

Journal of Applied Remote Sensing

Use of MODIS and Landsat time series data to generate high-resolution temporal synthetic Landsat data using a spatial and temporal reflectance fusion model

Mingquan Wu
Zheng Niu
Changyao Wang
Chaoyang Wu
Li Wang



Use of MODIS and Landsat time series data to generate high-resolution temporal synthetic Landsat data using a spatial and temporal reflectance fusion model

Mingquan Wu,^a Zheng Niu,^a Changyao Wang,^a
Chaoyang Wu,^{a,b} and Li Wang^a

^aInstitute of Remote Sensing Applications, Chinese Academy of Sciences, State Key Laboratory of Remote Sensing Science, Datun Street, Chaoyang District, P. O. Box 9718, Beijing 100101, China

wumq@irsa.ac.cn

^bUniversity of Toronto, Department of Geography, 100 St. George Street, Room 5047, Toronto, Ontario, Canada M5S 3G3

Abstract. Due to cloud coverage and obstruction, it is difficult to obtain useful images during the critical periods of monitoring vegetation using medium-resolution spatial satellites such as Landsat and Satellite Pour l'Observation de la Terre (SPOT), especially in pluvial regions. Although high temporal resolution sensors, such as the Advanced Very High Resolution Radiometer (AVHRR) and Moderate Resolution Imaging Spectroradiometer (MODIS), can provide high-frequency data, the coarse ground resolutions of these sensors make them unsuitable to quantify the vegetation growth processes at fine scales. This paper introduces a new data fusion model for blending observations of high temporal resolution sensors (e.g., MODIS) and moderate spatial resolution satellites (e.g., Landsat) to produce synthetic imagery with both high-spatial and temporal resolutions. By detecting temporal change information from MODIS daily surface reflectance images, our algorithm produced high-resolution temporal synthetic Landsat data based on a Landsat-7 Enhanced Thematic Mapper Plus (ETM+) image at the beginning time (T_1). The algorithm was then tested over a $185 \times 185 \text{ km}^2$ area located in East China. The results showed that the algorithm can produce high-resolution temporal synthetic Landsat data that were similar to the actual observations with a high correlation coefficient (r) of 0.98 between synthetic imageries and the actual observations. © 2012 Society of Photo-Optical Instrumentation Engineers (SPIE). [DOI: 10.1117/1.JRS.6.063507]

Keywords: Landsat; MODIS; data fusion; synthetic imagery.

Paper 11048 received Mar. 21, 2011; revised manuscript received Jan. 1, 2012; accepted for publication Jan. 5, 2012; published online Mar. 7, 2012.

1 Introduction

The Moderate Resolution Imaging Spectroradiometer (MODIS) onboard the Terra and Aqua satellites is an example of a high-temporal resolution sensor but with moderate spatial resolutions varying from 250 m to a few kilometers. These sensors are used to extract time series of vegetation indices (VIs), such as the Simple Ratio (SR), Normalized Different Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), and MERIS Terrestrial Chlorophyll Index (MTCI). Such VIs are used to monitor spatial and temporal dynamics of vegetation biophysical variables, including Leaf Area Index (LAI),¹⁻⁴ fraction of absorbed photosynthetically active radiation (fAPAR), foliar pigment content,^{5,6} and land cover changes.⁷⁻¹² However, when land surfaces exhibit large spectral variability at scales smaller than the resolution of the satellite data, mixed pixel issues will arise. The signals recorded by those remote sensors are usually a mixture of several types of land use with different spectrums and these images are often composed of a mixture of different components. Therefore, there is a pressing need to enhance the spatial resolution of these productions to conduct fine-scale applications.

Data from moderate-resolution ground satellites, such as Landsat and Satellite Pour l'Observation de la Terre (SPOT), are widely used for extracting VIs, and monitoring land cover.^{13–16} However, it is very difficult to obtain useful images at critical vegetation periods due to cloud coverage or duty cycle limitations.¹⁷ The probability of acquiring cloud-free Landsat images for a given year (cloud cover <10%) can be as low as 10%.¹⁸

A solution for fine-scale vegetation research is to blend the data from both high-temporal resolution sensors (e.g., MODIS) and moderate ground resolution satellites (e.g., Landsat) to generate synthetic observations with characteristics of both.¹⁹ In recent decades, several approaches have been proposed to enhance the temporal frequency of high-resolution spatial satellite observations.

Neglecting both geolocation errors and differences in sensor systems which are considered to be different in atmospheric correction and sensors (e.g., bandwidth, acquisition time, and spectral response function), the surface reflectance $R(i, t)$ of a coarse resolution pixel at time t can be aggregated from fine resolution pixels of surface reflectance $r(j, t)$ based on

$$R(i, t) = \sum_{j=1}^n r(j, t), \quad (1)$$

where n is the number of fine resolution pixels. Since there are n unknowns in Eq. (1), it is ill-posed mathematically, and a unique solution cannot be determined without additional information. Land cover type is the most commonly used ancillary information, which assumes that the spectral properties of each land cover class are constant.

According to unmixing theory, the reflectance of a coarse-resolution spatial pixel measured by a sensor is a composite value. The response of each coarse spatial resolution pixel is assumed to be a linear combination of the responses of each land-cover class contributing to the mixture.²⁰ The coarse spatial reflectance $R(i, t)$ of the landscape thus consists of k discrete land-cover class c , weighted by their class fractional cover as

$$R(i, t) = \sum_{c=0}^k f_c(i, c) \times \bar{r}(c, t) + \xi(i, t) \quad (2)$$

$$\text{Constrained : } \sum_{c=0}^k f_c(i, c) < 1; \quad \text{and} \quad f_c(i, c) \geq 0 \text{ for all,}$$

where $f_c(i, c)$ is the fractional cover of class c in pixel i , which is usually assumed not to change over time, $\bar{r}(c, t)$ is the mean reflectance of class c , and $\xi(i, t)$ is the residual error term. Since the pixels belong to the same land-cover class having the similar reflectance, it can be assumed that the surface reflectance of each land-cover class is constant within the whole scene. Thus, surface reflectance $r(j, t)$ of fine resolution homogeneous pixels belonging to land-cover class c were deemed to be equal with the mean reflectance $\bar{r}(c, t)$ of class c . Ordinary least squares techniques can be used to estimate the mean reflectance of classes by Eq. (2). The fractional cover $f(i, c)$ is usually obtained from higher resolution spatial land-cover maps.^{21–24}

