Real-Time Spatiotemporal Spectral Unmixing of MODIS Images

Qunming Wang, Xinyu Ding, Xiaohua Tong, *Senior Member, IEEE*, and Peter M. Atkinson

***Abstract*— Mixed pixels are a ubiquitous problem in remote sensing images. Spectral unmixing has been used widely for mixed pixel analysis. However, up to now, most spectral unmixing methods require endmembers and cannot consider fully intraclass spectral variation. The recently proposed spatiotemporal spectral unmixing (STSU) method copes with the aforementioned problems through exploitation of the available temporal information. However, this method requires coarse-to-fine spatial image pairs both before and after the prediction time and is, thus, not suitable for important real-time applications (i.e., where the fine spatial resolution data after the prediction time are unknown). In this article, we proposed a real-time STSU (RSTSU) method for real-time monitoring. RSTSU requires only a single coarse-to-fine spatial resolution image pair before, and temporally closest to, the prediction time, coupled with the coarse image at the prediction time, to extract samples automatically to train a learning model. By fully incorporating the multiscale spatiotemporal information, the RSTSU method inherits the key advantages of STSU; it does not need endmembers and can account for intraclass spectral variation. More importantly, RSTSU is suitable for real-time analysis and, thus, facilitates the timely monitoring of land cover changes. The effectiveness of the method was validated by experiments on four Moderate Resolution Imaging Spectroradiometer (MODIS) datasets. RSTSU utilizes and enriches the theory underpinning the advanced STSU method and enhances greatly the applicability of spectral unmixing for time-series data.**

***Index Terms*— Machine learning, real time, spatiotemporal spectral unmixing (STSU), spectral unmixing.**

I. Introduction

THE mixed pixel problem has been recognized as a long-standing issue in remotely sensed images [1]. This type of pixel contains multiple land cover classes and its remotely sensed spectrum is a combination of the spectra of the constituent land cover classes. Conventional hard classification [2]–[4] allocates directly each mixed pixel to a single class, and the geometric precision of land cover mapping is, thus, greatly affected. Spectral unmixing has been developed for coping with the mixed pixel problem, which estimates the proportions of all land cover classes in the pixel [5]. Different from traditional hard classification, spectral unmixing can extract subpixel level information in land cover mapping [6]–[9] and provide more detailed land cover information for interpretation. Consequently, spectral unmixing has attracted increasing attention.

Researchers have developed a series of spectral unmixing models to solve the mixed pixel problem, which can be categorized into linear and nonlinear models [10]–[12]. Among the linear models, the linear spectral mixture model (LSMM) has been applied widely due to its clear physical meaning [13], [14]. Nonlinear models can be advantageous, but there is always a risk of collinearity, and the fitting tends to be more complicated. As a result, the prediction error of nonlinear models can be larger than that of linear models in some cases. Therefore, the choice of the mixture model depends mainly on the degree of nonlinearity of the relationship between spectral reflectance and cover proportion for each land cover class in the pixel.

Both the linear and nonlinear approaches require endmembers (represented by the pure spectra of each land cover class) to construct the corresponding mathematical model. Both supervised and unsupervised approaches are available for endmember estimation. The unsupervised methods need to estimate the representative spectra of each land cover, and the commonly used methods include N-FINDR [15], pixel purity index [16], iterative error analysis [17], and vertex component analysis [18]. The supervised methods obtain endmembers through visual interpretation of the images or referring to known spectral libraries.

LSMM methods consider each mixed pixel as a linear combination of the endmembers, and the task is essentially to predict the linear coefficients. Most LSMM-based methods assume that the spectrum of each land cover class can be represented by a single fixed endmember. However, the situation exists widely in remote sensing images whereby the same pure land cover class can produce different spectra [19], that is, intraclass spectral variation is prominent. Thus, it is not sufficient to use a single endmember to represent the spectrum of each land cover class. Indeed, intraclass spectral variation is the main obstacle increasing the accuracy of spectral unmixing. To overcome intraclass spectral variation, researchers have proposed several multiple endmember-based methods, in which multiple endmembers are used to represent the spectra of each land cover class. Multiple endmember-based

Manuscript received May 14, 2021; revised July 17, 2021 and August 7, 2021; accepted August 25, 2021. Date of publication September 2, 2021; date of current version January 26, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 41971297 and Grant 42171345, and in part by Tongji University under Grant 02502350047. *(Corresponding author: Xiaohua Tong.)*

Qunming Wang, Xinyu Ding, and Xiaohua Tong are with the College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China (e-mail: wqm11111@126.com; xhtong@tongji.edu.cn).

Peter M. Atkinson is with the Faculty of Science and Technology, Lancaster University, Lancaster LA1 4YR, U.K., and also with the Geography and Environment, University of Southampton, Southampton SO17 1BJ, U.K.

Digital Object Identifier 10.1109/TGRS.2021.3108540

methods include the multiple endmember spectral mixture analysis method [20], endmember bundles method [21], and Monte Carlo method [22]. They can address the problem of intraclass spectral variation to a certain extent, but often need a large number of endmembers exemplars.

It is always a great challenge to extract a large number of endmembers, especially for regions with great heterogeneity, where a large number of mixed pixels exist. In recent years, with the development of machine learning, support vector machines (SVM) [23], [24], neural networks [25]–[29], and many other learning-based methods have been applied widely to spectral unmixing. These learning-based methods have the advantage of not requiring pure endmembers, rather fitting the relation between the proportions of the mixed pixels and the corresponding spectra by training the learning model using pixels with known proportions (i.e., training samples). For example, Immitzer *et al.* [30] proposed to use the 2-m WorldView-2 data to aid the spectral unmixing of 30-m Landsat data. More specifically, a classified WorldView-2 land cover map was degraded to 30 m to simulate coarse proportions in the corresponding subregion within an entire Landsat region. The 30-m known proportions were integrated with the spectra of the Landsat data to construct training samples for machine learning. Although these methods do not depend on pure endmembers, they do require a large number of pixels with known proportions to construct the training samples, which can be a great challenge in practice.

Most existing spectral unmixing models were developed based on single time data, and there is limited research on spectral unmixing of time-series data. Moreover, the few spectral unmixing methods for handling time-series data cannot monitor dynamic changes in land cover, and the use of temporal information is limited [31]–[34]. Thus, the potentially valuable information in the temporal domain is generally always not fully explored. To cope with the aforementioned problems of intraclass spectral variation and the inability to obtain (a larger number of) endmembers, Wang *et al.* [35] proposed a spatiotemporal spectral unmixing (STSU) model to fully exploit the inherent spatial and temporal information in time-series data. The advanced model extends spectral unmixing from the traditional spatial domain to the spatiotemporal domain and fully utilizes the multiscale spatiotemporal information. Specifically, STSU utilizes temporally neighboring fine spatial resolution images to detect land cover changes and extracts the known coarse class proportions of unchanged pixels to construct training samples for machine learning. Different from the conventional LSMM, the STSU model can extract training samples at the prediction time to construct a learning model, which is free of the need for endmembers and fully considers the intraclass spectral variation. However, this model requires fine spatial resolution images both before and after the predicted data for construction of training samples. Thus, STSU is suitable for spectral unmixing of historical time-series data, rather than near real-time data (i.e., where the fine spatial resolution data after the prediction time are unknown). For clarity, we refer to this type of data as "real time" in this article.

In practical applications, it is necessary to interpret real-time remote sensing images, such as to quickly identify and monitor dynamic changes in land cover. As acknowledged widely, there are rapid changes in the global coverage of vegetation. For example, there exists large-scale human-made logging in the Amazon, Brazil, where the coverage of forest is changing every day. In such rapidly changing areas, monitoring based on real-time data is especially important. Thus, it is of great theoretical significance and application value to develop STSU methods that are suitable for real-time data and can monitor land cover changes in a timely fashion. Actually, the spectral unmixing of real-time data and historical data is quite different in theory, technique, and application.

Based on the requirement of timely monitoring, this article proposes a real-time STSU (RSTSU) method for spectral unmixing of real-time remote sensing images in an operational context. Moderate Resolution Imaging Spectroradiometer (MODIS) data were considered in this article. Due to the coarse spatial resolution of MODIS data, it is usually a great challenge to obtain pure endmembers or information on class proportions for supervised training. We used the temporally closest (before the prediction time) MODIS data (the Landsat data at the same time are also known) for change detection and then extracted the spectral information and proportions of the unchanged MODIS pixels to construct training samples, where the coarse proportions were obtained by upscaling the classified Landsat map. The unchanged pixels represent the type of pixels in which the land cover distributions are the same across time, but the spectra can be different due to vegetation phenology and changes in conditions of data acquisition. Note that the spatial resolution of Landsat data is 30 m, much finer than 500 m of MODIS data. Thus, the former can provide an important reference for proportion at the spatial resolution of 500 m (the 30-m classified land cover map of Landsat data can be upscaled to 500 m to produce the coarse proportion at the corresponding time). However, the temporal resolution of Landsat is 16 days, much coarser than daily of MODIS data. Thus, Landsat data are temporally sparse, and we need to perform spectral unmixing of MODIS data to produce real-time proportion images based on the daily temporal resolution of MODIS images.

For model training, this article exploits an effective machine learning method (i.e., least squares SVM (LSSVM) [36]) to make full use of the limited MODIS training samples to construct an effective learning model and to produce reliable proportion predictions of the MODIS data. The contributions of this article are summarized as follows.

1) RSTSU is proposed for spectral unmixing of real-time MODIS images. Different from STSU, RSTSU is undertaken with only one MODIS-Landsat image pair before the real-time MODIS data, which extends STSU for handling real-time images and enhances its applicability for spectral unmixing. Interpretation of real-time images is crucial for dynamic monitoring of land cover changes (e.g., inundation, fire, and many other emergent incidents). The RSTSU method has great potential for widespread applications that cannot be achieved by STSU.
2) The RSTSU method is undertaken based on change detection and machine learning. It does not require endmember extraction and can take intraclass spectral

variation into account. Meanwhile, the RSTSU method also takes advantage of multiscale spatiotemporal information to increase the accuracy of spectral unmixing. RSTSU is a simple and practical solution that will be understood conveniently by practitioners so that it will be potentially used widely.

3) The RSTSU method was applied to monitor fraction vegetation cover (FVC) [34]–[36], which is an important variable in analysis of global change. RSTSU provides a practical solution to obtain real-time FVC data at large scale, which cannot be realized by STSU.

The remainder of this article is divided into four parts. Section II introduces the principles of the proposed RSTSU method in detail, including the methods of extracting unchanged MODIS pixels (i.e., training samples) and predicting the proportion of changed pixels based on LSSVM. In Section III, experimental results for four datasets are provided to demonstrate the applicability of the new approach. Meanwhile, RSTSU is compared with two benchmark methods. Section IV further discusses the proposed approach, followed by a conclusion in Section V.

