

A Spatiotemporal Fusion Based Cloud Removal Method for Remote Sensing Images With Land Cover Changes

Huanfeng Shen¹, Senior Member, IEEE, Jingan Wu, Qing Cheng¹, Mahemujiang Aihemaiti, Chengyue Zhang, and Zhiwei Li¹

Abstract—Cloud contamination greatly limits the potential utilization of optical remote sensing images for geoscience applications. Many solutions have been developed to remove the clouds from multispectral images. Among these approaches, the temporal-based methods which borrow complementary information from multitemporal images outperform the other methods. However, the common fundamental supposition of the temporal-based methods decides that they are only suitable for scenes with phenological changes, while they perform poorly in cases with significant land cover changes. In this paper, a cloud removal procedure based on multisource data fusion is developed to overcome this limitation. On the basis of the temporal-based approaches, which employ a cloud-free image as reference, this method further introduces two auxiliary images with similar wavelengths and close acquisition dates to the reference and target (contaminated) images into the reconstruction process. The temporal variability of the land cover is captured from the two auxiliary images through a modified spatiotemporal data fusion model, and thus, the serious errors produced by the temporal-based methods can be avoided. Moreover, a residual correction strategy based on the Poisson equation is used to enhance the spectral coherence between the recovered and remaining regions. The experiments confirmed that the proposed method can perform very well for cases with significant land cover changes. Compared with some state-of-the-art approaches, it produces lower bias and more robust efficacy. In conclusion, our method will act as an important technical supplement to the current cloud removal framework, and it provides the possibility to handle scenes with significant land cover changes.

Index Terms—Cloud removal, land cover changes, optical remote sensing images, residual correction, spatiotemporal data fusion.

Manuscript received July 23, 2018; revised November 30, 2018 and January 13, 2019; accepted January 28, 2019. Date of publication March 5, 2019; date of current version March 25, 2019. This work was supported by the National Natural Science Foundation of China under Grants 41601357 and 61671334. (*Corresponding author: Qing Cheng*)

H. Shen is with the School of Resource and Environmental Sciences and the Collaborative Innovation Center for Geospatial Technology, Wuhan University, Wuhan 430072, China (e-mail: shenhf@whu.edu.cn).

J. Wu, C. Zhang, and Z. Li are with the School of Resource and Environmental Sciences, Wuhan University, Wuhan 430072, China (e-mail: wujg1992@sina.com; cowherbseed@gmail.com; lizw@whu.edu.cn).

Q. Cheng is with the School of Urban Design, Wuhan University, Wuhan 430072, China (e-mail: qingcheng@whu.edu.cn).

M. Aihemaiti is with the College of Water Conservancy and Civil Engineering, Xinjiang Agricultural University, Urumqi 830052, China (e-mail: amraom@126.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSTARS.2019.2898348

I. INTRODUCTION

A LARGE volume of remotely sensed observations is now acquired every year, and geoscience applications based on these data facilitate the monitoring of the land surface environment and the understanding of Earth dynamics [1]–[2]. However, a huge gap still exists between the data we acquire and the data we require [3]. A very serious obstacle to the further utilization of optical remote sensing images is the cloud and cloud shadow contamination issue (referred to as cloud contamination, hereafter). The passive radiant energy of cloud-covered ground features cannot pass through clouds and be captured by satellite-borne sensors during the acquisition process, resulting in missing information in optical images. It has been reported that Landsat ETM+ images are 35% cloud-contaminated globally [4], and the problem is even worse in humid tropical areas. Therefore, cloud removal is a vital topic for the subsequent applications.

Much effort has been devoted to addressing this issue. Essentially speaking, cloud removal is a process of missing information reconstruction. According to the complementary information source, the existing reconstruction methods can be classified into four categories: spatial-based methods, spectral-based methods, temporal-based methods, and multisource-based methods [5]. A brief review of cloud removal is presented below. It is noteworthy that gap filling is identical to cloud removal, in essence, and thus, some approaches for gap filling are also included in the following discussion.

Spatial-based methods exploit the cloud-free regions of a target image to provide complementary information. Spatial interpolation approaches were one of the first type of methods to be applied to estimate missing pixels [6]–[7]; however, they only work for small gaps. Subsequently, some advanced mathematical techniques have been employed to better solve the issue [8]–[11]. Technically, these methods propagate the geometric structures of ground features from cloud-free regions to cloud-covered regions; thus, realizing the goal of reconstruction. Generally speaking, the spatial-based methods are capable of yielding plausible cloud-free visualization for small gaps [12]–[13]; however, they perform poorly for large-area contamination, and thus, they are unsuitable for quantitative analysis and further applications. The spectral-based methods make full use of the correlation between the contaminated bands and

auxiliary clear bands for the reconstruction. For example, a haze-optimized transformation method was proposed by Zhang *et al.* [14] to radiometrically correct the visible bands of Landsat images contaminated by thin cloud and haze. A similar strategy can be found in some of the research into the recovery of missing data of Aqua MODIS band 6 [15]–[17]. However, thick clouds usually have serious contamination in all bands, and no complete clear bands can be employed as complementary information. As a result, the spectral-based methods only work for thin clouds, and they cannot deal with thick clouds.

Remote sensing systems with a regular revisit cycle can provide us with multitemporal images for a given location. The temporal correlation between multitemporal images is fully utilized for reconstruction by the temporal-based methods. For time-series data with a dense acquisition frequency, temporal filter methods can effectively reconstruct the missing regions resulting from cloud contamination. These methods are based on the fact that the dense time series tends to display regular fluctuations if the data are ordered in a chronological way. Related work can be found in the studies of Beck *et al.* [18], Malambo and Heatwole [19], and Yang *et al.* [20]. Nevertheless, the temporal filter methods are only applicable to dense time-series data, which often come with a coarse spatial resolution. By contrast, temporal replacement methods have been proposed to be used specifically with high and medium spatial resolution images having a sparse temporal frequency. For this category, a cloud-free multitemporal image is adopted as a reference to recover the contaminated pixels in the target image. An ideal situation is to directly replace the missing pixels with observed pixels from the same region in the reference image [21]. Unfortunately, the radiometric differences in multitemporal images often result in spectral incoherence between the recovered and remaining regions in the synthetic results. To minimize the differences, some studies have further adjusted the brightness of the observed pixels in the reference image in accordance with that in cloud-free regions in the target image, using strategies such as local linear histogram matching (LLHM) [22], the modified neighborhood similar pixel interpolator (MNSPI) [23], and weighted linear regression (WLR) [24]. Besides the temporal replacement methods, another effective way to replace a missing pixel is to directly utilize same-class pixels in the cloud-free regions of the target image itself, with the reference image serving as a guidance to locate similar pixels. Related cases can be found in Meng *et al.* [25], Jin *et al.* [26], and Cheng *et al.* [3]. Recently, some advanced techniques have also been used in this category, such as sparse representation [27]–[30] and deep learning [31]–[32], and they are expected to generate accurate recovered results. In general, compared with other categories, the temporal-based methods are less influenced by the spatial size and landscape heterogeneity of the missing areas, and they can produce synthetic cloud-free results with a higher fidelity. However, the temporal-based methods assume that the spatial coverage of the land cover is fixed during the acquisition interval. Limited by this common assumption, the temporal-based methods are appropriate for scenes with phenological changes, but they cannot handle cases with significant land cover changes (especially those with unfixed object boundaries).