However, in many cases, the model assumption is violated since the surface reflectance can vary spatially. As a result, the true reflectance of a land-cover class in a selected pixel can be substantially different from the mean value estimated by Eq. (2). The reflectance of a land-cover class can be related to differences in environmental factors (e.g., altitude, morphology, soil type).²⁵ Therefore, previous studies have developed methods to address the problem of land cover spatial variability in the disaggregation of remote sensing images.^{25–28} These methods involved two steps: 1. identification of a subset S of coarse resolution pixels to be used for the disaggregation of the surface reflectance of the target coarse resolution pixel and 2. the application of Eq. (2) (or a variant of the equation) to determine the target coarse resolution pixel to estimate disaggregated reflectance in the subset S . Zhukov et al.²⁶ proposed a method that only considered the pixels contained within a window of a specific size centered on the target. In contrast, Maselli²⁷ suggested to weight the different coarse resolution pixels using an inverse exponential function of their Euclidean spatial distance to the target, so that the farther a

pixel is from the target, the less it influences the disaggregation process. Both studies were based on the assumption that the surface reflectance of a land-cover class does not show great variations between pixels. Lorenzo et al.²⁵ developed a new method based on the work of Zhukov et al.²⁶ and Maselli,²⁷ which selected and weighted pixels on the basis of their Euclidean distances from the target pixel and also on the basis of the spectral similarity of their subcomponents with those of targets. Gao et al.²⁸ introduced the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM), which selected and weighted pixels based on Euclidean distance, spectral and temporal similarity with the target pixel. Roy et al.²⁹ used a semi-physical fusion approach that used the MODIS BRDF/Albedo land surface characterization product and Landsat ETM+ data to predict ETM+ reflectance based on the assumption that the temporal reflectance variation of MODIS reflectance is similar to the variation of Landsat ETM+ reflectance. Hilker et al.³⁰ introduced a data fusion model to produce synthetic imagery and the detection of changes, called Spatial Temporal Adaptive Algorithm for mapping Reflectance Change (STAARCH). Liu et al.³¹ used STARFM method to produce time-series synthetic reflectance images for deriving urban environmental variables such as normalized difference vegetation index, normalized difference water index, and land-surface temperature for public health studies. Walker et al.³² demonstrated the feasibility of using the STARFM algorithm to assemble an imagery time series at Landsat spatial resolution and MODIS temporal resolution in vegetated dry-land ecosystems.

Although these methods are used to identify the subset S of coarse resolution pixels and they can reduce the difference between the true reflectance of a selected pixel and the mean value estimated by Eq. (2) to a certain extent, the difference cannot be eliminated completely due to the heterogeneity of land surface. Thus, the reflectance estimated by these proposed methods is the mean value \bar{r} of the land-cover class in the whole image or an identified subset S , which is not equal to the real surface reflectance r of fine resolution pixels. Additionally, these proposed models leave the spatial details within the coarse resolution pixels unresolved in the decomposed mean value \bar{r} .

To generate real surface reflectance r of a fine resolution pixel, Wu et al.³³ introduced a Landsat TM image simulation method by considering the temporal change information of each land-cover class which was not considered in the linear mixed model. However, this method was based on the original linear mixed model which does not consider the spatial variation. Zhu et al.³⁴ introduced an enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions which also considered the temporal change information. However, the reflectance change of each class was supposed as linear from the beginning to the end, which was not always true considering the temporal change feature of vegetation.

In this paper, our aim was to generate the real surface reflectance r of a fine resolution pixel considering both the spatial variation and the nonlinear temporal change information. To achieve this, we proposed and validated a novel spatial and temporal fusion model to generate synthetic image with both high temporal and spatial resolutions. The Spatial Temporal Data Fusion Approach (STDFA) is based on the assumption that the temporal variation properties of each land-cover class are constant. Under this assumption, we can generate the surface reflectance r of each fine resolution image through a Surface Reflectance Calculation Model (SRCM). This model incorporates inputs from a previous fine resolution image and temporal change information detected from temporal coarse resolution images using an improved linear mixed model which considers the spatial variation. Through the SRCM model, the STDFA algorithm can produce surface reflectance r values at very fine resolution and can dramatically improve the accuracy of the results.

The STDFA algorithm was applied to daily 500 m MODIS surface reflectance productions, and multi-temporal Landsat-7 ETM+ images were used for classification, estimation of surface reflectance r of fine-resolution pixel from mean values of \bar{r} , and accuracy assessment.

2 Approach

The input of the STDFA algorithm includes two scenes of Landsat images and a time series of MODIS from the beginning to the end of the observation period. The main output of this method