II. Methods

A. Overview of the RSTSU Method

Assume that we are predicting the proportion of MODIS data at T_n, and the MODIS and Landsat data at the previous time T_m are known. The whole process of RSTSU is divided into two stages. In the first stage, change detection is performed based on the spectral difference between the MODIS data at T_m and T_n and the unchanged MODIS pixels (the land cover distributions within these pixels are the same across time, but the spectra can be different due to vegetation phenology and changes in sunlight and atmospheric conditions) are extracted, whose proportions can be obtained through degradation of the classified Landsat data at T_m. The proportions coupled with the corresponding observed spectra are used to construct training samples for training the learning model at T_n. The second stage predicts the proportions of all the remaining MODIS pixels (i.e., identified as changed pixels) at T_n based on the trained learning model. The flowchart is shown in Fig. 1.

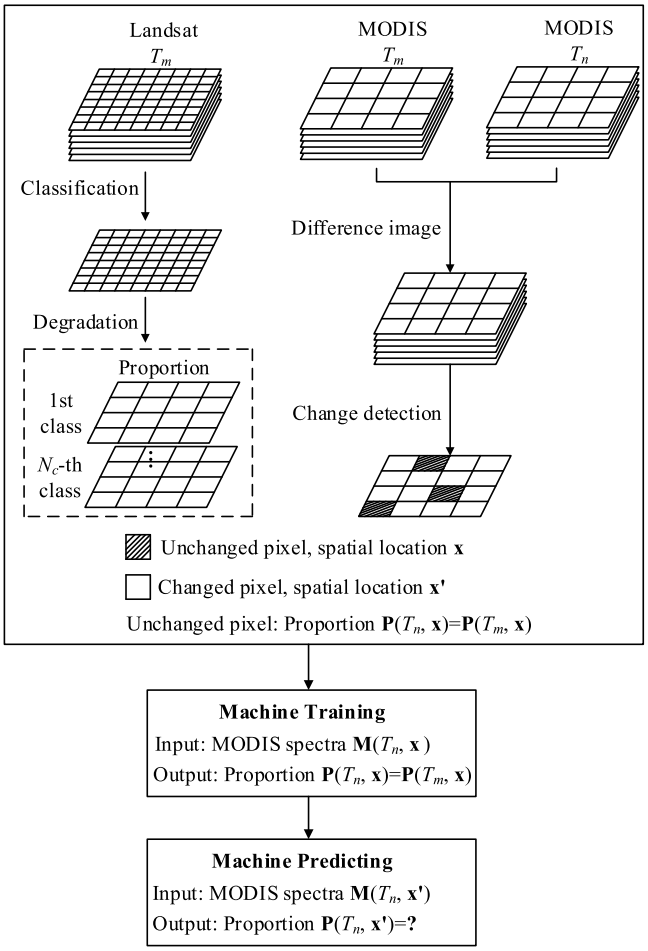


Fig. 1. Methodology of RSTSU.

B. Spectral Differences Between the MODIS Images

In the first stage, we quantified the changes to the MODIS spectra from T_m to T_n, thus providing a basis for extracting the unchanged MODIS pixels subsequently. The spectral change in the MODIS data is calculated as follows:

$$\begin{aligned}\Delta\mathbf{M}(T_{m\to n},\mathbf{y}) &= \mathbf{M}(T_n,\mathbf{y})-\mathbf{M}(T_m,\mathbf{y})\\ &= \begin{bmatrix} M_1(T_n,\mathbf{y})\\ \cdots\\ M_b(T_n,\mathbf{y})\\ \cdots\\ M_B(T_n,\mathbf{y})\end{bmatrix} - \begin{bmatrix} M_1(T_m,\mathbf{y})\\ \cdots\\ M_b(T_m,\mathbf{y})\\ \cdots\\ M_B(T_m,\mathbf{y})\end{bmatrix}\end{aligned} \quad (1)$$

where y is the spatial location of the MODIS pixel to be labeled (i.e., whether changed or unchanged), $\Delta\mathbf{M}(T_{m\to n},\mathbf{y})$ is the spectral difference of the MODIS pixel at y from T_m to T_n, and $\mathbf{M}(T_n,\mathbf{y})$ and $\mathbf{M}(T_m,\mathbf{y})$ are the spectra for the MODIS pixel at y spatially, but at T_n and T_m temporally. $M_b(T_n,\mathbf{y})$ and $M_b(T_m,\mathbf{y})$ are the spectral values at y in the bth band ($b=1,2,\ldots,B$, where B is the number of bands) at T_n and T_m, respectively.

Furthermore, we calculated the modulus ΔM of the spectral change $\Delta\mathbf{M}(T_{m\to n},\mathbf{y})$

$$\Delta M = \|\Delta\mathbf{M}(T_{m\to n},\mathbf{y})\|. \quad (2)$$

The larger the value of ΔM, the larger the likelihood that the land cover has changed. On the contrary, the smaller the value of ΔM, the smaller the likelihood of land cover changes, that is, the pixel is more likely to be an unchanged pixel.

C. Extraction of Unchanged MODIS Pixels (Training Samples)

Given the modulus ΔM of the MODIS spectral change from T_m to T_n, the whole modulus image can be classified into two classes by image segmentation [40]–[42], in which pixels with smaller modulus values are determined as unchanged pixels and the rest as changed pixels. In this article, we used the classic OTSU algorithm for segmentation and to extract unchanged pixels as training samples for machine learning. More precisely, OTSU is a well-established algorithm for

image binarization, which can automatically select a threshold to effectively divide the original image into foreground and background. Moreover, the basis of the threshold determination in the OTSU algorithm is to maximize the variance between the foreground and background parts. For the modulus image, pixels with values smaller than the threshold are determined as unchanged pixels, while the other pixels are determined as changed pixels.

For an extracted unchanged MODIS pixel, the spatial location is recorded as $\mathbf{x}$, and the location of a changed pixel is recorded as $\mathbf{x}'$. For an unchanged pixel, the proportion of each land cover class at T_n is equal to that at T_m, that is,

$$\mathbf{P}(T_n, \mathbf{x}) = \mathbf{P}(T_m, \mathbf{x}) \tag{3}$$

where $\mathbf{P}$ is a vector composed of the proportions of all land cover classes and $\mathbf{P}(T_n, \mathbf{x})$ and $\mathbf{P}(T_m, \mathbf{x})$ are the proportions of the unchanged pixel $\mathbf{x}$ at T_n and T_m, respectively. $\mathbf{P}(T_m, \mathbf{x})$ can be obtained based on the Landsat data at the corresponding time. Specifically, the Landsat data at T_m are classified first, and the classification map of each land cover class is then upscaled to the MODIS spatial resolution to produce the coarse proportion at T_m. The details are shown in Fig. 1. Therefore, according to (3), the proportion of the unchanged pixel at T_n, that is, $\mathbf{P}(T_n, \mathbf{x})$, can be predicted, which can be further combined with the corresponding MODIS spectra $\mathbf{M}(T_n, \mathbf{x})$ to construct training samples at T_n. Based on the training samples, the relationship between the proportion and the MODIS spectra can be fit by a learning machine.

D. Proportion Prediction of Changed MODIS Pixels

SVM is an advantageous learning machine for cases involving a small number of samples and nonlinear and high-dimensional pattern recognition problems [24], [43]–[44]. Among various SVM versions, LSSVM turns the inequality constraints in SVM into equality constraints, which facilitates the calculation [36], [45]. Therefore, LSSVM was selected as the machine learning model in this article. After the training process, the proportions of the changed MODIS pixels at T_n are predicted as

$$\hat{\mathbf{P}}(T_n, \mathbf{x}') = \sum_{i=1}^{n} \alpha_i^* K\left(\mathbf{M}(T_n, \mathbf{x}), \mathbf{M}(T_n, \mathbf{x}')\right) + b^* \tag{4}$$

where α_i^* and b^* are the Lagrange operator and the classification threshold in the LSSVM model, respectively, and both are obtained through the training process. $\hat{\mathbf{P}}(T_n, \mathbf{x}')$ is the proportion prediction of the changed pixel at $\mathbf{x}'$, and $\mathbf{M}(T_n, \mathbf{x})$ and $\mathbf{M}(T_n, \mathbf{x}')$ are the MODIS spectra of the unchanged pixel $\mathbf{x}$ and changed pixel $\mathbf{x}'$ at T_n, respectively. K is a kernel function. Finally, combining with the predictions based on (3), we can obtain the proportion predictions of all MODIS pixels at T_n.

Algorithm RSTSU

Inputs:

1) $\mathbf{M}(T_n, \mathbf{y})$: the spectra of MODIS pixel at $\mathbf{y}$ spatially and T_n temporally.
2) $\mathbf{M}(T_m, \mathbf{y})$: the spectra of MODIS pixel at $\mathbf{y}$ spatially and T_m temporally.
3) $\mathbf{P}(T_m, \mathbf{y})$: the proportions of MODIS pixel at $\mathbf{y}$ spatially and T_m temporally.

Output:

$\hat{\mathbf{P}}(T_n, \mathbf{x}')$: the proportion prediction for the changed pixel at $\mathbf{x}'$ spatially and T_n temporally.

Steps:

1) Calculate the spectral difference between two MODIS pixels $\mathbf{M}(T_n, \mathbf{y})$ and $\mathbf{M}(T_m, \mathbf{y})$ by Eq. (1), and then obtain the modulus ΔM through Eq. (2).
2) The OTSU algorithm is performed on the image composed of ΔM for all pixels to extract unchanged pixels (training samples). The spatial location of unchanged pixel is recorded as $\mathbf{x}$, and the proportion $\mathbf{P}(T_n, \mathbf{x})$ can be acquired by Eq. (3).
3) The LSSVM model is trained to fit the relationship between the spectra of extract unchanged MODIS pixel $\mathbf{M}(T_n, \mathbf{x})$ and the corresponding proportion $\mathbf{P}(T_n, \mathbf{x})$.
4) The trained LSSVM model is used to predict the proportions of changed pixels (i.e., $\hat{\mathbf{P}}(T_n, \mathbf{x}')$) by Eq. (4).

III. Experiments

A. Datasets and Experimental Setup

In the experiments, we used the MODIS and Landsat data of four regions. The first two regions (denoted as Regions 1 and 2) are both located in Daxing, Beijing, China. The used datasets were provided by Li *et al.* [46]. The latter two regions are Zibo, Shandong, China (denoted as Region 3), and Lichuan, Jiangxi, China (denoted as Region 4). For each of the four regions, two 480-m MODIS images at different time points and 30-m Landsat images at the corresponding time points were acquired, as shown in Fig. 2. All the Landsat and MODIS data contain six bands (blue, green, red, near-infrared response (NIR), SWIR1, and SWIR2). Due to the coarse spatial resolution and regular global coverage, MODIS images have been widely used for monitoring of typical land cover classes of interest at the global scale, as can be observed from the popular MODIS products of snow, impervious surface (urban), vegetation, water, and so on. For the four regions, the vegetation (e.g., crops and forest) is the main land cover class. By spectral unmixing of the MODIS images, the FVC [37]–[39] can be predicted, which is a crucial variable in analysis of global change. Therefore, we divided the main land cover classes into vegetation and nonvegetation (i.e., urban) in the four regions.