The multisource-based methods may provide us with a possible way to solve this problem via the introduction of observations from other data sources. Unfortunately, there appears to have been no satisfactory solutions for this problem until now. To recover the contaminated regions in a target image, some studies have employed optical images acquired from different remote sensing systems [33]–[34]. For this approach to work, the auxiliary images must have a similar wavelength and spatial resolution to the target image. Essentially, the reconstruction methods in this case are identical to the above-mentioned temporal-based methods, and thus, they cannot generate satisfactory results when major land cover changes occur. Moreover, synthetic aperture radar (SAR) images, which are free from cloud disturbance, have been introduced as auxiliary data by Eckardt *et al.* [35], Huang *et al.* [33], and Li *et al.* [37]. However, SAR images commonly suffer from severe speckle noise, and using them as an auxiliary is likely to decrease the reconstruction accuracy. Another interesting study was done by Roy *et al.* [38], which used multitemporal Landsat-MODIS fusion for gap filling. Although, to the best of our knowledge, they were the first to present the idea that spatiotemporal data fusion has the potential to reconstruct the missing areas in optical images, their work [38] only aimed to develop a new spatiotemporal data fusion model for the production of land surface reflectance with both high spatial and temporal resolutions. It did not adequately transfer the spatiotemporal data fusion model into the cloud removal objective. Thus, they did not provide any experiment for cloud removal.

In this paper, based on the idea of multisource data fusion, we propose a new cloud removal method that fully integrates the multisource data into the cloud removal objective. In the proposed method, a modified spatiotemporal data fusion model is introduced to effectively integrate multisensor observation data and generate initial recovered results with high spatial consistency. Furthermore, a residual correction strategy based on the Poisson equation follows to unite the information from the target image to enhance the spectral coherence between the recovered and remaining regions. In short, the proposed cloud removal method effectively combines the spatiotemporal data fusion model with the residual correction process to produce recovered results with high spatial and spectral consistency. The proposed method is then applied to scenes with significant land cover changes, for which the widely used temporal-based methods show poor recovery ability. The experiments undertaken in this study confirmed the efficacy of the proposed method and revealed that it can acquire plausible visual performance and satisfactory quantitative accuracy for scenes with significant land cover changes.

II. METHOD

A. Fundamental Framework

Suppose we have a cloud-contaminated image (denoted as the target image P_t) acquired on the target date t , then the data source of the given image is termed the primary source. To better handle significant land cover changes, an additional data source, defined as the auxiliary source, is introduced in the proposed

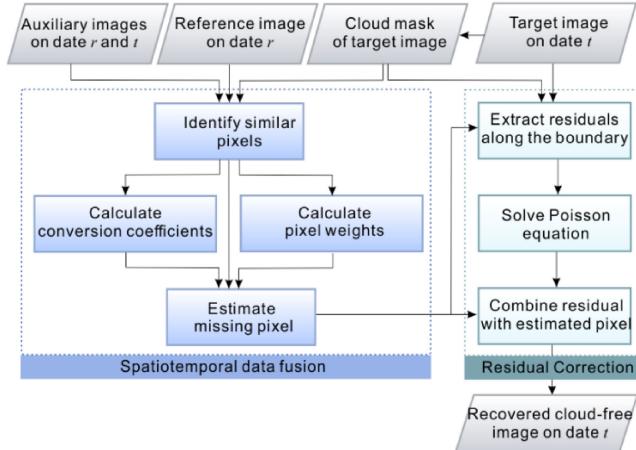


Fig. 1. Flowchart of the proposed cloud removal method.

framework. The auxiliary source is required to have similar wavelengths to that of the primary source and have a relatively high temporal frequency to match the acquisition dates. Due to the tradeoff between the spatial and temporal resolutions, the auxiliary images are usually of a medium or low spatial resolution. More specifically, to recover the target image P_t , we need to select a cloud-free multitemporal image (denoted as the reference image P_r) acquired on the reference date r . Moreover, two cloud-free coarse-resolution images (denoted as auxiliary images A_t and A_r) from the auxiliary source need to be provided. Note that the acquisition dates of the auxiliary images are supposed to be the same as or adjacent to the target date t and reference date r , respectively, which is why the auxiliary source should have a relatively high temporal resolution. Before the implementation of the proposed method, two preprocessing steps are required: upsampling of the two auxiliary images to the same spatial resolution as the target image; and acquisition of a cloud mask of the target image to locate the pixels which need to be reconstructed.

Generally speaking, the basic framework of the proposed method consists of two stages. In the first stage, the reference image P_r and the auxiliary images A_t and A_r are jointly fused to fill the contaminated regions in the target image P_t , and thus, a preliminary cloud-free result is generated. A recently developed fusion method, termed the spatial and temporal nonlocal filter-based data fusion model (STNLFFM), is modified to better handle this case [40]. In the second stage, given that the preliminary recovered regions may display slight spectral incoherence with the remaining regions due to the slight radiometric differences between multiple sensors, a residual correction strategy is used to enhance the spectral consistency. Fig. 1 displays the overall flowchart of the proposed method.

B. Spatiotemporal Data Fusion

Spatiotemporal data fusion aims to counteract the tradeoff between the spatial and temporal resolutions of different sensors, and it was originally developed to generate a Landsat reflectance time series at a daily frequency by blending Landsat and MODIS

observations [39]. Recently, Cheng *et al.* [40] came up with STNLFFM, and the model proved to be superior to some of the classical data fusion algorithms. In this paper, a modified version of STNLFFM is provided to better solve the problem. To be more specific, two improvements are made. First, two fine and coarse spatial resolution image pairs are required for the original STNLFFM as input, which is inappropriate for our situation. In the modified version, the whole weight is removed to support the fusion with only one image pair. Second, the individual weight was previously calculated based on patches, while, in our version, a pixel-scale estimation strategy is considered to increase the computational efficiency.