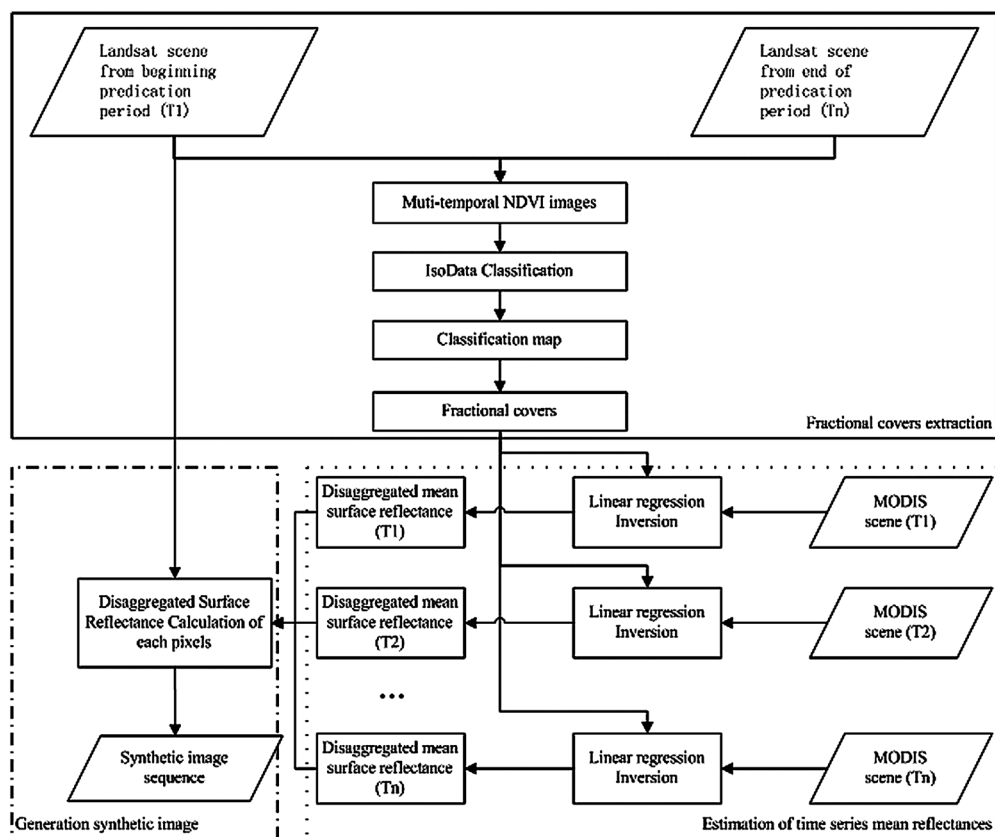


Fig. 1 Flow chart of the STDA algorithm. Processing steps of the three main blocks are explained in Secs. 2.1–2.3.

is a sequence synthetic image with spatial resolution similar to the fine resolution image (Landsat ETM+) and temporal resolution similar to MODIS. The algorithm includes three steps: 1. Mapping medium resolution spatial Landsat-7 ETM+ images by the IsoData classification method to extract fractional covers of each coarse MODIS pixels, 2. Extracting time series mean reflectance \bar{r} by solving Eq. (2) using ordinary least squares techniques with inputting time series daily 500 m MODIS surface reflectance images and fractional covers, and 3. Calculating surface reflectance r using an SRCM by inputting time series mean reflectance values \bar{r} and a previous fine resolution image. The flow chart of the algorithm is shown in Fig. 1.

2.1 Fractional Cover Extraction

Similar to the traditional spatial and temporal fusion method, fractional covers are extracted from high resolution classification map. However, in order to satisfy the assumption that the temporal variation properties of each Landsat ETM+ pixels in the same class are constant, a new classification method considering the temporal change information of different dates was used. Two Landsat ETM+ images were used in the classification. First, we obtained two NDVI images from Landsat ETM+ images at time T_1 and T_n . Second, a classification map was produced with the IsoData classification method of multi-NDVI images. Due to fact that the multi-NDVI images were acquired at multi-temporal images, this classification method is based on the temporal change information, indicating that the pixels belong to the same class have the same temporal changes.

Finally, we extracted fractional covers of classes within MODIS pixels using ARCGIS 9.2 by intersecting the classification map with the 480 m grid and rationing the area occupied by each class in the different pixels to the total pixel area.⁶ Fractional cover values were stored in a different raster image at MODIS spatial resolution for each class.

2.2 Estimation of Time Series Mean Reflectance

Using fractional cover values and surface reflectance of each MODIS pixel as the input parameters, Eq. (2) can be solved with the ordinary least squares technique and outputting the disaggregated mean surface reflectance value \bar{r} for each class. By inputting time series MODIS surface reflectance products from the beginning time (T_1) to the end time (T_n) and fractional cover values (which are assumed to not vary between dates), time series mean reflectance values can be obtained by solving the linear system of Eq. (2). These values can be used to calculate the surface reflectance r of each fine resolution pixel.

However, due to spatial variation, the solution of Eq. (2) was conducted in a rectangular window centered to the MODIS target pixel. Meanwhile, this procedure also removes the impact of geolocation errors associated between the coarse resolution images and the fine resolution images. To select a suitable window, rectangles of different lengths were tested in a pre-study stage which was conducted in Landsat7-ETM+ band4. First, synthetic images were produced using windows at different lengths. Then, several indicators like the coefficient (r), variance, mean absolute difference (MAD), bias, RMSE between the real Landsat images and the synthetic image were calculated to determine the best length n of the window (Table 1). Through testing the length of the window size from 5 to 60 MODIS pixels with 5 or 10 pixels for sample intervals, we identified that a 40×40 MODIS pixels window was the most suitable area for the solution of Eq. (2) as it had the highest coefficient r between the real Landsat image and the synthetic image.

2.3 Surface Reflectance Calculation

Since Eq. (2) only provides the mean values for different classes, further calculations are required to determine the reflectance of each fine resolution pixel. This can be accomplished by SRCM. The model is based on the assumption that the temporal variation properties of each Landsat ETM+ pixels in the same class are constant.

Time series surface reflectance of fine resolution pixels a and b of class c can be expressed as the vector $r_{a,c} = (r_{a,c,t_1}, r_{a,c,t_2}, \dots, r_{a,c,t_n})^T$ and $r_{b,c} = (r_{b,c,t_1}, r_{b,c,t_2}, \dots, r_{b,c,t_n})^T$, where r_{a,c,t_i} is the surface reflectance of pixel a at time t_i and r_{b,c,t_i} is the surface reflectance of pixel b at time t_i . The time comparability factors for pixels a and b are

$$\text{TCF}_{(a,c)} = \frac{\sum_{i=1}^n r_{a,c,t_i}}{n} \quad (3)$$

$$\text{TCF}_{(b,c)} = \frac{\sum_{i=1}^n r_{b,c,t_i}}{n} \quad (4)$$

The shape factors of pixels a and b at time t_i are

$$\text{SF}_{a,c,t_i} = r_{a,c,t_i} - \text{TCF}_{(a,c)}, \quad (5)$$

$$\text{SF}_{b,c,t_i} = r_{b,c,t_i} - \text{TCF}_{(b,c)}. \quad (6)$$

Table 1 A comparison between the observed surface reflectance of Landsat-7 ETM+ Band4 and synthetic imageries produced with different window size.

Window size (MODIS pixels)	r	Variance	MAD	RMSE	Bias
5	0.9763	0.0063	0.0487	0.0795	-0.0019
10	0.9788	0.0057	0.0478	0.0760	-0.0066
20	0.9789	0.0057	0.0470	0.0754	-0.0053
30	0.9789	0.0057	0.0483	0.0762	-0.0089
40	0.9804	0.0054	0.0456	0.0736	-0.0068
50	0.9788	0.0057	0.0476	0.0759	-0.0066
60	0.9755	0.0067	0.0544	0.0828	-0.0135

Under the assumption that the temporal variation properties of the Landsat ETM+ pixels in the same class c are constant, we can calculate: $SF_{(a,c,t_i)} = SF_{(b,c,t_i)}$, and $r_{a,c,t_j} - r_{a,c,t_i} = r_{b,c,t_j} - r_{b,c,t_i}$.

