), and an editorial board member of Spatial Statistics (Elsevier).
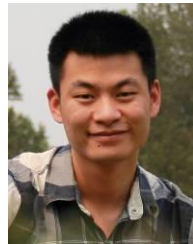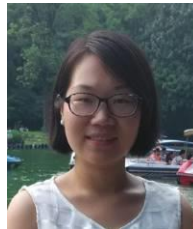
Yihang Zhang received his B.S. in Land Resource Management from the China University of Geosciences, Wuhan, China in 2012, and received his Ph.D. degrees in Physical Geography from the Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan, in 2017. His research interests are super-resolution mapping and downscaling of remotely sensed imagery.

Lihui Wang received the B.S. degree in Geographic Information System from the China University of Geosciences, Wuhan, China in 2007. He received the M.S. and Ph.D. degrees in Physical Geography from the Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan, in 2010 and 2013, respectively. He is currently a research assistant with the Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan, China. His research interests are remote sensing applications in land resources and ecological environment.

Lingfei Shi received the B.S. degree in geographic information system from Henan Polytechnic University, Jiaozuo, China, in 2013. She is currently working toward the Ph.D. degree in physical geography in the Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan, China. Her research interests include super-resolution mapping and spatio-temporal analysis based on remotely sensed imagery.

Xinyan Li received her B.S. in Land Resources Management from the China University of Geosciences, Wuhan, China in 2016. She is currently pursuing the Ph.D. degree in physical geography at the Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan, China. Her research interest is super-resolution mapping of remotely sensed imagery.
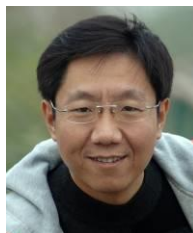
Yun Du received his B.S. in Geomorphology and Quaternary Geology from Nanjing University, Nanjing, China in 1989, and his M.S. in Physical Geography from the Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan, China in 1992. He received his Ph.D. in Historical Geography from Wuhan University, Wuhan, China in 1999. He is currently a professor at the Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan, China. His research interests are remote sensing application and hydrology.