For a given pixel location, the STNLFFM method employs a linear model to describe the temporal variability between the target image P_t and the reference image P_r . The temporal relationship is then assumed to be scale-invariant, so we can obtain the linear model with the aid of the two auxiliary images A_t and A_r . Furthermore, the STNLFFM method makes full use of the similar pixels to increase the robustness of the prediction. On this basis, the reflectance of a missing pixel can be recovered with the following equation:

$$\hat{P}_t(x, y, B) = \sum_{i=1}^N W(x_i, y_i, B) \times [a(x_i, y_i, B) \times P_r(x_i, y_i, B) + b(x_i, y_i, B)] \quad (1)$$

where $\hat{P}_t(x, y, B)$ denotes the estimated reflectance of a missing pixel located in (x, y) in band B , and it is the unknown that is to be calculated. $P_r(x_i, y_i, B)$ denotes the observed reflectance of the i th similar pixel in band B for the reference image P_r . N is the total number of similar pixels. $a(x_i, y_i, B)$ and $b(x_i, y_i, B)$ are the conversion coefficients which describe the temporal variability, and they can be derived from the two auxiliary images based on the above assumptions. $W(x_i, y_i, B)$ is the weight of the i th similar pixel, and it determines how much the i th similar pixel contributes to the estimation of the missing pixel.

1) Identifying Similar Pixels: Selecting similar pixels is the prerequisite for the calculation of conversion coefficients and pixel weights, and the quality of the similar pixels directly affects the accuracy of the subsequent steps. To guarantee the effectiveness of the search procedure, we take advantage of the reference image and the auxiliary images to locate similar pixels. For a given missing pixel, a local moving window (with a size w) centered in its location is first applied to select potential candidates in the close neighborhood, and then two conditions are set according to the spectral similarity and temporal consistency as follows:

$$|P_r(x, y, B) - P_r(x_i, y_i, B)| < d \times 2^{P_r(x, y, B)} \quad (2)$$

$$\begin{aligned} & ||A_t(x, y, B) - A_r(x, y, B)|| - |A_t(x_i, y_i, B) \\ & - A_r(x_i, y_i, B)| < \sigma \end{aligned} \quad (3)$$

where d is a free parameter that determines the threshold in the first condition in (2). It may slightly vary for different data sources, and it is set as 0.01 for the Landsat TM case, as suggested in [40]. σ is the uncertainty of the temporal difference

between two coarse-resolution auxiliary images. It is primarily caused by the difference in observation conditions. In this paper, it is set as 0.005 for MODIS, as in the study of Gao *et al.* [39]. A pixel is picked if it meets the above two conditions. It is noteworthy that the temporal consistency condition further removes those pixels which exhibit different temporal variability from the missing pixel, and it plays an important role in scenes with significant land cover changes.

It should be noted that the window size w has a direct impact on the fusion performance. If the size is set too small, we cannot identify enough similar pixels for estimating the conversion coefficients in the next step. If it is too large, the computational burden will be increased significantly. After carrying out a trial-and-error test, we decided to set the window size to be 41 in our experiment to balance the fusion accuracy and the computational efficiency.

2) *Calculating Conversion Coefficients*: Based on the supposition that the linear model is scale-invariant, the auxiliary images can be explored to derive the conversion coefficients. Since similar pixels exhibit temporal variability that is consistent with the missing pixel, it is reasonable to assume that they share the same conversion coefficient set. Here, a restricted least-squares model is applied to the similar pixels in auxiliary images A_t and A_r to obtain a and b .

3) *Calculating Pixel Weights*: Although the selected pixels will share evident similarity with the missing pixel, slight differences still exist. A pixel with less difference provides more credible information and, thereby, it should contribute more to the final estimate. In the modified version of STNLFFM, the whole weight is removed to support the fusion with only one coarse-and-fine image pair. Meanwhile, a pixel-based individual weight is used to increase the computational efficiency, with the original form of the nonlocal means filter model preserved. The weight in this paper is calculated as

$$S(x_i, y_i, B) = \exp\left(-\frac{|A_r(x_i, y_i, B) - A_t(x, y, B)|}{h^2}\right) \quad (4)$$

$$W(x_i, y_i, B) = \frac{S(x_i, y_i, B)}{\sum_{i=1}^N S(x_i, y_i, B)} \quad (5)$$

where $S(x_i, y_i, B)$ is the spectral similarity of the i th similar pixel. In STNLFFM, the calculation of $S(x_i, y_i, B)$ takes the basic form of a nonlocal filter [40], h is a parameter related to the noise level of the auxiliary images. Based on our pretesting, we find it has little influence on the fusion results, so h is set to 0.15 in our experiment, as in [40]. After normalization with (5), the numerical range of the i th similar pixel weight is $[0, 1]$, and the sum of all the similar pixel weights equals 1.

4) *Estimating Missing Pixels*: After the calculation of conversion coefficients and pixel weights, the estimate of the missing pixel can be generated through (1), and, thus, all the contaminated patches can be preliminarily recovered. Additionally, it should be noted that pixels along the boundaries outside the contaminated patches also need to be estimated, because they are a necessary input for the residual correction in the second stage.

C. Residual Correction

Although we assume that the introduced auxiliary source is spectrally comparable with the primary source, it is likely that there will still be some radiometric differences between multiple sensors. As a result, the recovered patches generated from the spatiotemporal data fusion may display slight spectral inconsistency with the cloud-free regions in target image P_t , especially in the edge areas. Therefore, a residual correction step is required to minimize the incoherence.