Since the mean reflectance $\bar{r}(c, t_i)$ of class c at time t_i is the mean value of pixels that belong to class c

$$\bar{r}(c, t_i) = \frac{1}{m} \sum_{k=1}^m r(k, c, t_i), \quad (7)$$

where m is the number of pixels belonging to class c . Thus, from time t_i to time t_j , the variation $\Delta\bar{r}(c, t_j - t_i)$ of mean reflectance of class c is equal to the variation $\Delta r(k, c, t_j - t_i)$ of pixel k :

$$\Delta\bar{r}(c, t_j - t_i) = \Delta r(k, c, t_j - t_i) \quad (8)$$

or

$$\bar{r}(c, t_j) - \bar{r}(c, t_i) = r(k, c, t_j) - r(k, c, t_i). \quad (9)$$

Since $\bar{r}(c, t_1)$ and $\bar{r}(c, t_j)$ can be calculated by Eq. (2) from time t_1 to time t_j , and $r(k, c, t_1)$ can be obtained from the Landsat scene at the initial time, we can calculate $r(k, c, t_j)$ with Eq. (9). Equation (9) is the SRCM.

3 Algorithm Testing and Validation

3.1 Study Area

The study area is located at Jiangsu Province, China (Fig. 2). The area is covered by a scene of Landsat ETM+ image (path 120, row 38).

The study area is located in the middle-lower Yangtze Plain and has hills in the west and south and some cities in the north regions. The climate is a typical humid subtropical monsoon

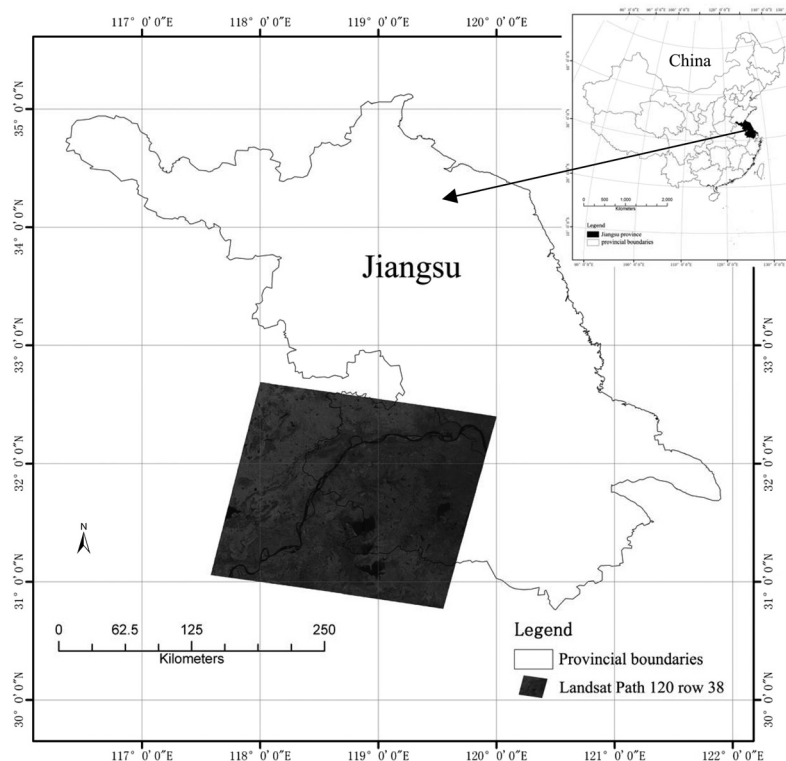


Fig. 2 Location of the study area.

climate with a mean annual temperature of 15.3 °C, an annual precipitation of 1106.5 mm, and rainy season from late June or early July. The flat areas in the region are mainly devoted to agriculture, with a prevalence of winter wheat, rapeseed plant, and paddy fields. The rugged areas in the region are mainly covered by subtropical evergreen forests.

This area is a complex heterogeneous region with diverse land cover types, including water, farmland, forest, and city, and the plots in this area are very cracked which are usually less than 1000 m².

3.2 Landsat-7 ETM+ Data and Preprocessing

Three Landsat-7 ETM+ images acquired in 2002 were used in this study (Table 2). All of the images were acquired with clear sky conditions, except the one acquired on October 24 which had a little cloud in the southeast. The images were atmospherically corrected using the 6S radiative transfer code. Atmospherically corrected images were then georeferenced using a second-order polynomial warping approach. This approach was based on the selection of an appropriate number of Ground Control Points (GCPs) using a 1:10,000 topographic map by the nearest neighbor resampling method with the position error within 0.5 for Landsat-7 ETM+ pixels.

3.3 MODIS Data and Preprocessing

Three daily MODIS surface reflectance products (MOD09GA, 500 m) obtained in clear sky conditions were used in this study (Table 2). Those MODIS products were acquired at the same time as the Landsat-7 ETM+ data to minimize spectral differences caused by temporal difference. The MOD09GA product has six bands at 500 m spatial resolution (Table 3). All MODIS images were reprojected from the native Sinusoidal projection to a UTM-WGS84 reference system and were resized to select the study area using MODIS Reprojection Tool (MRT) software. We also resized the spatial resolution from 500 m to 480 m with a nearest neighbor resampling method in MRT. All MODIS data were then georeferenced by a second-order polynomial warping approach based on the selection of an appropriate number of GCPs on

Table 2 The main characteristics of the three Landsat ETM+ images and MODIS products used in the STDFA algorithm.

Landsat7-ETM+			MODIS	
Acquisition date	Path/row	Main usage	Acquisition date	Main usage
10/08/2002	120/38	Classification	10/08/2002	Estimation of time series mean reflectance
10/24/2002	120/38	Validation	10/24/2002	
11/09/2002	120/38	Classification	11/09/2002	

Table 3 Landsat ETM+ bandwidth and MODIS bandwidth.