The same processing was carried out for the Landsat data of the four regions. First, K-means-based unsupervised classification (coupled with visual interpretation and Google map information) was implemented for the two Landsat images, producing two 30-m classification maps covering vegetation and nonvegetation classes, which leads to satisfactory classification accuracy (e.g., with an overall accuracy of above 0.90 for Region 3) and ensure the reliability of the experimental setup. Then, the 30-m classification maps were degraded to 480-m MODIS spatial resolution. Specifically, each 16×16 Landsat pixels were aggregated to a MODIS pixel by averaging the

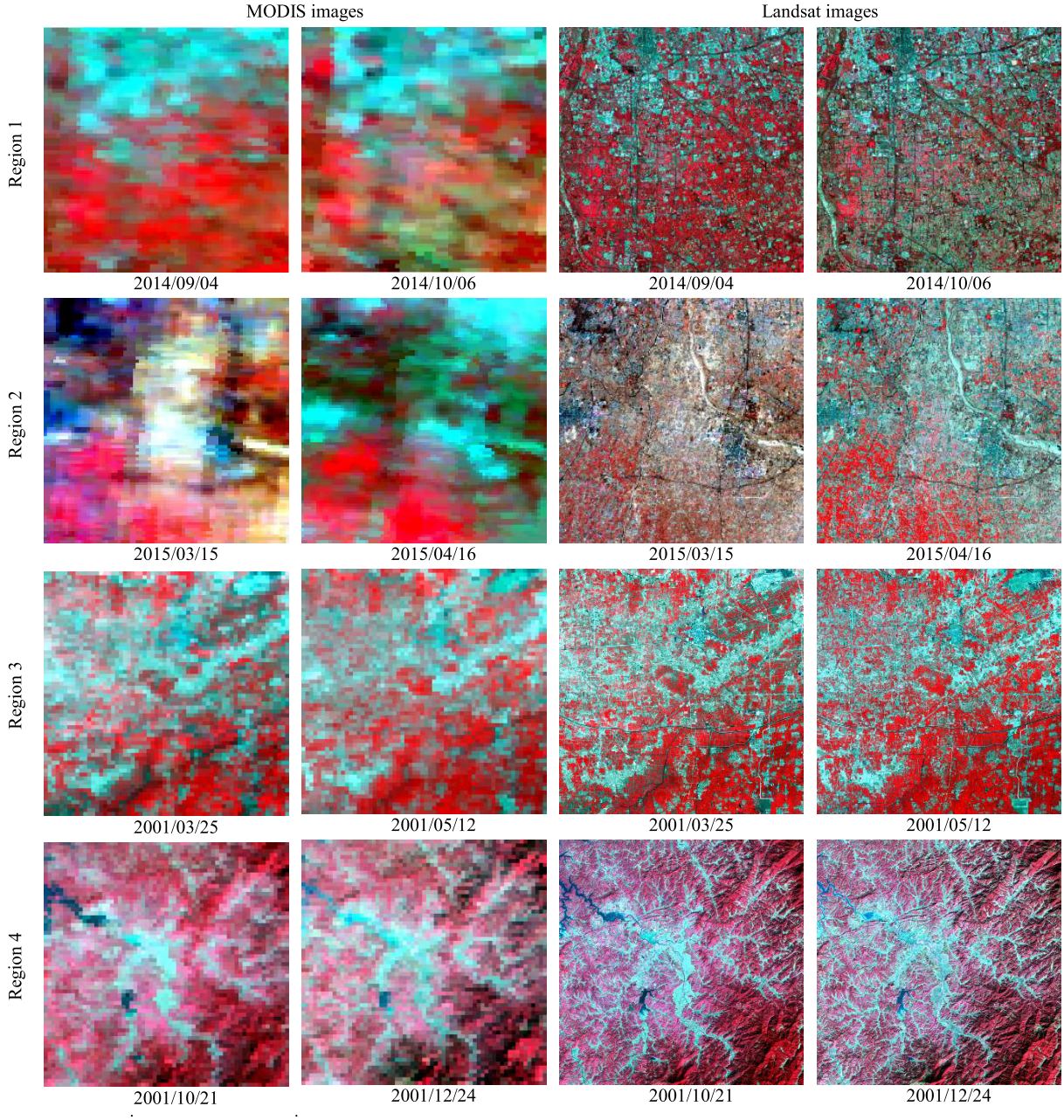


Fig. 2. MODIS and Landsat images of the four regions (NIR–red–green as RGB channels).

256 pixel values. Admittedly, the relation between MODIS and Landsat data cannot be perfectly characterized by the average process. However, uncertainty in other models also exists and the average model has been the most widely used choice up to now. Moreover, for each region and each method, we used the same degradation method for Landsat images at both known and prediction times, that is, for training and predicting samples as well as each spectral unmixing method, the uncertainty caused by this issue is the same, which can be eliminated in the method comparison. Fig. 3 shows the classification images and vegetation proportion images for the four regions.

The data in Region 1 were acquired on September 4 and October 6, 2014. The spatial sizes of the MODIS and Landsat images are 60×60 pixels and 960×960 pixels, respectively. The used MODIS dataset was derived from the MOD02HKM product, which was atmospherically corrected by the Quick Atmospheric Correction algorithm [46]. For Region 2, the data were acquired on March 15 and April 16, 2015. The image sizes of MODIS and Landsat data are 80×80 and 1280×1280 pixels, respectively. The used MODIS product and preprocessing are the same as for Region 1. The data in Region 3 were acquired on March 25 and May 12, 2001. The corresponding MODIS (from the MOD09GA product) and Landsat images have spatial sizes of 62×62 and 992×992 pixels, respectively. The data in Region 4 were acquired on October 21 and December 24, 2001, including two MODIS data with 63×63 pixels (from the MOD09GA product) and Landsat data with 1008×1008 pixels.

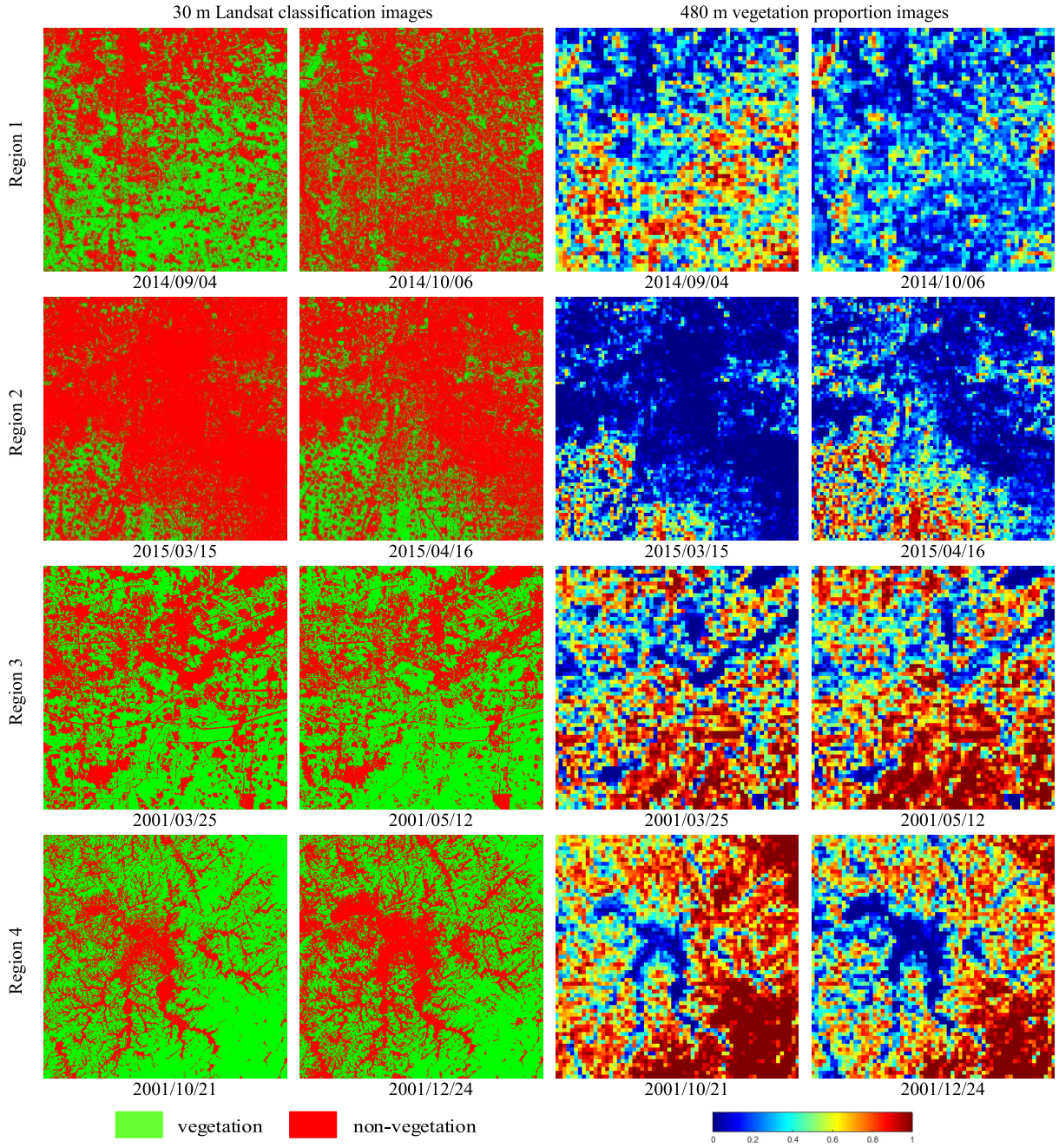


Fig. 3. 30-m Landsat classified images and 480-m vegetation proportion images of the four regions.

For each region, spectral unmixing was performed on the MODIS data at the later time, which is also used the MODIS and Landsat data at the former time. Simultaneously, the 480-m proportion data produced from the Landsat data at the latter time were used as the reference for accuracy evaluation. Note that normally, the MODIS-Landsat image pair temporally closest to the prediction time needs to be used, and the time interval varies for each of the four regions.

B. Evaluation Indices

In this article, we used correlation coefficient (CC), mean absolute error (MAE), and root-mean-square error (RMSE) to evaluate the accuracy of spectral unmixing

$$\mathrm{CC} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \tag{5}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i)^2} \tag{6}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|X_i - Y_i| \tag{7}$$

where X_i is the prediction, Y_i is the reference, $\bar{X}$ and $\bar{Y}$ are the mean values of prediction and reference, respectively, and n is the total number of samples.