To this end, a patch-based adjustment strategy is embedded in the proposed method. This strategy was originally developed for seamless image cloning by Pérez *et al.* [41]. It mathematically formulates the problem as Poisson equation and solves it using a global optimization process. For the target image P_t , we denote the spatial coverage of a cloud-contaminated patch as Ω (the boundary of the patch is $\partial\Omega$), while the cloud-free coverage around the contaminated patch is termed Ω^* . The cloud-free image after residual correction can be viewed as an intensity function, which consists of two parts. The first part is an unknown function f defined over Ω , with $f = \{P_t(x, y, B) | (x, y) \in \Omega\}$, which is what we want to obtain. The second part is a known function f^* defined over Ω^* , with $f^* = \{P_t(x, y, B) | (x, y) \in \Omega^*\}$. The preliminary recovered patch in the first stage is defined as the function g over Ω , with $g = \{\hat{P}_t(x, y, B) | (x, y) \in \Omega\}$. Specifically, the gradient of the preliminary recovered patch g is employed as a guidance vector field to guide the optimization. A boundary condition is used to solve the optimization equation, and the adjusted patch f can be obtained as follows:

$$\min_f \int \int_{\Omega} |\nabla f - \nabla g|^2, \text{ with } f|_{\partial\Omega} = f^*|_{\partial\Omega} \quad (6)$$

where $\nabla = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y})$ is the gradient operator. The adjusted recovered patch ρ derived from (6) satisfies two conditions: the first is that the gradient of the adjusted patch f is as close to that of the preliminary recovered patch g as possible, and the second is that the transition from the adjusted patch f to the cloud-free regions in the target image P_t is smooth. To solve (6), a helpful alternative way using Poisson equation is provided. We further define a residual function f' over Ω , such that $f = g + f'$. Since the pixels along the boundary $\partial\Omega$ are also estimated in the first stage, the mismatch $(f^* - g)$ over $\partial\Omega$ can be obtained, with $f^* - g = \{P_t(x, y, B) - \hat{P}_t(x, y, B) | (x, y) \in \partial\Omega\}$. The residual function f' can then be viewed as a membrane interpolation of the mismatch $(f^* - g)$ along the boundary $\partial\Omega$. Thus, the residual function f' can be solved by the following Poisson equation with a boundary condition:

$$\Delta f' = 0 \text{ over } \Omega \text{ with } f'|_{\partial\Omega} = (f^* - g)|_{\partial\Omega} \quad (7)$$

where $\Delta = (\frac{\partial^2}{\partial x^2}, \frac{\partial^2}{\partial y^2})$ is the Laplacian operator. The residual function f' can be interpolated with (7), and the adjusted patch (i.e., the unknown function f) can be produced by adding the residual back to the preliminary recovered patch g . A brief implementation procedure for the residual correction strategy is presented below.

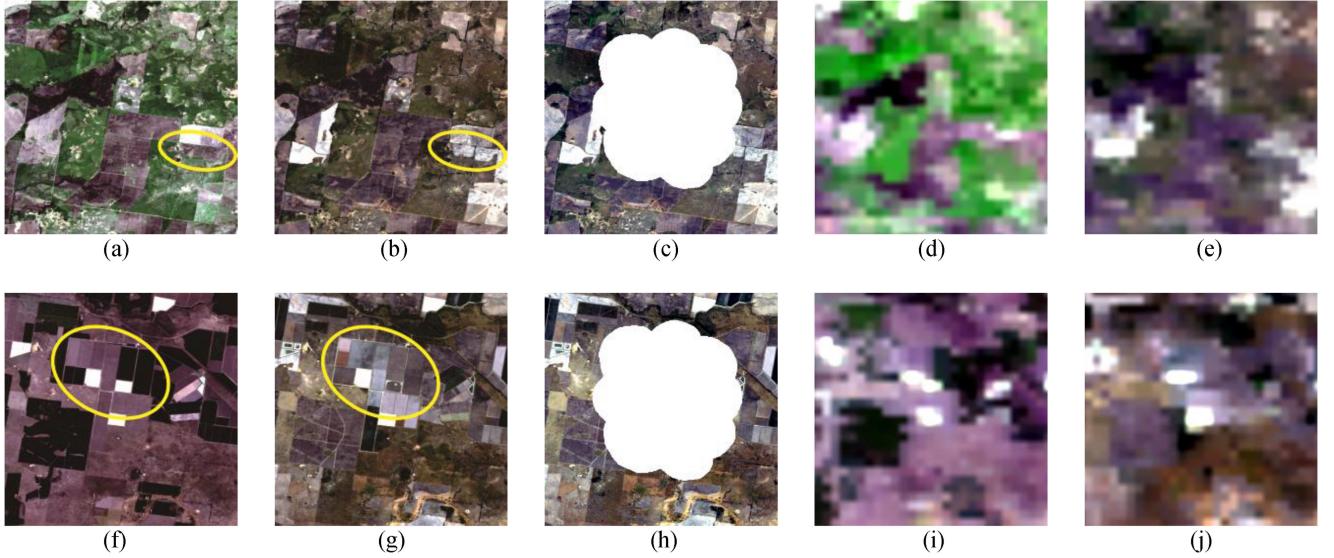


Fig. 2. Two datasets used for the algorithm testing (with the red, green, and blue bands as RGB). The first row displays the data of site 1. (a)–(b) Landsat images of December 28, 2004, and February 14, 2005, respectively. (c) Cloud-contaminated image simulated from (b). (d)–(e) Resampled MODIS images for December 28, 2004, and February 14, 2005, respectively. The second row displays the data of site 2. (f)–(g) Landsat images of August 22, 2004, and April 3, 2005, respectively. (h) Cloud-contaminated image simulated from (g). (i)–(j) Resampled MODIS images for August 22, 2004, and April 3, 2005, respectively.

1) Extracting Residuals Along the Patch Boundary: As mentioned above, a mask identifying the location of clouds is required beforehand. With the aid of the Laplacian convolution kernel, external boundaries $\partial\Omega$ of the cloud patches can be accurately extracted from the mask. Since we also estimate the pixels along the boundary through the spatiotemporal data fusion in the first stage, the residuals ($f^* - g$) along the boundary can be obtained between the estimated pixels and the original observed pixels in the target image P_t .

2) Solving Poisson Equation: A Laplacian coefficient matrix should be first derived. With the residuals ($f^* - g$) along the boundary $\partial\Omega$, we can successfully build Poisson equation [i.e., (7)]. The residual function f' defined over the contaminated region Ω can be calculated by solving the equation, and we can thus obtain the residual of each pixel in the preliminary recovered patch.

3) Combining Residuals With the Preliminary Estimated Pixels: The residual of each pixel is added back to the preliminary recovered pixel in the first stage, and then the adjusted pixel after residual correction is estimated. All the contaminated pixels in the target image P_t are replaced by the adjusted pixels, and thus, the final cloud-free image is produced.

III. EXPERIMENTS AND RESULTS

Landsat 5 images (bands 1–5, and 7) were selected to conduct the experiments, and MODIS images (bands 3, 4, 1, 2, 6, and 7), which show good spectral compatibility with Landsat images, were used as the auxiliary source. A time-series dataset, which has been extensively used in data fusion research [42]–[43], was employed in the experiments. This dataset contains 14 available Landsat-MODIS image pairs of the Lower Gwydir Catchment in Australia, observed from April 2004 to April 2005. All the

Landsat images were acquired by the TM sensor, and they were atmospherically corrected using the algorithm proposed by Li *et al.* [44]. MODIS Terra MOD09GA Collection six daily reflectance products were used and geographically rectified with the Landsat data. Given that the Landsat images in the dataset were resampled to 25 m, it was necessary that we resample the MODIS images to the same spatial resolution using the cubic convolution method. All the test data used were typical subsets cropped from the time-series dataset.