Landsat ETM+		MODIS	
Band	Bandwidth (nm)	Band	Bandwidth (nm)
1	450 to 520	3	459 to 479
2	530 to 610	4	545 to 565
3	630 to 690	1	620 to 670
4	780 to 900	2	841 to 876
5	1550 to 1750	6	1628 to 1652
7	2090 to 2350	7	2105 to 2155

simulated MODIS images with a nearest neighbor resampling method. The simulated MODIS images were simulated from georeferenced Landsat ETM+ images with the pixel aggregate resampling method, and the position error was within 0.7 MODIS pixels.

3.4 Results and Validation

The STDFA model was applied to two Landsat-7 ETM+ images acquired on October 8, and November 9, 2002. One synthetic image with six bands at 30 m spatial resolution on October 24, 2002 was outputted by the STDFA model.

Figure 3(a) shows the actual observation of Landsat ETM+ band4 surface reflectance and Fig. 3(b) shows the synthetic surface reflectance imagery for the near infrared band. Obviously, the actual observation and synthetic imagery of Landsat ETM+ band4 is so similar that it is hard to discriminate each one from the other by visual interpretation.

From the histogram of bands, the reflectance of all bands followed a Gaussian distribution with the peak value at 0.28 in Landsat ETM+ band4. Therefore, correlation analysis was used to quantitatively evaluate the similarity between the actual observations and synthetic imagery. Several indicators were calculated, including the coefficient (r), variance, mean absolute difference (MAD), bias, RMSE, and p -value (Table 4). Figure 4 shows the scatter plots between the actual observations and synthetic imagery of six Landsat-7 ETM+ bands, which fit well to the 1:1 line. These results imply that there is high correlation between the actual observations and synthetic imagery.

Table 4 A comparison between the observed surface reflectance of Landsat-7 ETM+ data and synthetic imagery.

ETM+ band	r	Variance	MAD	RMSE	Bias	p -value
1	0.8875	0.0246	0.1038	0.1579	0.0187	0.00008
2	0.9451	0.0129	0.0765	0.1156	0.0204	<0.00001
3	0.9363	0.0135	0.0813	0.1203	0.0310	0.00060
4	0.9804	0.0054	0.0456	0.0736	-0.0068	0.00001
5	0.9548	0.0114	0.0710	0.1108	0.0291	0.00002
7	0.9318	0.0157	0.0943	0.1432	0.0695	<0.00001

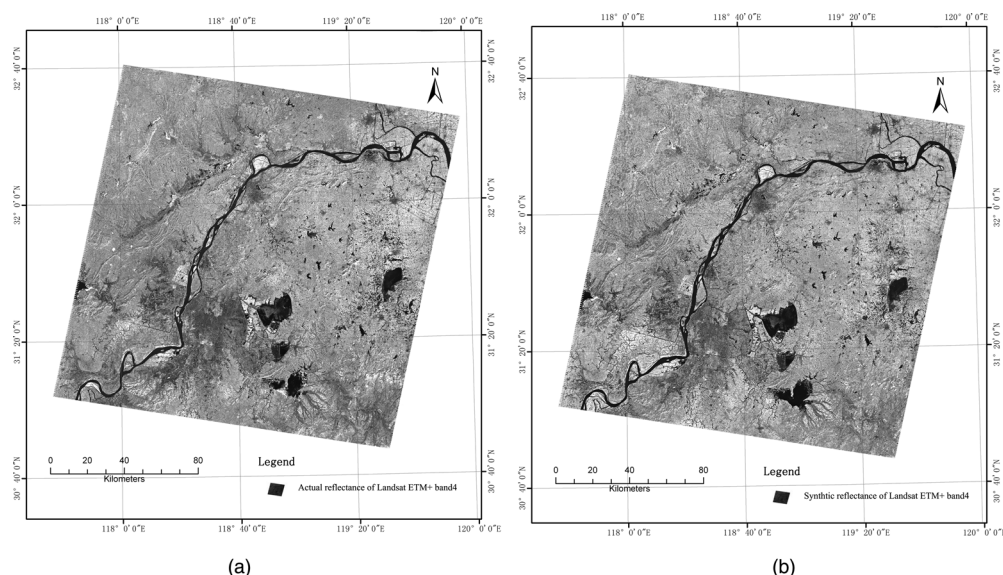


Fig. 3 Actual observation (a) and synthetic imagery (b) on October 24, 2002 for Landsat ETM+ band4.

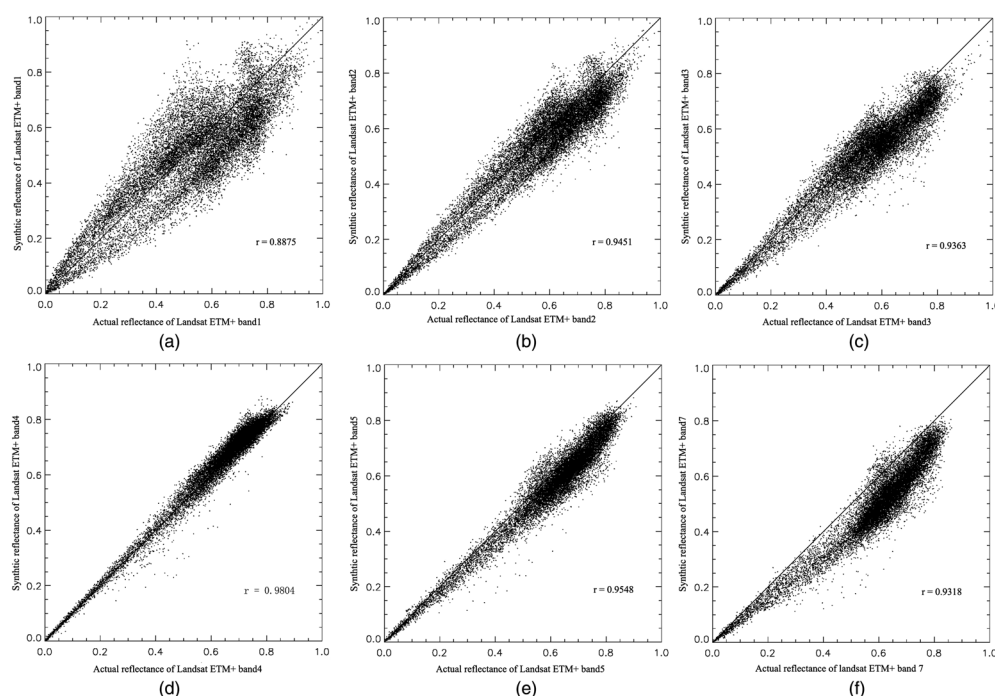


Fig. 4 The scatter plots of the actual observation and synthetic imagery of six Landsat-7 ETM+ bands.