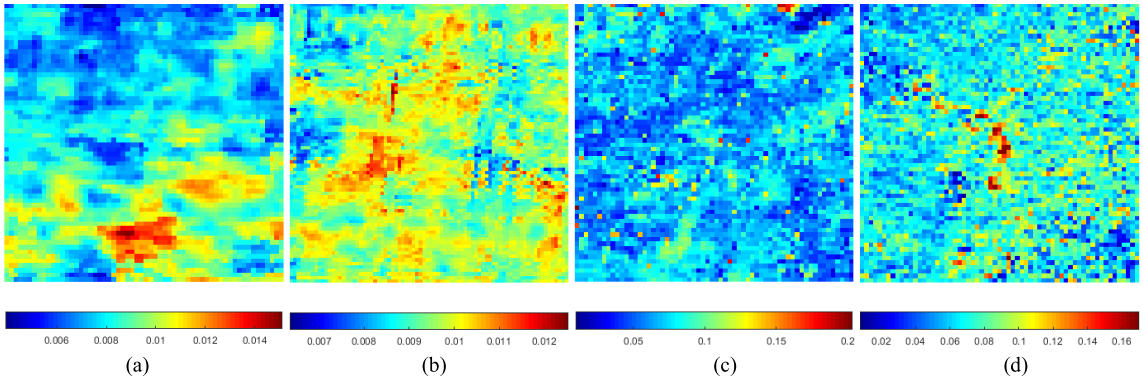


Fig. 4. Modulus images of the MODIS difference data for the four regions. (a) Region 1. (b) Region 2. (c) Region 3. (d) Region 4.

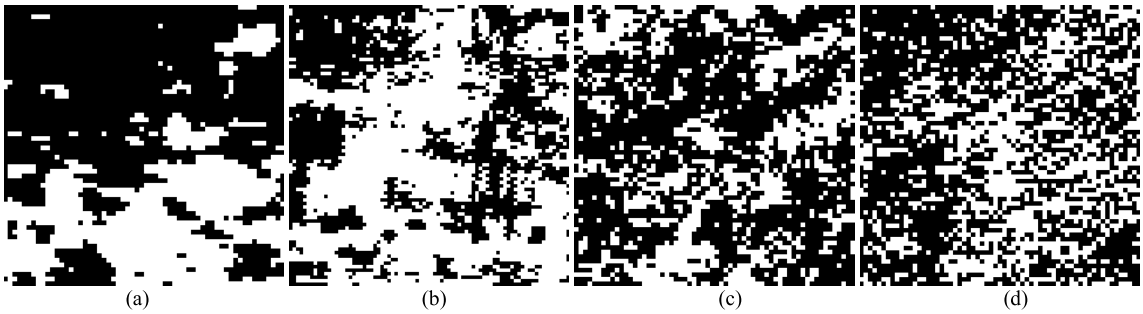


Fig. 5. Segmentation results of OTSU (black and white represent the unchanged and changed pixels, respectively). (a) Region 1. (b) Region 2. (c) Region 3. (d) Region 4.

TABLE I
NUMBER OF DETECTED UNCHANGED MODIS PIXELS IN THE FOUR REGIONS

	Total number of MODIS pixels	Number of unchanged pixel (percentage of the pixels)
Reigon 1	3600	2285 (63.47%)
Region 2	6400	2683 (41.92%)
Region 3	3844	2435 (63.35%)
Region 4	3969	2374 (59.81%)

TABLE II
ACCURACIES OF THE PROPORTIONS FOR THE UNCHANGED MODIS PIXELS IN REGIONS 1–4

	Region 1	Region 2	Region 3	Region 4
CC	0.7357	0.8361	0.9418	0.9230
RMSE	0.1811	0.1642	0.1244	0.1591
MAE	0.1251	0.1128	0.0872	0.1195

C. Results of Training Sample (Unchanged Pixel) Extraction

The modulus images of the MODIS difference data for the four regions are shown in Fig. 4. Based on the modulus results, Fig. 5 shows the unchanged pixels (i.e., training samples) extracted by the OTSU algorithm, where black and white represent the extracted unchanged and changed pixels, respectively. Comparing Figs. 4 and 5, it can be seen that the spatial locations of the extracted unchanged pixels in Fig. 5 correspond mainly to the blue parts in the modulus images in Fig. 4. This conforms to the basic assumption that the smaller the difference, the less likely the land cover changes, indicating the rationality of the OTSU method. Table I lists the number of unchanged pixels extracted by the OTSU algorithm and the percentage relative to all MODIS pixels. Table II shows the accuracies of the proportions for the unchanged MODIS pixels in Regions 1–4. For the four regions, the percentages for the unchanged pixels are 63.47%, 41.92%, 63.35%, and 59.81%, and the corresponding CCs are 0.7357, 0.8361, 0.9418, and 0.9230. The above results show that the unchanged pixels in each region account for the majority of the whole image, suggesting that the proposed scheme can extract sufficient training samples. Meanwhile, as training samples for the subsequent machine learning, the accuracy (in terms of CC) of the proportions for unchanged pixels is mostly above 0.8, revealing the reliability of the training data.

For each of the four regions, we selected a number of unchanged, pure vegetation pixels in the MODIS images at the two times based on the results in Fig. 5 and then averaged the spectra for each time. The spectral profiles of vegetation are shown in Fig. 6. It can be seen that the spectral profiles between the MODIS images at two times present obvious difference, that is, for unchanged pixels, even the land cover distributions (i.e., vegetation proportions) at the two times are the same, the spectra of vegetation change across time, due to vegetation phenology and changes in atmospheric conditions and other factors. Thus, the proposed RSTSU method can account for the spectral variation across time for unchanged pixels.

D. Proportion Predictions

1) Benchmark Methods: In this section, we compare the proposed RSTSU method with two spectral unmixing methods

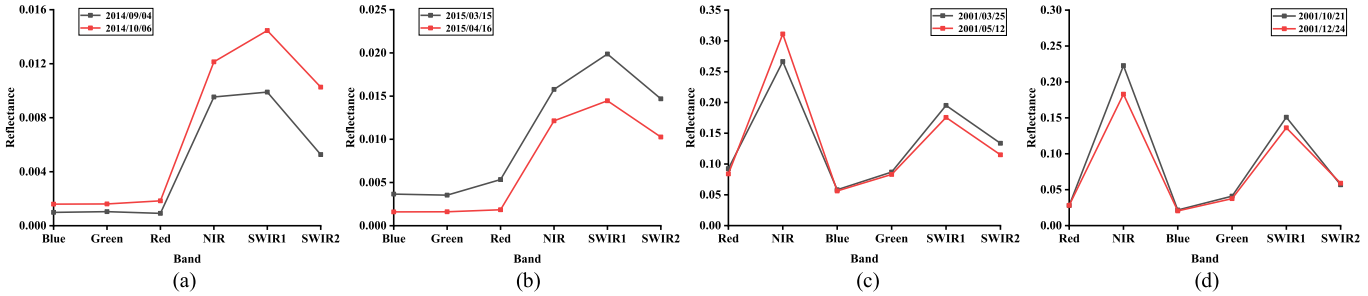


Fig. 6. Spectral profiles of vegetation for Regions 1–4. (a) Region 1. (b) Region 2. (c) Region 3. (d) Region 4.

(denoted briefly as LSMM and LSSVM) to show the advantages of RSTSU. The two methods are briefly described as follows.

LSMM: From T_m to T_n, if an unchanged pixel is a pure pixel at T_m, then the pixel is also recognized as a pure pixel at T_n. After finding all pure pixels at T_n, the spectra of all pure pixels of each class in the MODIS image were averaged and treated as the spectrum of the endmember for LSMM-based spectral unmixing. Due to the great spatial heterogeneity in Regions 1 and 2, pure pixels rarely exist or almost do not exist. Therefore, we assumed that if for one pixel, the proportion of vegetation or nonvegetation is larger than 90%, then it is regarded as sufficiently a pure pixel and the spectra of these pixels were extracted for use in characterizing the endmembers.

LSSVM: The spectra of all MODIS pixels at T_m and the corresponding proportions (obtained from the degradation of the Landsat classification map) were used to construct training samples and to train a learning model. Based on the trained model, we predicted the proportion of all MODIS pixels at T_m.

2) Results: For the proposed RSTSU method, the Gaussian radial basis was used as a kernel function in LSSVM. In Regions 1–4, the penalty factors were 1, 1000, 10, and 1000, respectively, and the kernel parameters were set to one for all regions. Fig. 7 shows the vegetation proportion results of the four regions under the three methods (LSMM, LSSVM, and RSTSU). Comparing them with the corresponding reference proportion, the error maps of the three methods (represented by the absolute value of the error) were produced, as shown in Fig. 8. We can draw the following conclusion based on the analysis in Figs. 7 and 8.

First, the LSMM results are dominated by the clustering of large color blocks. See, for example, the results for Regions 1 and 2. The results suggest that LSMM is inadequate for regions with great spatial heterogeneity, as the results are the most different from the reference. This is because the LSMM method used a fixed endmember to describe each land cover class (i.e., the average spectrum of each land cover class), ignoring intraclass spectra variation.

Second, for the LSSVM method, the difference between the result and the reference is smaller than that for LSMM because LSSVM used multiple samples to describe the intraclass spectral variation, even if they were acquired at a different time. Due to the land cover changes and spectral differences between the two time points, however, the training samples extracted at the previous time, in LSSVM, cannot fully represent the data at the predicted time, that is, the training samples contain great uncertainty and the performance of spectral unmixing is affected greatly. This can be observed clearly from the result for Region 1.

Finally, the difference between the RSTSU results and the reference is the smallest, suggesting that RSTSU is the most accurate method among the three methods. Different from LSSVM, the RSTSU method extracted unchanged pixels based on the MODIS spectra at the prediction time, which are closer to the spectra of the pixels to be unmixed in the same region. Fig. 9 shows the scatterplots between the vegetation proportion predictions and the reference proportions for the three methods and the four regions. It can be seen that the results of LSMM are scattered and sometimes lie on the coordinate axis, and the scatterplots of LSSVM are less dispersed than those of LSMM, but in some cases, the distribution is biased toward a coordinate axis. In particular, in Region 1, many points in the LSMM scatterplot are close to the x-coordinate axis and the distribution of the LSSVM scatterplot is biased toward the x-coordinate axis. In contrast, the result of RSTSU is more concentrated along the $y = x$ line, indicating that the RSTSU result is closer to the reference.

We used three indices to evaluate quantitatively the accuracy of the proportion predictions for the three methods, specifically the CC, RMSE, and MAE. The results are shown in Table III. It can be seen clearly that the CC of the RSTSU method is obviously larger than that of LSMM and LSSVM, and the RMSE and MAE of RSTSU are smaller, indicating that the accuracy of RSTSU is greater than LSSVM and LSSVM. Specifically, for Regions 1–4, the CCs of RSTSU are 0.17, 0.22, 0.11, and 0.10 larger than for LSMM and 0.12, 0.06, 0.10, and 0.06 larger than for LSSVM. The RMSEs of RSTSU are 0.16, 0.02, 0.08, and 0.03 smaller than for LSMM for Regions 1–4. Correspondingly, the MAEs of RSTSU are 0.03, 0.01, 0.06, and 0.006 smaller than for LSSVM.