To assess the reconstruction performance of the proposed method, benchmarking was introduced. The following four temporal-based methods were employed to enable a comparative analysis: LLHM [22], MNSPI [23], WLR [24], and patch matching-based multitemporal group sparse representation (PM-MTGSR) [30]. Also, given that spatiotemporal data fusion can be considered an effective strategy for reconstruction, we also compared our work with STARFM proposed by Gao *et al.* [39]. We first got a fused result from STARFM, using the reference and auxiliary images as input, and then replaced the contaminated regions in the target image with the fused result.

A. Experiments With Simulated Cloud Contamination

Since we mainly target special scenes dominated by significant land cover changes, two typical sites were selected. As displayed in Fig. 2, two 500×500 subsets covering a same overall area of 156.25 km^2 were cropped to conduct the simulated tests. Each group dataset comprised four observed images, with two from the Landsat source and two from the MODIS source. For the first site, images acquired on December 28, 2004 (as the reference date), and February 14, 2005 (as the target date), were selected. Fig. 2(a) and (b) shows that obvious land cover changes exist in the region marked with yellow. A

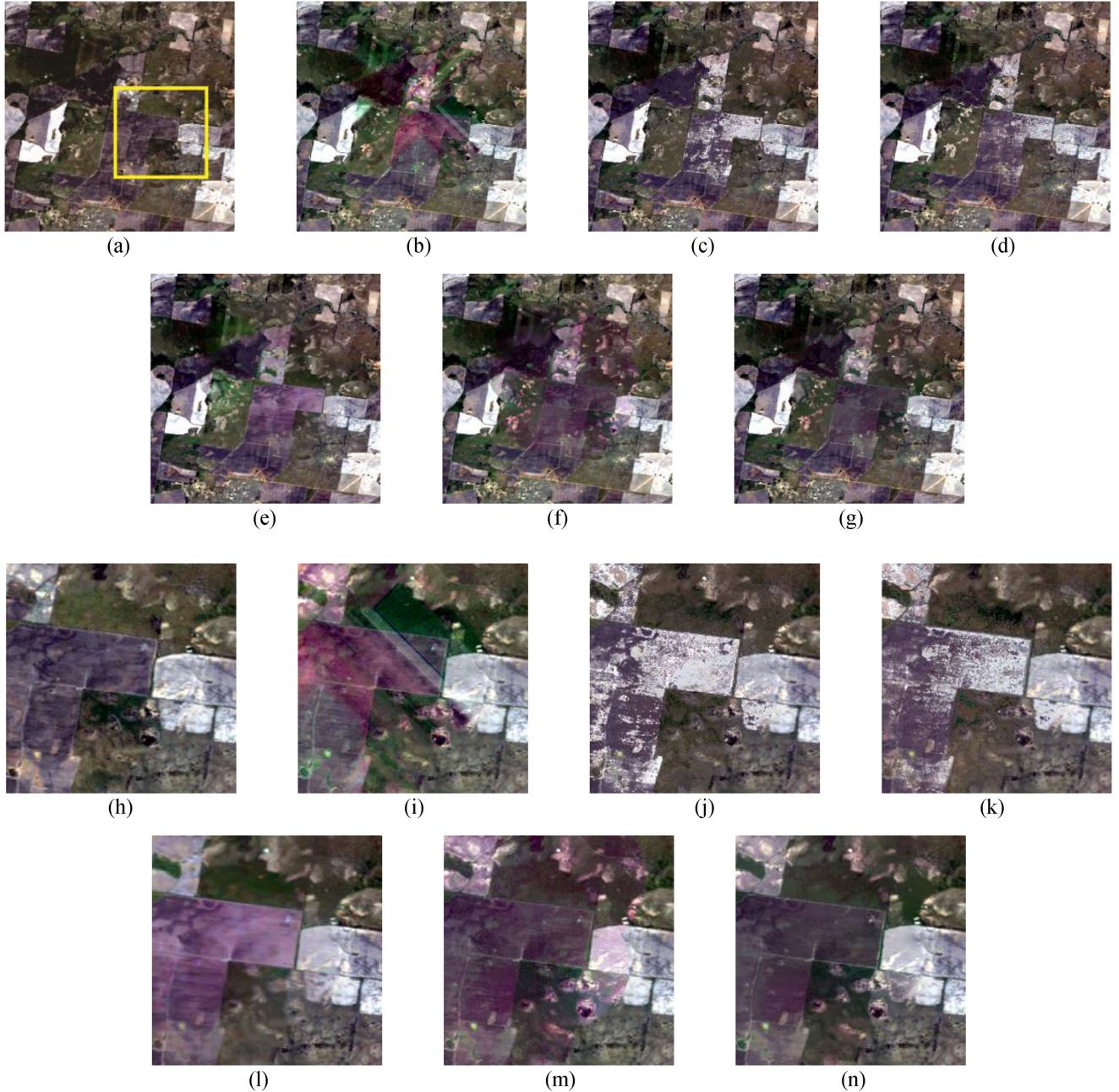


Fig. 3. Recovered results for site 1 (with the red, green, and blue bands as RGB). (a) Original cloud-free Landsat image of February 14, 2005. (b)–(g) Images recovered by LLHM, MNSPI, WLR, PM-MTGSR, STARFM, and the proposed method, respectively. (h)–(n) Zoomed-in views of the subset region marked in yellow in (a).

cloud-contaminated region, accounting for approximately 35% of the whole image, was simulated in the central part of the original clear image, as shown in Fig. 2(c). For the second site, images acquired on August 22, 2004 (as the reference date), and April 3, 2005 (as the target date), were used. Significant land cover changes can be found in the marked region [see Fig. 2(f) and (g)]. Fig. 2(h) shows a simulated obscured image with the same manual cloud mask as Fig. 2(c). It is noteworthy that the landscape of the second site is more fragmented, and it experienced more significant changes due to its longer acquisition interval between multitemporal images.

The recovered images generated by the six methods were compared with the original clear images, from both visual and quantitative fidelity aspects. Two widely used indices—the root-mean-square error (RMSE) and the correlation coefficient (CC)—were adopted to allow the quantitative assessment. A lower RMSE and a higher CC indicate better consistency of recovered result with the target image.