3.5 Discussion

The correlation analysis showed that the STDFA algorithm proposed in this paper can obtain high spatial and temporal resolution images from coarse resolution images and two-fine resolution images.

To demonstrate that the new STDFA algorithm can well capture the temporal change information, the traditional spatial and temporal fusion method³³ was also tested in this study area. Table 5 shows the correlation analysis of traditional STFM. Table 5 suggests that the new STDFA algorithm has a higher correlation coefficient r than that of the traditional method.

The SRCM model, which estimates surface reflectance at a fine resolution, ensures the strong results of the algorithm. Essentially, the STDFA algorithm simulates fine-resolution images by using the temporal change information for each fine-resolution pixel. After the temporal change information was detected from sequence coarse resolution images, the reflectance of each fine resolution pixel could be calculated from the fine-resolution image at the beginning time T_1 . Furthermore, this new STDFA algorithm can capture nonlinear temporal change information which has been ignored in the conditional STFM. While most conventional methods can only calculate the mean reflectance of each class, our algorithm can obtain the reflectance for each fine resolution pixel.

Table 5 A comparison between the observed surface reflectance of Landsat-7 ETM+ data and synthetic imagery using STFM method.

ETM+ band	r	Variance	MAD	RMSE	Bias	p -value
1	0.7121	0.0003	0.0135	0.0196	0.0093	0.0501
2	0.7956	0.0004	0.0140	0.0211	0.0072	0.0362
3	0.8915	0.0004	0.0153	0.0195	0.0016	0.0082
4	0.8974	0.0009	0.0224	0.0348	-0.0183	0.0250
5	0.8691	0.0013	0.0283	0.0399	-0.0182	0.02380
7	0.8298	0.0014	0.0302	0.0400	-0.0145	0.0564

Although we have observed good results using the proposed method, there are still issues that can affect the precision. Those issues include factors related to the sensors, disaggregation method, method assumptions, and image quality.

Factors related to the sensors may have influenced the precision of the method. These factors include differences of bandwidth, acquisition time, spectral response functions (Table 3), geolocation errors, and differences in atmospheric corrections. To evaluate the influences of those differences in sensor systems, a prestudy was conducted in a $58 \times 63 \text{ km}^2$ randomly selected subset area of Landsat ETM+ scene path 120, row 38. In this prestudy stage, instead of using the actual MODIS images, simulated MODIS reflectance images were input into the STDFA. Those simulated MODIS reflectance images were produced by degrading Landsat-7 ETM+ images that were atmospherically corrected and georeferenced with the pixel aggregate resampling method. Because the simulated MODIS reflectance images were produced by Landsat-7 ETM+ images, there would be no difference in sensor systems between the simulated coarse resolution images and Landsat-7 ETM+ images. Through correlation analysis of the synthetic images produced by inputting simulated MODIS reflectance images into STDFA, we found that the correlation coefficient r between synthetic image produced by inputting simulated MODIS images and actual Landsat ETM+ image was 0.946 in near infrared band, while inputting actual MODIS images and Landsat ETM+ image would give 0.939. Therefore, the differences in sensor systems will only lead to a 0.01 reduce in correlation coefficient r .

The disaggregation method may also have influenced the precision of this method. Although temporal information and spatial variability were considered, the residual error in disaggregation still could not be removed completely. There were some outliers due to the disaggregation. If those outliers were not deleted, the correlation coefficient r was 0.9405 in near infrared band. However, the correlation will be improved by excluding outliers with r of 0.9804. Thus, the residual error in disaggregation can lead to a 0.0399 decrease in correlation coefficient r .

This method is based on two assumptions. One is that the temporal variation properties of each class are constant within the rectangular window used in the algorithm. The other is the fractional covers of classes do not change from the beginning time to the end, which is the basic assumption of all temporal and spatial fusion methods. So those two assumptions require an unchanged land-use classes, which, regrettably, is very difficult to satisfy when applied in a large area.

Additionally, the linear mixture model may have changed the precision in this study. Most disaggregation methods are based on a linear mixture model, where the spectral signature of a mixed pixel (such as MODIS pixels) is assumed to be the linear mixture of the end members. This is generally the case when the surface consists of snow³⁵ or bare soil, but is not always true when spectral mixing of vegetation end members occurs. Only a part of the radiation is directly reflected (dependent on the wavelength), and the rest is absorbed or transmitted to other layers, from which it can be reflected to the sensor again. As a result, the reflected radiation interacts with more than one object, resulting in nonlinear spectral mixing.^{35–38} Nonlinear spectral mixing is common when vegetation is assumed to occur in the study area. Thus, a nonlinear spectral mixing model is an important direction where future research should focus.

Lastly, image quality can also influence precision. In this study, the correlation coefficient (r) between disaggregated synthetic surface reflectance imagery and actual observation of Landsat-7 ETM+ images in band1 was systematically lower than other bands. Similar results were also found for other spatial and temporal fusion methods.^{30,33} This may be caused by the random noise in images because of the low signal-to-noise ratio in this band compared to other bands.

4 Conclusions

A new method for the determination of spatially sub-pixelated surface reflectance from multi-temporal coarse resolution and two fine-resolution, remote-sensing images was developed and tested on daily MODIS surface reflectance images and two Landsat7 ETM+ images in east China. Disaggregated synthetic surface reflectance imageries were calculated for each fine-resolution pixel by three steps: 1. High-spatial resolution land use map was produced from time series finer-resolution images using the IsoData classification method to extract

fractional covers of each coarse pixels. 2. The mean surface reflectance \bar{r} estimation for each class was calculated by solving linear system equations based on linear unmixing model from time series MODIS daily surface reflectance images. 3. Surface reflectance r was calculated for each fine resolution pixel from the mean surface reflectance \bar{r} based on the assumption that the temporal variation properties of each pixel belonging to the same class are constant. Validation of these results was accomplished by the comparison of disaggregated synthetic surface reflectance imageries to actual observations of Landsat-7 ETM+ images.