Based on the results of the proportion maps, error maps, scatterplots maps, and accuracy evaluation, it can be concluded that RSTSU is consistently the most accurate method among the three models.

3) Computational Complexity: The computational times of the three spectral unmixing methods for Regions 1–4 are shown in Table IV. All experiments were undertaken using MATLAB (R2018b) based on a desktop with an Intel Core i9-9900K CPU at 3.60 GHz. It can be seen that the computational time of RSTSU is shorter than that of LSSVM, as the

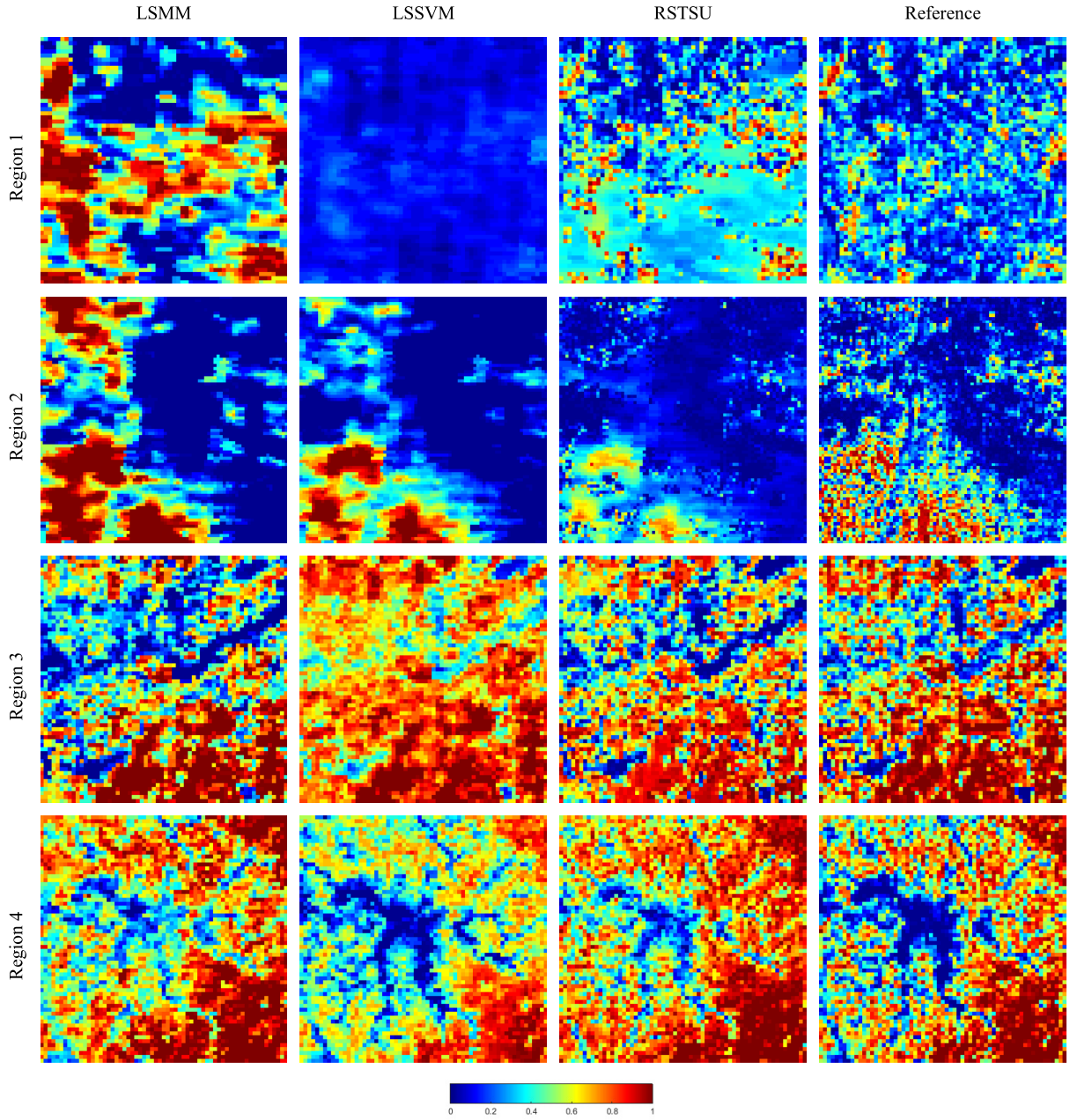


Fig. 7. Vegetation proportion maps of the three spectral unmixing methods for Regions 1–4.

number of training samples in RSTSU is smaller than that of LSSVM. In addition, RSTSU generally requires more time than LSMM.

E. Influence of Training Samples in RSTSU

The reliability of training samples (that is, the detected unchanged pixels) is affected unavoidably by the change detection process. Therefore, to examine the influence of the uncertainty in the training samples, we used the proportion without error (referring to the real proportion at the prediction time) in the training samples directly. Table V shows the accuracies of the RSTSU results with ideal training samples for Regions 1–4. Compared with Table III, it can be seen that there are errors in the process of selecting training samples for RSTSU, and the predictions based on ideal training samples are obviously more accurate than those produced using training samples with uncertainty. More precisely, the decrease in RMSE ranges from 0.04 to 0.10.

F. Comparison Between the Choices of LSSVM and Random Forest (RF) in RSTSU

For the proposed RSTSU method, we selected LSSVM as the machine learning model, which can deal with the problem involving a small number of samples. However, the learning model in RSTSU is not limited to LSSVM, and many different machine learning models can be slotted in. RF is a common machine learning model, which has also been applied widely

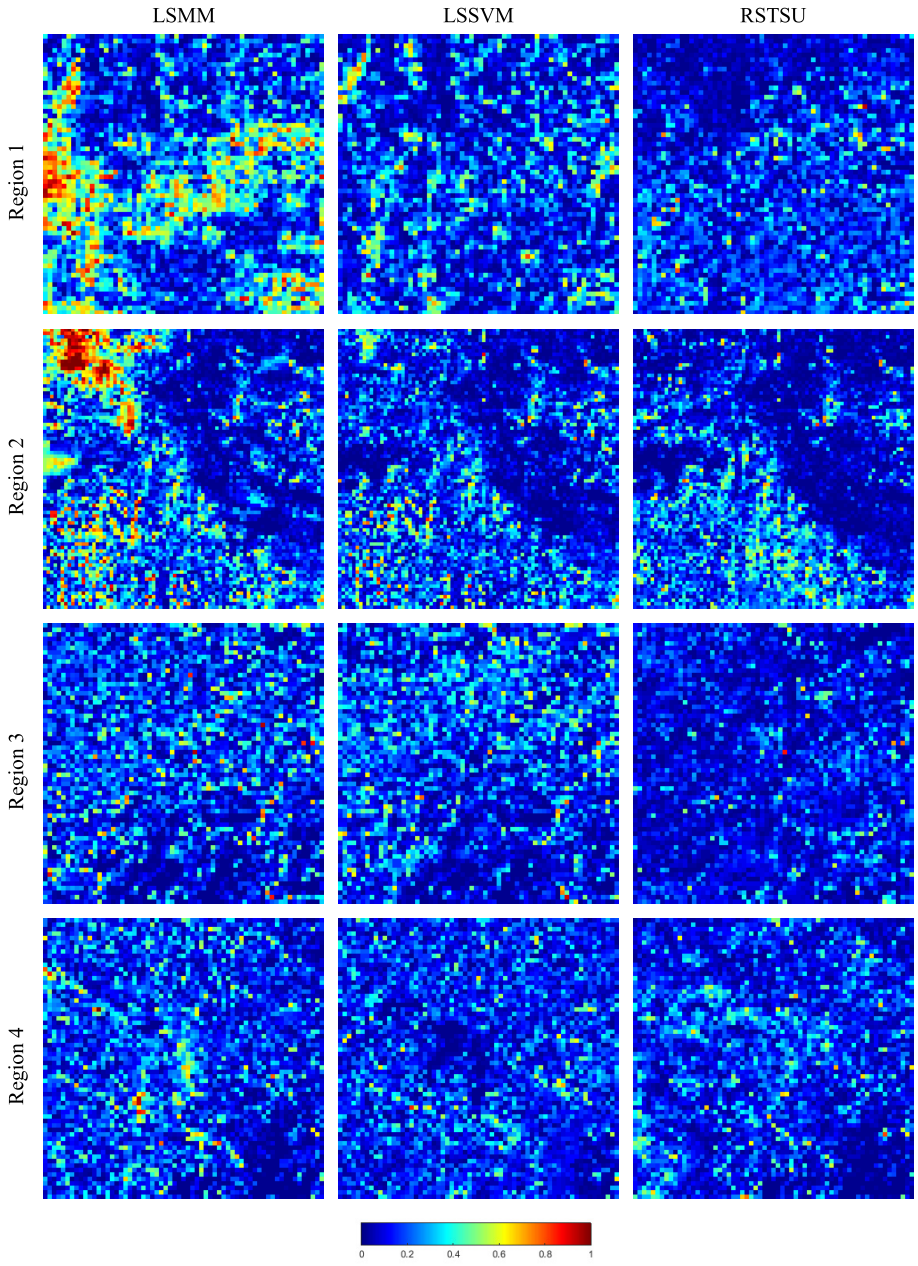


Fig. 8. Vegetation proportion error maps (in absolute value) of the three spectral unmixing methods for Regions 1–4.

in recent years. Thereby, we examined the choice of RF by replacing LSSVM with it in RSTSU.

Table VI shows the accuracies of the spectral unmixing results produced by the RF model for Regions 1–4. The vegetation proportion maps and error maps of the RF-based RSTSU method are shown in Fig. 10. Fig. 11 shows the corresponding scatterplots. Compared with Table III, it can be found that even if the machine learning model is replaced by RF in RSTSU, the spectral unmixing results are still more accurate than the benchmark methods. Meanwhile, it can be seen from Figs. 10 and 11 that the spectral unmixing results of the RF model are also close to those of LSSVM-based RSTSU. Therefore, the implementation and applicability of RSTSU is not limited to a specific learning model.