For the first site, the original cloud-free image and six recovered images are displayed in Fig. 3, together with a detailed view of a subset region. Judging from all the recovered images, we can see that the LLHM method [see Fig. 3(b)] produces serious directional artifacts and obvious spectral discrepancy in the

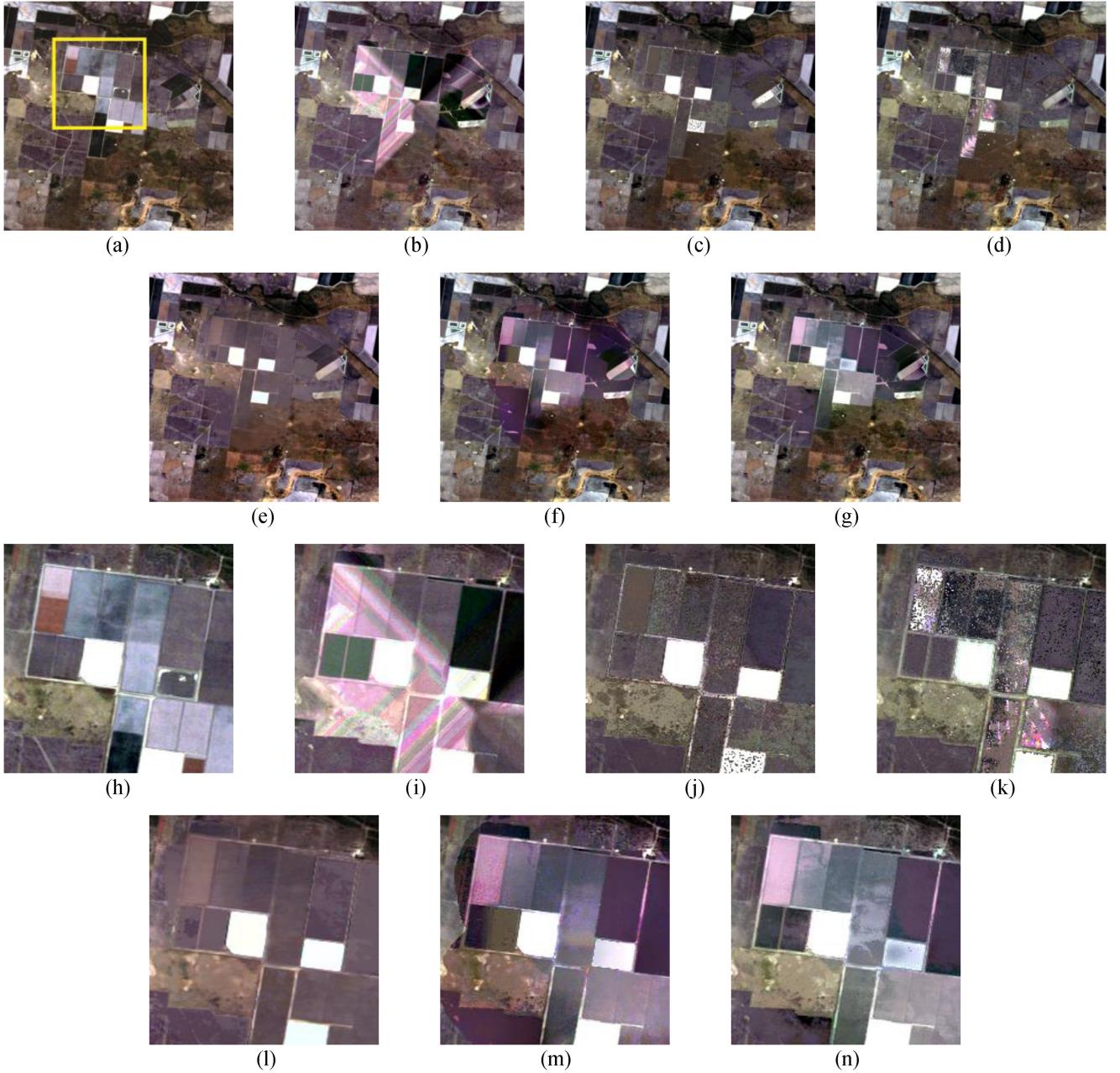


Fig. 4. Recovered results for site 2 (with the red, green, and blue bands as RGB). (a) Original cloud-free Landsat image of April 3, 2005. (b)–(g) Images recovered by LLHM, MNSPI, WLR, PM-MTGSR, STARFM, and the proposed method, respectively. (h)–(n) Zoomed-in views of the subset region marked in yellow in (a).

central part of the contaminated region. As all the valid pixels in the moving window are involved in the coefficient estimation, and that without any optimal filtration, the recovered image shows obvious errors. The results of the MNSPI [see Fig. 3(c)] and WLR [see Fig. 3(d)] methods reveal some close parallels. The detailed zoomed-in regions in Fig. 4(j) and (k) exhibit evident reconstruction mistakes. This is primarily caused by the incorrect temporal information borrowed from the adjacent areas with significant changes. PM-MTGSR obtains a plausible result, but the recovered spectral signals [shown in the zoomed in image in Fig. 3(l)] are slightly different from that in the target image. The two spatiotemporal-based methods generate results that are

visually closest to the target image. Specifically, the STARFM result is spectrally inconsistent in the boundary of the reconstruction region, as shown in Fig. 3(m). The proposed method avoids this by implementing the residual correction strategy.

The six recovered images and the original clear image of the second site are shown in Fig. 4; again, a detailed region is provided to facilitate a visual comparison. In this case, the restored area of the LLHM method [see Fig. 4(b)] presents serious directional artifacts, and LLHM cannot capture the significant temporal changes. The MNSPI [see Fig. 4(c)] and WLR [see Fig. 4(d)] methods provide similar results again. Both their restored results suffer from noise-like errors, and the changed

TABLE I
QUANTITATIVE ASSESSMENT OF THE SIMULATED TEST FOR THE FIRST SITE

	LLHM	MNSPI	WLR	PM-MTGSR	STARFM	Proposed
RMSE	Red	0.009	0.013	0.011	0.009	0.007
	Green	0.012	0.016	0.014	0.011	0.009
	Blue	0.016	0.019	0.017	0.014	0.012
	NIR	0.040	0.030	0.027	0.023	0.024
	SWIR-1	0.040	0.028	0.027	0.025	0.023
	SWIR-2	0.033	0.027	0.025	0.022	0.027
	Mean	0.025	0.022	0.020	0.017	0.016
CC	Red	0.793	0.743	0.775	0.813	0.861
	Green	0.748	0.643	0.700	0.759	0.840
	Blue	0.732	0.685	0.727	0.764	0.827
	NIR	0.822	0.813	0.850	0.898	0.879
	SWIR-1	0.655	0.753	0.759	0.767	0.817
	SWIR-2	0.785	0.852	0.859	0.870	0.889
	Mean	0.756	0.749	0.778	0.812	0.855