By considering the nonlinear temporal change similarities and spatial variations in the disaggregation, the algorithm provided good estimates of subpixel surface reflectance values, with correlation coefficients r ranging from 0.88 to 0.98 between disaggregated synthetic surface reflectance imageries and actual observations of Landsat-7 ETM+ images.

In summary, this study demonstrates that the proposed algorithm can be efficiently used for the disaggregation of time series of coarse resolution images, given that two high spatial-resolution images are collected (one at the beginning and the other at the end of the period). The introduction of temporal change information in the disaggregation and the use of the SRCM model, which estimates each fine resolution pixel surface reflectance r from time series mean surface reflectance \bar{r} and high spatial resolution image at the beginning time (T_1), ensured accurate estimation of subpixel surface reflectance time series.

The disaggregated surface reflectance temporal profiles estimated by the algorithm can be used for very precise monitoring of the seasonal cycle of vegetation. It is possible to determine the beginning, the duration, and the amplitude of the growth phase of vegetation with the high spatial resolution and high temporal resolution data produced by the algorithm. Thus, this algorithm can improve the accuracy of several applications, such as land-cover mapping and environmental modeling applications. From these applications, specific parameters (e.g., NDVI or Leaf Area Index) can be derived from surface reflectance.

Finally, although the algorithm has been developed on MODIS 500 daily surface reflectance product and Landsat7 ETM+ images, it can be applied to images acquired by other coarse spatial resolution satellite sensors, such as MERIS, AVHRR, or SPOT VEGETATION, and other medium resolution spatial satellite sensors (e.g., SPOT), to produce subpixel surface reflectance time series at a regional scale.

Acknowledgments

This work was supported by the Major State Basic Research Development Program of China (2010CB950603), the National Natural Science Foundation of China (41001209), and the State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing Applications of Chinese Academy of Sciences.

References

1. A. Verger, F. Baret, and M. Weiss, "A multisensor fusion approach to improve LAI time series," *Rem. Sens. Environ.* **115**(10), 2460–2470 (2011), <http://dx.doi.org/10.1016/j.rse.2011.05.006>.
2. X. Xiao et al., "Quantitative relationships between field-measured Leaf Area Index and vegetation index derived from VEGETATION images for paddy rice fields," *Int. J. Remote Sens.* **23**(18), 3595–3604 (2002), <http://dx.doi.org/10.1080/01431160110115799>.
3. Q. Wang et al., "On the relationship of NDVI with leaf area index in a deciduous forest site," *Rem. Sens. Environ.* **94**(2), 244–255 (2005), <http://dx.doi.org/10.1016/j.rse.2004.10.006>.
4. D. Feng et al., "Algorithm for global leaf area index retrieval using satellite imagery," *IEEE Trans. Geosci. Rem. Sens.* **44**(8), 2219–2229 (2006), <http://dx.doi.org/10.1109/TGRS.2006.872100>.
5. C. Wu, Z. Niu, and S. Gao, "The potential of the satellite derived green chlorophyll index for estimating midday light use efficiency in maize, coniferous forest and grassland," *Ecol. Indic.* **14**(1), 66–73 (2012), <http://dx.doi.org/10.1016/j.ecolind.2011.08.018>

6. C. Wu, J. M. Chen, and N. Huang, "Predicting gross primary production from the enhanced vegetation index and photosynthetically active radiation: evaluation and calibration," *Rem. Sens. Environ.* **115**(12), 3424–3435 (2011), <http://dx.doi.org/10.1016/j.rse.2011.08.006>.
7. P. S. Thenkabail, M. Schull, and H. Turrall, "Ganges and Indus river basin land use/land cover (LULC) and irrigated area mapping using continuous streams of MODIS data," *Rem. Sens. Environ.* **95**(3), 317–341 (2005), <http://dx.doi.org/10.1016/j.rse.2004.12.018>.
8. Z. Xia et al., "Land cover classification of the North China Plain using MODIS EVI time series," *ISPRS J. Photogramm. Rem. Sens.* **63**(4), 476–484 (2008), <http://dx.doi.org/10.1016/j.isprsjprs.2008.02.005>.
9. H. Carrão, P. Gonçalves, and M. Caetano, "Contribution of multispectral and multitemporal information from MODIS images to land cover classification," *Rem. Sens. Environ.* **112**(3), 986–997 (2008), <http://dx.doi.org/10.1016/j.rse.2007.07.002>.
10. B. P. Salmon et al., "The use of a Multilayer Perceptron for detecting new human settlements from a time series of MODIS images," *Int. J. Appl. Earth Obs. Geoinf.* **13**(6), 873–883 (2011), <http://dx.doi.org/10.1016/j.jag.2011.06.007>.
11. D. Wang et al., "Spatio-temporal pattern analysis of land use/cover change trajectories in Xihe watershed," *Int. J. Appl. Earth Obs. Geoinf.* **14**(1), 12–21 (2011), <http://dx.doi.org/10.1016/j.jag.2011.08.007>.
12. Z. Li and J. M. Fox, "Mapping rubber tree growth in mainland Southeast Asia using time-series MODIS 250 m NDVI and statistical data," *Appl. Geogr.* **32**(2), 420–432 (2011), <http://dx.doi.org/10.1016/j.apgeog.2011.06.018>.
13. T. A. Schroeder et al., "Mapping wildfire and clear cut harvest disturbances in boreal forests with Landsat time series data," *Rem. Sens. Environ.* **115**(6), 1421–1433 (2011), <http://dx.doi.org/10.1016/j.rse.2011.01.022>.
14. Y. Julien, J. A. Sobrino, and J.-C. Jiménez-Muñoz, "Land use classification from multitemporal Landsat imagery using the Yearly Land Cover Dynamics (YLCD) method," *Int. J. Appl. Earth Obs. Geoinf.* **13**(5), 711–720 (2011), <http://dx.doi.org/10.1016/j.jag.2011.05.008>.
15. J. Aguirre-Gutiérrez, A. C. Seijmonsbergen, and J. F. Duivenvoorden, "Optimizing land cover classification accuracy for change detection, a combined pixel-based and object-based approach in a mountainous area in Mexico," *Appl. Geogr.* **34**, 29–37 (2012), <http://dx.doi.org/10.1016/j.apgeog.2011.10.010>.
16. M. A. Wulder et al., "Landsat continuity: issues and opportunities for land cover monitoring," *Rem. Sens. Environ.* **112**(3), 955–969 (2008), <http://dx.doi.org/10.1016/j.rse.2007.07.004>.
17. J. C. Ju and D. P. Roy, "The availability of cloud-free Landsat ETM plus data over the conterminous United States and globally," *Rem. Sens. Environ.* **112**(3), 1196–1211 (2008), <http://dx.doi.org/10.1016/j.rse.2007.08.011>.
18. D. Leckie, "Advances in remote sensing technologies for forest survey and management," *Can. J. For. Res.* **20**(4), 464–483 (1990), <http://dx.doi.org/doi:10.1139/x90-063>.
19. R. S. Lunetta et al., "North American landscape characterization dataset development and fusion issues," *Photogramm. Eng. Remote Sens.* **64**(8), 821–829 (1998).
20. J. J. Settle and N. A. Drake, "Linear mixing and the estimation of groundcover proportion," *Int. J. Remote Sens.* **14**(6), 1159–1177 (1993), <http://dx.doi.org/10.1080/01431169308904402>.
21. K. W. Oleson et al., "Unmixing multiple land cover type reflectances from coarse spatial resolution satellite data," *Rem. Sens. Environ.* **54**(2), 98–112 (1995), [http://dx.doi.org/10.1016/0034-4257\(95\)00100-F](http://dx.doi.org/10.1016/0034-4257(95)00100-F).
22. F. Maselli, M. A. Gilabert, and C. Conese, "Integration of high and low resolution NDVI data for monitoring vegetation in Mediterranean environments," *Rem. Sens. Environ.* **63**(3), 208–218 (1998), [http://dx.doi.org/10.1016/S0034-4257\(97\)00131-4](http://dx.doi.org/10.1016/S0034-4257(97)00131-4).
23. W. Lei and S. Uchida, "Use of linear spectral mixture model to estimate rice planted area based on MODIS data," *Rice Sci.* **15**(2), 131–136 (2008), [http://dx.doi.org/10.1016/S1672-6308\(08\)60031-1](http://dx.doi.org/10.1016/S1672-6308(08)60031-1).