IV. Discussion

A. Comparison Between RSTSU and STSU

The RSTSU method proposed in this article performs change detection based on MODIS data at the previous time and the prediction time and extracts unchanged MODIS pixels as training samples. Therefore, the RSTSU method can achieve spectral unmixing for data acquired in real time. Different from RSTSU, the STSU method [35] requires images both before and after the prediction time to extract unchanged pixels. Thus, STSU is suitable only for processing historical data. Meanwhile, the STSU method is based on a core assumption: the change of each land cover proportion in the period covering three time points is a monotonous process, that is, the boundary change of the land cover is a process of either

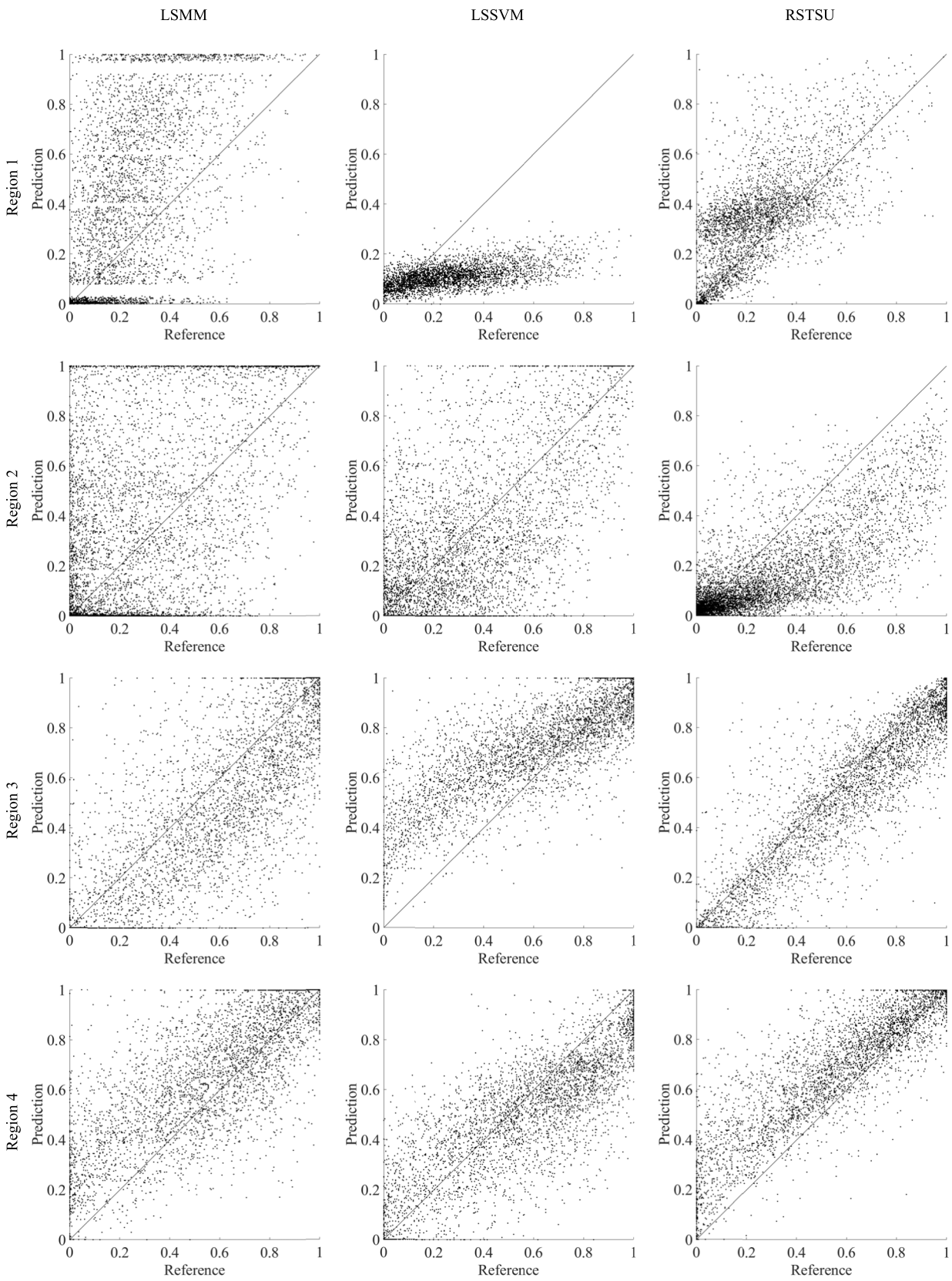


Fig. 9. Scatterplots of vegetation proportion (prediction versus reference) for the three methods in Regions 1–4.

gradual expansion or contraction (e.g., for areas dominated by vegetation cover, the study period cannot include middle of season). On the contrary, the RSTSU method needs only the auxiliary data before the prediction time and is free of the requirement for the above assumptions. Thus, the applicability of RSTSU is more general.

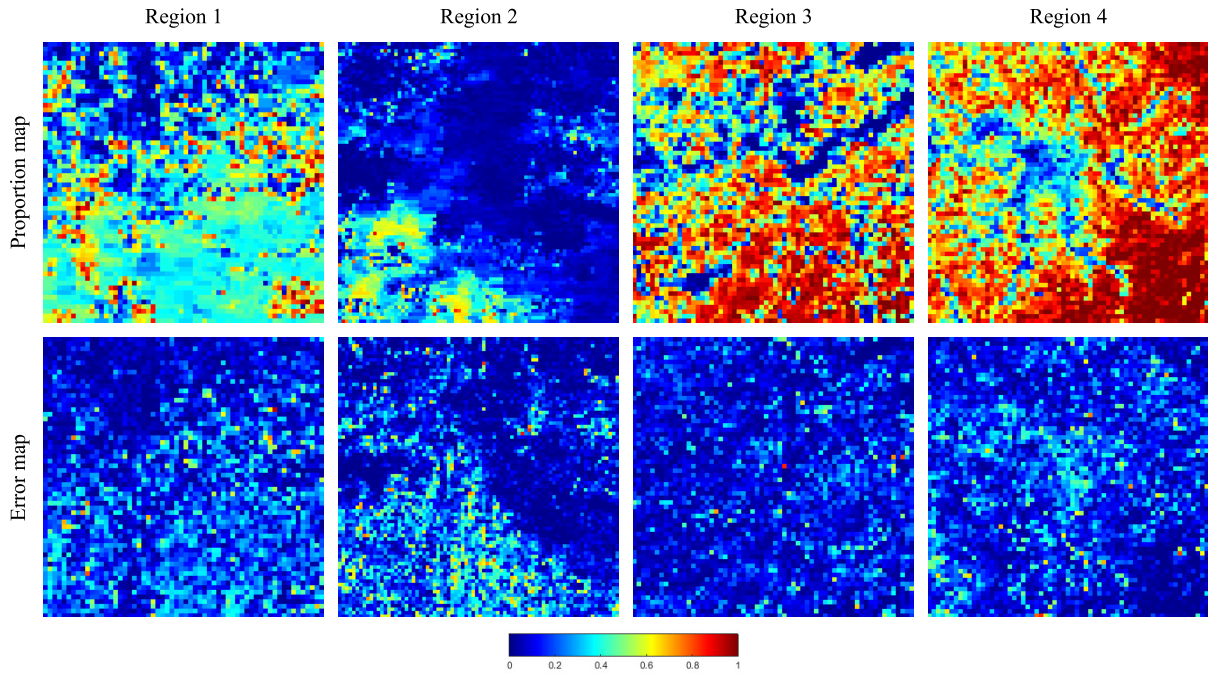


Fig. 10. Vegetation proportion maps and error maps of the RF-based RSTSU model for Regions 1–4.

TABLE III

ACCURACIES OF THE PROPORTION PREDICTIONS FOR REGIONS 1–4 (THE VALUES IN BOLD REPRESENT THE MOST ACCURATE RESULT IN EACH CASE)

		LSMM	LSSVM	RSTSU
Region 1	CC	0.4863	0.5392	**0.6600**
	RMSE	0.3444	0.2264	**0.1821**
	MAE	0.2723	0.1688	**0.1373**
Region 2	CC	0.5498	0.7083	**0.7708**
	RMSE	0.2268	0.2152	**0.2042**
	MAE	0.3152	0.1538	**0.1434**
Region 3	CC	0.7816	0.7971	**0.8993**
	RMSE	0.2228	0.2188	**0.1453**
	MAE	0.1668	0.1662	**0.1056**
Region 4	CC	0.7757	0.8177	**0.8814**
	RMSE	0.2121	0.1844	**0.1831**
	MAE	0.1592	0.1435	**0.1368**

TABLE IV

COMPUTATIONAL TIMES OF DIFFERENT MODELS FOR REGIONS 1–4

	Region 1	Region 2	Region 3	Region 4
LSMM	12.85s	23.83s	13.99s	14.41s
LSSVM	77.03s	399.98s	97.44s	103.23s
RSTSU	15.71s	40.20s	19.05s	19.01s

It needs to be emphasized that although RSTSU is proposed for real-time data, it is also applicable to historical data in theory (i.e., the case for STSU). Therefore, it is necessary to compare the performances of the two methods under the same conditions of processing historical data. The dataset in Region 1 was considered as an example. Specifically, based on the RSTSU method, we used the data on October 6, 2014, to predict the proportion of the MODIS data on September 4, 2014. Meanwhile, for STSU, both the data on October 6 and August 19, 2014, were used for spectral unmixng of the MODIS data on September 4, 2014. The results are shown in Table VII. It can be seen that the CC of the STSU method is 0.07 larger than the RSTSU method. The main reason is that there are obvious changes in vegetation cover between September and October, and the STSU method also uses the data in August to define more reliably the unchanged pixels in the data in September. Thus, the training samples are more reliable, leading to more accurate predictions of proportions. Therefore, STSU is still a preferable choice for spectral unmixing of historical time-series data.

TABLE V

ACCURACIES OF RSTSU RESULTS WITH IDEAL TRAINING SAMPLES FOR REGIONS 1–4

	Region 1	Region 2	Region 3	Region 4
CC	0.8929	0.8461	0.9394	0.9268
RMSE	0.0845	0.1390	0.1032	0.1132
MAE	0.0389	0.0805	0.0453	0.0557

TABLE VI

ACCURACIES OF THE PROPORTION PREDICTIONS OF THE RF-BASED RSTSU MODEL FOR REGIONS 1–4

	Region 1	Region 2	Region 3	Region 4
CC	0.6358	0.7540	0.8950	0.8735
RMSE	0.1982	0.2151	0.1476	0.1889
MAE	0.1511	0.1508	0.1071	0.1416

B. Change Detection Methods

In the first stage of RSTSU, we calculated the spectral differences between the two MODIS images to provide a basis for extracting the unchanged MODIS pixels. The operation

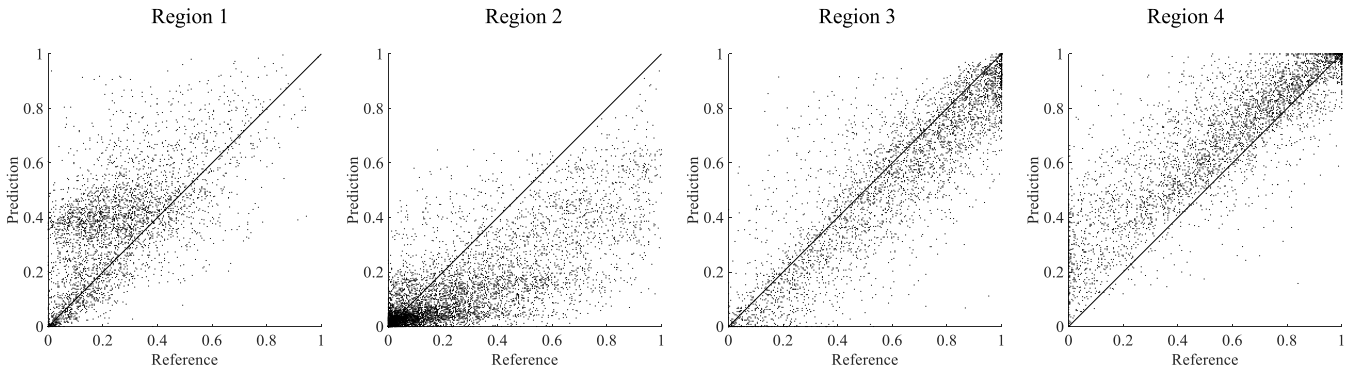


Fig. 11. Scatterplots of vegetation proportion (prediction versus reference) for the RF-based RSTSU model in Regions 1–4.