TABLE II
QUANTITATIVE ASSESSMENT OF THE SIMULATED TEST FOR THE SECOND SITE

	LLHM	MNSPI	WLR	PM-MTGSR	STARFM	Proposed
RMSE	Red	0.024	0.015	0.020	0.016	0.012
	Green	0.027	0.020	0.027	0.021	0.016
	Blue	0.045	0.026	0.038	0.024	0.020
	NIR	0.186	0.073	0.091	0.044	0.058
	SWIR-1	0.066	0.041	0.044	0.037	0.029
	SWIR-2	0.092	0.033	0.053	0.028	0.029
	Mean	0.073	0.035	0.045	0.028	0.026
CC	Red	0.608	0.749	0.723	0.734	0.881
	Green	0.666	0.777	0.765	0.745	0.906
	Blue	0.627	0.748	0.716	0.742	0.891
	NIR	-0.471	0.164	0.071	0.726	0.521
	SWIR-1	0.482	0.704	0.675	0.746	0.845
	SWIR-2	0.025	0.701	0.400	0.789	0.790
	Mean	0.323	0.640	0.558	0.747	0.808

spectral signature is not well recovered. The detailed region shown in Fig. 4(j) and (k) reveals that the WLR result is even more unsatisfactory in this case. PM-MTGSR [see Fig. 4(e)] fails to reconstruct the changed land cover. In comparison, both STARFM [see Fig. 4(f)] and the proposed method [see Fig. 4(g)] produce a relatively favorable result when compared with that of the other methods. The changed land covers are basically acquired, and the recovered spectral signatures are much closer to those of the original clear image than other methods. It should be noted that the recovery ability of the proposed method for small ground features is not optimal, because MODIS observations with a 500-m spatial resolution are spectrally mixed in heterogeneous areas and are insufficient to provide temporal information of small features.

A quantitative assessment of all the recovered results was performed. Tables I and II list the RMSE and CC values for the two tests. The conclusions from the visual inspection are also supported by the quantitative indices. For both tests, the proposed method obtains the lowest RMSE and highest CC values in most bands, indicating that it achieves improved performance. Additionally, as mentioned above, the second site experienced more dramatic changes in land cover due to its longer acquisition interval. For this case, the proposed method is superior to temporal-based methods, with better RMSE (lower)

and CC (higher) values. It can be concluded that the proposed method shows an evident improvement over the other methods for scenes with significant land cover changes. Regarding the reconstruction ability in each band, we find that the proposed method achieves best performance in most bands except NIR. A much narrower NIR bandwidth is employed in the MODIS sensor than that in the TM case (MODIS: 841–876 nm; TM: 760–900 nm), and the inconsistency in NIR band may go against the fundamental assumption of spatiotemporal fusion that both target source (i.e., TM) and auxiliary source (i.e., MODIS) are spectrally comparable. As a consequence, the proposed method may not perform as well as expected in NIR band.

B. Experiments With Time-Series Images

To further explore the robustness of the proposed algorithm, a set of simulated experiments based on a Landsat-MODIS time series was conducted. The second site from Section III-A. was used, due to the dramatic changes it experienced during the acquisition interval. The time series, which consists of 14 available Landsat-MODIS image pairs, indicates that the landscape underwent different kinds of changes over time. Specifically, the site was mainly dominated by phenological changes before December 12, 2004, while it experienced significant land cover changes thereafter.

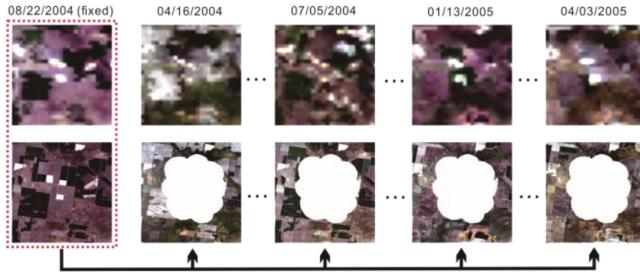


Fig. 5. Illustration of the cloud-removal procedure for the proposed method. Each contaminated image on the target date was recovered using a fixed Landsat-MODIS reference pair and the MODIS image on the target date.

To conduct the experiment, August 22, 2004, was fixed as the reference date, and the other 13 dates were used as target dates. Thirteen Landsat images were imposed with the same simulated contamination as that described in Section III-A. A cloud removal procedure was then carried out for each target image, assisted by the cloud-free observations on the fixed reference date. To be specific, for four temporal-based methods (i.e., LLHM, MNSPI, WLR, and PM-MGTSR), the contaminated regions of the 13 target images were recovered, respectively, using the fixed cloud-free Landsat image as reference; for STARFM and the proposed method, each target image was recovered using the fixed Landsat-MODIS reference image pair and the corresponding MODIS image acquired on the same date, as shown in Fig. 5. The evaluation of algorithm performance was conducted in a quantitative way. In addition to RMSE and CC, another two indices—the average absolute difference (AAD) and the universal image quality index (UIQI)—were adopted for further comparison. A lower AAD and a higher UIQI indicate better reconstruction accuracy.

The quantitative assessment of the recovered results is displayed in Fig. 6. We can see that for most cases before December 12, 2004 (i.e., cases dominated by phenological changes), PM-MGTSR performs better than the other methods. For most cases after December 12, 2004, when significant land cover changes occurred, the spatiotemporal fusion based methods acquire the lowest AAD and RMSE values and the highest CC and UIQI values; thus, demonstrating that they are superior to the temporal-based methods. Specifically, the proposed method performs slightly better than STARFM. The results indicate that the temporal-based methods are able to achieve satisfactory results with less input in cases dominated by phenological changes, whereas the proposed method is good at handling scenes with significant land cover changes.

C. Experiments With Observed Cloud Contamination

The cloud removal performance is mainly judged by visual inspection for images with observed cloud cover. For scenes with phenological changes, we can distinguish the performance in terms of spectral coherence and spatial adjacency. Nevertheless, when major land cover changes occur, we cannot confirm what exactly the land covers have turned into, even if we had multitemporal images as a reference, so the recovered results of this kind of scene are often unverifiable.