24. D. B. Lobell and G. P. Asner, "Cropland distributions from temporal unmixing of MODIS data," *Rem. Sens. Environ.* **93**(3), 412–422 (2004), <http://dx.doi.org/10.1016/j.rse.2004.08.002>.
25. L. Busetto, M. Meronib, and R. Colombo, "Combining medium and coarse spatial resolution satellite data to improve the estimation of subpixel NDVI time series," *Rem. Sens. Environ.* **112**(1), 118–131 (2008), <http://dx.doi.org/10.1016/j.rse.2007.04.004>.
26. B. Zhukov et al., "Unmixing-based multisensor multiresolution image fusion," *IEEE Trans. Geosci. Rem. Sens.* **37**(3), 1212–1226 (1999), <http://dx.doi.org/10.1109/36.763276>.
27. F. Maselli, "Definition of spatially variable spectral endmembers by locally calibrated multivariate regression analyses," *Rem. Sens. Environ.* **75**(1), 29–38 (2001), [http://dx.doi.org/10.1016/S0034-4257\(00\)00153-X](http://dx.doi.org/10.1016/S0034-4257(00)00153-X).
28. F. Gao et al., "On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance," *IEEE Trans. Geosci. Remote Sens.* **44**(8), 2207–2218 (2006), <http://dx.doi.org/10.1109/TGRS.2006.872081>.
29. D. P. Roy et al., "Multi-temporal MODIS—Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data," *Rem. Sens. Environ.* **112**(6), 3112–3130 (2008), <http://dx.doi.org/10.1016/j.rse.2008.03.009>.
30. T. Hilker et al., "A new data fusion model for high spatial- and temporal- resolution mapping of forest disturbance based on Landsat and MODIS," *Rem. Sens. Environ.* **113**(8), 1613–1627 (2009), <http://dx.doi.org/10.1016/j.rse.2009.03.007>.
31. H. Liu and Q. Weng, "Enhancing temporal resolution of satellite imagery for public health studies: a case study of West Nile Virus outbreak in Los Angeles in 2007," *Rem. Sens. Environ.* **117**, 57–71(2011), <http://dx.doi.org/10.1016/j.rse.2011.06.023>.
32. J. J. Walker et al., "Evaluation of Landsat and MODIS data fusion products for analysis of dryland forest phenology," *Rem. Sens. Environ.* **117**, 381–393 (2012), <http://dx.doi.org/10.1016/j.rse.2011.10.014>.
33. M. Wu et al., "Simulate TM image from MODIS data, using Linear Mixture Model," in *ICMT2010*, 1064–1067, IEEE Computer Society, Ningbo, China (2010).
34. X. Zhu et al., "An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions," *Rem. Sens. Environ.* **114**(11), 2610–2623 (2010), <http://dx.doi.org/10.1016/j.rse.2010.05.032>.
35. Z. Malenovsky et al., "Scaling dimensions in spectroscopy of soil and vegetation," *Int. J. Appl. Earth Obs. Geoinf.* **9**(2), 137–164 (2007), <http://dx.doi.org/10.1016/j.jag.2006.08.003>.
36. T. H. Painter et al., "Retrieval of subpixel snow-covered area and grain size from imaging spectrometer data," *Rem. Sens. Environ.* **85**(1), 64–77 (2003), [http://dx.doi.org/10.1016/S0034-4257\(02\)00187-6](http://dx.doi.org/10.1016/S0034-4257(02)00187-6).
37. T. W. Ray and B. C. Murray, "Nonlinear spectral mixing in desert vegetation," *Rem. Sens. Environ.* **55**(1), 59–64 (1996), [http://dx.doi.org/10.1016/0034-4257\(95\)00171-9](http://dx.doi.org/10.1016/0034-4257(95)00171-9).
38. S. Cherchali, O. Amram, and G. Flouzat, "Retrieval of temporal profiles of reflectances from simulated and real NOAA-AVHRR data over heterogeneous landscapes," *Int. J. Remote Sens.* **21**(2), 753–775 (2000), <http://dx.doi.org/10.1080/014311600210551>.



Mingquan Wu is a research assistant at The State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, China. His research interests include multisensor data fusion, retrieving vegetation parameters through inversion of remote sensing models, and mapping crops like rice and wheat.

Biographies and photographs of the other authors are not available.