TABLE VII
ACCURACIES OF RSTSU AND STSU UNDER THE SAME CONDITION (DATA IN REGION 1 AS AN EXAMPLE)

	RSTSU	STSU
CC	0.7018	0.7747
RMSE	0.2142	0.1775
MAE	0.1602	0.1312

of direct subtraction cannot eliminate the noise effect in the original MODIS data, which leads to uncertainty in the change detection process. A common solution to eliminate the noise effect in change detection is to borrow the spatial information in the neighborhood. However, the spatial resolution of MODIS data is very coarse, and the spatial relation between the adjacent pixels is not as obvious as that in finer spatial resolution data such as Landsat or Sentinel-2 images. Thus, spatial context was not considered in the spectral unmixing process in this article. Even though, in Table II, it can be seen that the training samples extracted using only the spectral difference images are reliable, and the resultant RSTSU predictions are more accurate than the benchmark methods.

The change detection module of the RSTSU method in this article uses the OTSU algorithm, which is an image binarization segmentation algorithm based on threshold estimation. Since the performance of the OTSU algorithm is robust to changes in image contrast and brightness, it is a common choice for automatic threshold selection. However, a limitation of this method is that the performance of segmentation is compromised when the sizes of the target (cluster of unchanged pixels as defined in this article) and the background (cluster of changed pixels) in the image are very different or the grayscales of the target and the background have a large overlap. Therefore, there exists inevitable uncertainty in the training samples extracted by the OTSU algorithm. For example, when the land cover change between two time points in the study region is small, the algorithm may extract fewer unchanged pixels, and some unchanged pixels may be incorrectly identified as changed pixels, which would potentially result in fewer training samples for machine learning and larger errors in the spectral unmixing results. Similarly, when the land cover changes greatly, the algorithm may extract more training samples than the actual number of unchanged pixels, which will lead to incorrectly classifying changed pixels as unchanged pixels and affect the final prediction of class proportions. Therefore, to increase the accuracy of change detection, it is necessary to consider and develop more reliable methods for extraction of unchanged pixels, which can consider the spectral changes caused by differences in vegetation phenology, sunlight, and climate. A potential strategy is to seek supervised information, that is, to use the known spectral information of some unchanged or changed pixels to assist in change detection. However, how to effectively collect supervised information from time-series data, especially in a large land cover scenario, will be an important consideration.

C. Uncertainty in Training Samples

The RSTSU method extracts training samples for machine learning and post-spectral unmixing. In this article, we chose the LSSVM model for learning because LSSVM can provide satisfactory performance based on a small number of training samples, that is, given a limited number of known mixed spectra (i.e., proportions of the mixed pixels are known), it can make full use of the information of each sample and account for intraclass spectral variation. However, the training samples extracted by change detection (such as the OTSU algorithm in this article) are not completely reliable, as shown in Table II earlier. Thus, the uncertainty of the training sample will also be propagated to the machine training process, affecting the final predictions, as shown in Table V. For the LSSVM model, we exploited all the unchanged pixels extracted from the change detection step as training samples for model training. During the learning process, we did not consider the impact of the uncertainty in the training samples. Specifically, we assigned the same penalty factor to each training sample. In theory, however, the reliability of each sample should not be the same. For example, the smaller ΔM for an unchanged pixel, the smaller the possibility of the corresponding land cover change (i.e., the larger the importance of the pixel as a training sample), and vice versa. In this regard, we can consider weighting different training samples in the learning model (e.g., different weights to the penalty factor in LSSVM). The weights can be determined according to the importance of each sample. On the other hand, not all training samples are effective for learning, that is, some samples may affect the training model negatively, especially when a number of outliers exist. Thus, we should select more reliable samples for

training. The process of selecting and weighting is expected to reduce the influence of the uncertainty in the training samples and further enhance the reliability of the model. The question of how to filter valid samples and develop an effective criterion to quantify the contribution of each sample in the learning model requires in-depth research in future.

D. Applicability of RSTSU

Implementation of the RSTSU method is simple, mainly including MODIS image change detection, machine learning, and proportion prediction. Moreover, each part of RSTSU is easy to implement, as presented in Section II. Thus, the method can be readily coded. Meanwhile, due to the simplicity of RSTSU, practitioners can understand RSTSU more conveniently so that RSTSU will be potentially applied widely. It is also general and can be applied to a large number of situations with different numbers of data and different scenarios. In addition, there has been very little research on spectral unmixing of real-time data up to now. RSTSU can handle real-time images and enhance the applicability of spectral unmixing. Experimental results validate consistently that the proposed RSTSU method can make full use of temporal information to obtain more accurate prediction of proportions for real-time data. Thus, the RSTSU method is a simple and practical solution to spectral unmixing of real-time data.

E. Image Pair in RSTSU

For the proposed RSTSU method, it is important to make full use of the temporal information for spectral unmixing. Therefore, the MODIS-Landsat image pair was exploited as auxiliary information to extract training samples for learning. With respect to the number of image pairs, we used one MODIS-Landsat image pair temporally closest to the prediction time. It should be stressed that in the image pair, the MODIS and Landsat data are acquired at the same time (i.e., on the same day). Because of the different revisit intervals of the two types of remote sensing images, as well as the influence of cloud contamination, it is normally difficult to acquire an effective image pair. Thus, we used only one MODIS-Landsat image pair for spectral unmixing of real-time data. Although there are multiple historical image pairs among the long time series, the large time interval between these image pairs and the predicted images could lead to great land cover changes, reducing the value of the image pair. However, if the land cover in the study region presents obvious periodical changes in the long time series, we may consider to exploit the use of multiple pairs, which may aid the spectral unmixing to a certain extent.

F. Multisource Auxiliary Data

In this article, for spectral unmixing of 480-m MODIS real-time data, the fine spatial resolution (i.e., 30 m) Landsat time-series data were used to extract supervised information on class proportions. In practice, however, the time interval between effective Landsat images can be relatively long (normally much longer than the temporal revisit frequency of 16 days) due to cloud contamination [47]–[52]. It should be emphasized that we used the 30-m Landsat time-series data to produce the land cover map before the prediction time and then obtained a 480-m proportion at the corresponding time. Thereby, this means that our method does not have specific requirements on the consistency of wavelengths or spectral resolutions for the data at different spatial scales. Thus, we can also consider the use of other effective fine spatial resolution data. For example, Sentinel-2 is a program operated with twin satellites, which can acquire fine spatial resolution (i.e., 10 m) images with high quality and with a potential revisit interval as frequent as every to five days [48]. We can analyze these fine spatial resolution data to obtain the required land cover maps and further auxiliary proportion information, which can be used for STSU of real-time MODIS data. Furthermore, RSTSU is also applicable for spectral unmixing of other data, such as Sentinel 3 images. Therefore, the RSTSU model can be promoted in more application scenarios and applied to spectral unmixing of more types of multiscale remote sensing time-series data, with the aid of various fine spatial resolution neighbors.

V. Conclusion

Existing spectral unmixing methods usually assume the existence of pure endmembers or fail to account fully for intraclass spectral variation. The very few available spectral unmixing methods incorporating temporal information were not designed for dynamic monitoring of land cover changes. The recently proposed STSU method addressed the above problems effectively, but it is suitable for spectral unmixing of historical time-series data only. Thus, this article proposed the RSTSU method for spectral unmixing of real-time data, which not only extends spectral unmixing from the traditional spatial domain to the spatiotemporal domain but also from the processing of historical data to real-time data. The RSTSU method uses a MODIS-Landsat pair temporally closest to the prediction time and performs change detection based on the MODIS data at two times to extract unchanged MODIS pixels (i.e., training samples) for machine learning. The trained model is applied to predict the proportion of the remaining changed MODIS pixels. The RSTSU method does not require endmember extraction and also takes the intraclass spectral variation into consideration. RSTSU is a suitable method for real-time monitoring of land cover changes. We selected MODIS data covering four different regions for experimental validation. The results show that the RSTSU method can produce greater accuracy than the commonly used LSMM method and the LSSVM method using training samples extracted at the time prior to the prediction time.

References

[1] R. A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*. San Diego, CA, USA: Academic, 1997.

[2] H. Wu, A. Lin, K. C. Clarke, W. Shi, A. Cardenas-Tristan, and Z. Tu, "A comprehensive quality assessment framework for linear features from volunteered geographic information," *Int. J. Geograph. Inf. Sci.*, vol. 2020, pp. 1–22, Oct. 2020.

[3] H. Wu *et al.*, "Examining the sensitivity of spatial scale in cellular automata Markov chain simulation of land use change," *Int. J. Geograph. Inf. Sci.*, vol. 33, no. 5, pp. 1040–1061, May 2019.

[4] A. Lin *et al.*, "A big data-driven dynamic estimation model of relief supplies demand in urban flood disaster," *Int. J. Disaster Risk Reduction*, vol. 49, Oct. 2020, Art. no. 101682.

[5] J. J. Settle and N. A. Drake, "Linear mixing and the estimation of ground cover proportions," *Int. J. Remote Sens.*, vol. 14, no. 6, pp. 1159–1177, 1993.

[6] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.

[7] Q. Wei, M. Chen, J.-Y. Tourneret, and S. Godsill, "Unsupervised nonlinear spectral unmixing based on a multilinear mixing model," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4534–4544, Aug. 2017.

[8] J. Liu and J. Zhang, "Spectral unmixing via compressive sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7099–7110, Nov. 2014.

[9] Q. Wang, C. Zhang, X. Tong, and P. M. Atkinson, "General solution to reduce the point spread function effect in subpixel mapping," *Remote Sens. Environ.*, vol. 251, Dec. 2020, Art. no. 112054.

[10] D. C. Heinz and C.-I. Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 529–545, Mar. 2001.

[11] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.

[12] R. Heylen, M. Parente, and P. Gader, "A review of nonlinear hyperspectral unmixing methods," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 1844–1868, Jun. 2014.

[13] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[14] C. Quintano, A. Fernández-Manso, Y. E. Shimabukuro, and G. Pereira, "Spectral unmixing," *Int. J. Remote Sens.*, vol. 33, no. 17, pp. 5307–5340, Sep. 2012.

[15] S. Dowl and M. Andrews, "On the convergence of N-FINDR and related algorithms: To iterate or not to iterate?" *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 1, pp. 4–8, Jan. 2011.

[16] X. Wu, B. Huang, A. Plaza, Y. Li, and C. Wu, "Real-time implementation of the pixel purity index algorithm for endmember identification on GPUs," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 5, pp. 955–959, May 2014.

[17] L. Wang, F. Wei, D. Liu, Y. Wang, and Q. Wang, "Endmember extraction based on modified iterative error analysis," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2013, pp. 1071–1074.

[18] J. M. P. Nascimento and J. M. B. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

[19] C. Zhang, L. Ma, J. Chen, Y. Rao, Y. Zhou, and X. Chen, "Assessing the impact of endmember variability on linear spectral mixture analysis (LSMA): A theoretical and simulation analysis," *Remote Sens. Environ.*, vol. 235, Dec. 2019, Art. no. 111471.

[20] 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.

[21] C. A. Bateson, G. P. Asner, and C. A. Wessman, "Endmember bundles: A new approach to incorporating endmember variability into spectral mixture analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 2, pp. 1083–1094, Mar. 2000.

[22] G. P. Asner and D. B. Lobell, "A biogeophysical approach for automated SWIR unmixing of soils and vegetation," *Remote Sens. Environ.*, vol. 74, no. 1, pp. 99–112, Oct. 2000.

[23] L. Wang and X. Jia, "Integration of soft and hard classifications using extended support vector machines," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 3, pp. 543–547, Jul. 2009.

[24] H. Wang, J. Li, and F. Yang, "Overview of support vector machine analysis and algorithm," *Appl. Res. Comput.*, vol. 31, no. 5, pp. 1281–1286, 2014.

[25] X. Zhang, Y. Sun, J. Zhang, P. Wu, and L. Jiao, "Hyperspectral unmixing via deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 11, pp. 1755–1759, Nov. 2018.

[26] Y. Su, J. Li, A. Plaza, A. Marinoni, P. Gamba, and S. Chakravortty, "DAEN: Deep autoencoder networks for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4309–4321, Jul. 2019.

[27] M. Wang, M. Zhao, J. Chen, and S. Rahardja, "Nonlinear unmixing of hyperspectral data via deep autoencoder networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 9, pp. 1467–1471, Sep. 2019.

[28] Y. Qu and H. Qi, "uDAS: An untied denoising autoencoder with sparsity for spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1698–1712, Mar. 2019.

[29] S. Ozkan, B. Kaya, and G. B. Akar, "EndNet: Sparse autoencoder network for endmember extraction and hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 482–496, Jan. 2019.

[30] M. Immitzer *et al.*, "Fractional cover mapping of spruce and pine at 1 ha resolution combining very high and medium spatial resolution satellite imagery," *Remote Sens. Environ.*, vol. 204, pp. 690–703, Jan. 2018.

[31] R. Zurita-Milla, L. Gomez-Chova, L. Guanter, J. G. P. W. Clevers, and G. Camps-Valls, "Multitemporal unmixing of medium-spatial-resolution satellite images: A case study using MERIS images for land-cover mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4308–4317, Nov. 2011.

[32] F. Yang, B. Matsushita, T. Fukushima, and W. Yang, "Temporal mixture analysis for estimating impervious surface area from multi-temporal MODIS NDVI data in Japan," *ISPRS J. Photogramm. Remote Sens.*, vol. 72, pp. 90–98, Aug. 2012.

[33] C. Deng and Z. Zhu, "Continuous subpixel monitoring of urban impervious surface using Landsat time series," *Remote Sens. Environ.*, vol. 238, Mar. 2020, Art. no. 110929.

[34] E. L. Bullock, C. E. Woodcock, and P. Olofsson, "Monitoring tropical forest degradation using spectral unmixing and Landsat time series analysis," *Remote Sens. Environ.*, vol. 238, Mar. 2020, Art. no. 110968.

[35] Q. Wang, X. Ding, X. Tong, and P. M. Atkinson, "Spatio-temporal spectral unmixing of time-series images," *Remote Sens. Environ.*, vol. 259, Jun. 2021, Art. no. 112407.

[36] J. A. K. Suykens, J. De Brabanter, L. Lukas, and J. Vandewalle, "Weighted least squares support vector machines: Robustness and sparse approximation," *Neurocomputing*, vol. 48, nos. 1–4, pp. 85–105, 2002.

[37] T. N. Carlson and D. A. Ripley, "On the relation between NDVI, fractional vegetation cover, and leaf area index," *Remote Sens. Environ.*, vol. 62, no. 3, pp. 241–252, 1997.

[38] Y. Liu, X. Mu, H. Wang, and G. Yan, "A novel method for extracting green fractional vegetation cover from digital images," *J. Vegetation Sci.*, vol. 23, no. 3, pp. 406–418, Jun. 2012.

[39] G. Yan *et al.*, "Improving the estimation of fractional vegetation cover from UAV RGB imagery by colour unmixing," *ISPRS J. Photogramm. Remote Sens.*, vol. 158, pp. 23–34, Dec. 2019.

[40] T. W. Ridler and S. Calvard, "Picture thresholding using an iterative selection method," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-8, no. 8, pp. 630–632, Aug. 1978.

[41] S. S. Reddi, S. F. Rudin, and H. R. Keshavan, "An optimal multiple threshold scheme for image segmentation," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-14, no. 4, pp. 661–665, Jul. 1984.

[42] S. Arora, J. Acharya, A. Verma, and P. K. Panigrahi, "Multilevel thresholding for image segmentation through a fast statistical recursive algorithm," *Pattern Recognit. Lett.*, vol. 29, no. 2, pp. 119–125, Jan. 2008.

[43] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag, 1995.

[44] M. Pal and G. M. Foody, "Feature selection for classification of hyperspectral data by SVM," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2297–2307, May 2010.

[45] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.*, vol. 9, no. 3, pp. 293–300, Jun. 1999.

[46] J. Li, Y. Li, L. He, J. Chen, and A. Plaza, "Spatio-temporal fusion for remote sensing data: An overview and new benchmark," *Sci. China Inf. Sci.*, vol. 63, no. 4, Apr. 2020, Art. no. 140301.

[47] Y. Tang, Q. Wang, K. Zhang, and P. M. Atkinson, "Quantifying the effect of registration error on spatio-temporal fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 487–503, 2020.

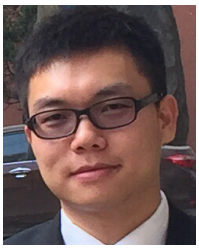
[48] Q. Wang, Y. Tang, X. Tong, and P. M. Atkinson, "Virtual image pair-based spatio-temporal fusion," *Remote Sens. Environ.*, vol. 249, Nov. 2020, Art. no. 112009.

[49] B. Chen, B. Huang, L. Chen, and B. Xu, "Spatially and temporally weighted regression: A novel method to produce continuous cloud-free Landsat imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 27–37, Jan. 2017.

[50] G. Gao and Y. Gu, "Multitemporal Landsat missing data recovery based on tempo-spectral angle model," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3656–3668, Jul. 2017.

[51] Q. Wang *et al.*, "Filling gaps in Landsat ETM+ SLC-off images with Sentinel-2 images," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 101, Sep. 2021, Art. no. 102365.

[52] Q. Wang, K. Peng, Y. Tang, X. Tong, and P. M. Atkinson, "Blocks-removed spatial unmixing for downscaling MODIS images," *Remote Sens. Environ.*, vol. 256, Apr. 2021, Art. no. 112325.



Qunming Wang received the Ph.D. degree from Hong Kong Polytechnic University, Hong Kong, in 2015.

He was a Lecturer and an Assistant Professor with Lancaster Environment Centre, Lancaster University, Lancaster, U.K., from 2017 to 2018. He is currently a Professor with the College of Surveying and Geo-Informatics, Tongji University, Shanghai, China. His three-year Ph.D. study was supported by the hyper-competitive Hong Kong Ph.D. Fellowship and his Ph.D. thesis was awarded as the Outstanding Thesis in the Faculty. He has authored or coauthored 60 peer-reviewed articles in international journals, such as *Remote Sensing of Environment*, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, and *ISPRS Journal of Photogrammetry and Remote Sensing*. His research interests include remote sensing, image processing, and geostatistics.

Dr. Wang was an Associate Editor of *Computers and Geosciences* from 2017 to 2020. He serves as an Associate Editor for *Science of Remote Sensing* (sister journal of *Remote Sensing of Environment*) and *Photogrammetric Engineering and Remote Sensing.*

Xinyu Ding received the B.S. degree from Jiangxi University of Science and Technology, Ganzhou, China, in 2020. She is currently pursuing the Ph.D. degree with Tongji University, Shanghai, China.

Her research interest includes remote sensing image processing, spectral unmixing, and time-series analysis.

Xiaohua Tong (Senior Member, IEEE) received the Ph.D. degree in traffic engineering from Tongji University, Shanghai, China, in 1999.

He is currently a Professor with the College of Surveying and Geoinformatics, Tongji University. He was a Research Fellow with Hong Kong Polytechnic University, Hong Kong, in 2006, and a Visiting Scholar with the University of California at Santa Barbara, Santa Barbara, CA, USA, from 2008 to 2009. His research interests include remote sensing, geographic information systems, uncertainty and spatial data quality, image processing for high-resolution, and hyperspectral images.

Peter M. Atkinson received the Ph.D. degree from the University of Sheffield, Sheffield, U.K., in 1990, and the M.B.A. degree from the University of Southampton, Southampton, U.K., in 2012.

He was a Professor of Geography with the University of Southampton, where he is currently a Visiting Professor. He is also Visiting Professor with the Chinese Academy of Sciences, Beijing, China. He is currently a Distinguished Professor of Spatial Data Science and the Dean of the Faculty of Science and Technology with Lancaster University, Lancaster, U.K. He has published over 300 peer-reviewed articles in international scientific journals and around 50 referred book chapters. The main focus of his research is in remote sensing, geographical information science and spatial (and space–time) statistics applied to a range of environmental science, and socioeconomic problems.

Prof. Atkinson is a fellow of the Learned Society of Wales. He previously held the Belle van Zuylen Chair of Utrecht University, Utrecht, The Netherlands. He was a recipient of the Peter Burrough Award from the International Spatial Accuracy Research Association and the NERC CASE Award from the Rothamsted Experimental Station. He has also edited nine journal special issues and eight books. He is the Editor-in-Chief of *Science of Remote Sensing*, a sister journal of *Remote Sensing of Environment*. He also sits on the Editorial Board of several further journals, including *Geographical Analysis*, *Spatial Statistics*, *International Journal of Applied Earth Observation and Geoinformation*, and *Environmental Informatics*.