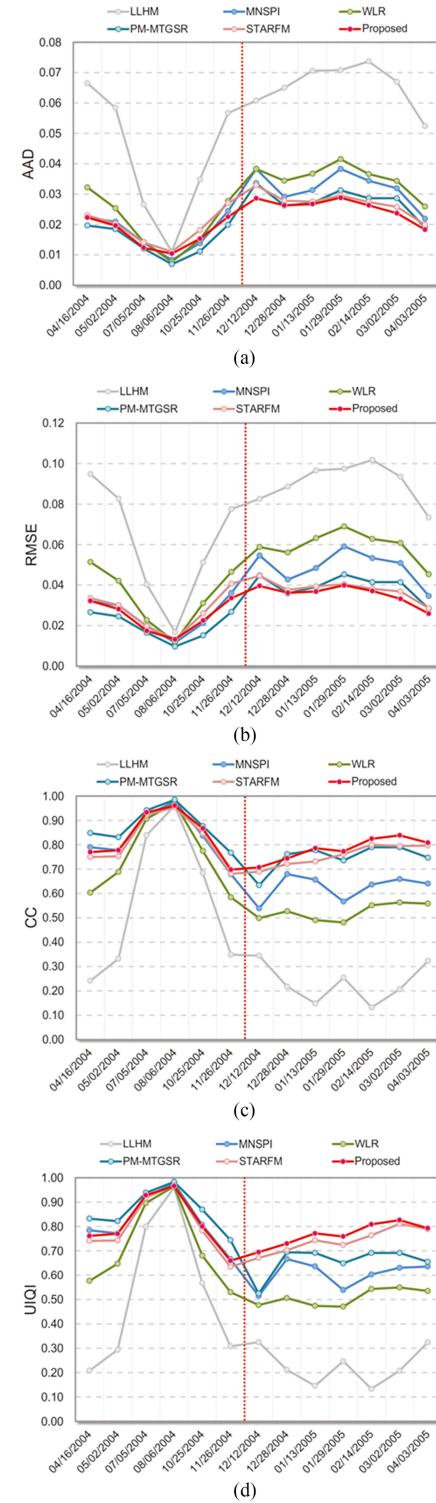


Fig. 6. Quantitative assessment results for the time-series experiments. (a) AAD. (b) RMSE. (c) CC. (d) UIQI. Red dashed line indicates the point when significant land cover changes occurred.

To solve this problem, a special case is provided in Fig. 7. The chosen site covers an approximate area of 265 km² (650 columns × 650 lines at 25-m resolution). The Landsat image of January 29, 2005, is covered with cloud, as shown in Fig. 7(b). We can see the approximate ground truth in the contaminated region through the thin cloud, and thus, the target image itself can be



Fig. 7. Data with observed clouds used for the algorithm testing (with the red, green, and blue bands as RGB). (a)–(b) Landsat images of October 25, 2004, and January 29, 2005, respectively. (c)–(d) Resampled MODIS images corresponding to (a) and (b). (e) Cloud and cloud shadow mask of (b) obtained by visual interpretation.

used for a visual comparison. The Landsat image acquired on October 25, 2004 [see Fig. 7(a)] was taken as the reference in this experiment. Significant land cover changes took place in the contaminated region during the acquisition interval, and the whole region turned into several land patches with distinctly different spectral characteristics. Two MODIS images acquired on the same date as the Landsat images were also used, as shown in Fig. 7(c) and (d). Notably, due to the cloud mobility, the MODIS image acquired on January 29, 2005 was actually cloud-free. Given that cloud detection was not our primary concern in this study, a cloud and cloud shadow mask of Fig. 7(b) produced by visual interpretation is provided in Fig. 7(e).

The observed cloud-contaminated image, the six restored cloud-free results, and images of the detailed subset region [see as indicated in Fig. 8(a)] are displayed in Fig. 8. For the result of the LLHM method [see Fig. 8(b)], the recovered area presents serious spectral inconsistency compared with the ground truth, and LLHM fails to capture the land cover changes. For the results of MNSPI, WLR, and PM-MTGSR, a significantly changed region [see Fig. 8(j), (k), and (l)] is zoomed in on for a detailed comparison. It can be seen that none of the methods effectively recover the white patch in Fig. 8(h). In comparison, both STARFM [see Fig. 8(f)] and the proposed method [see Fig. 8(g)] give the visually closest result to the ground truth, and the white patch is successfully recovered. This is primarily because the auxiliary MODIS images provide more reliable temporal information for this area. Also, the result of STARFM [see Fig. 8(m)] presents spectral distortion, whereas the result of the proposed method does not suffer from this problem. It should be mentioned that some small features cannot be captured by the proposed method well. The large zoom factor of approximately 20 between the Landsat and MODIS observations causes the information loss of small features in the MODIS images, resulting in the reduced performance.

IV. DISCUSSION

The temporal-based methods currently favored by the remote sensing community are incapable of handling scenes with significant land cover changes, especially those with unfixed object boundaries. There is, therefore, a need to develop new solutions to these special cases. In this paper, we have presented a cloud removal approach for scenes with significant land cover changes. The experimental comparison with five benchmark methods confirmed the reliability of the proposed method, i.e.,

it can effectively restore the changed land cover under clouds. The quantitative evaluation indicated that the proposed method can produce more accurate results and gain a more robust performance than other methods if the scene experienced great land cover changes. It's noteworthy that our method specializes in scenes with significant land cover changes. In practical applications, the selection of the cloud removal method (the temporal-based or the newly proposed) should be based on specific conditions. The temporal-based group is more suitable for phonological changes with fixed spatial tendency, whereas the proposed method is recommended for land cover changes having unfixed spatial tendency.

In this study, Landsat and MODIS data were chosen to conduct the experiments, but it should be noted that the proposed method could also be applied to other data source combinations. For example, the Sentinel-3A OLCI sensor revisits the same area with a temporal resolution of <2.8 days, and the frequency will even be shortened to <1.4 days after the launch of the twin satellite [45]. It therefore holds the potential to replace the MODIS data as auxiliary data in the reconstruction procedure. In addition, the twin Sentinel-2A and -2B satellites together deliver optical images every five days, and their high spatial resolution at 10 or 20 m produces even more detailed texture, so they could also be an alternative choice if the acquisition dates match up with the desired ones.

A. Technical Innovation of the Proposed Method

For scenes with significant land cover changes, the introduction of auxiliary images was the greatest contribution to the improved performance of the proposed method. There is a high possibility that the spatial coverage of objects varies over time in these scenes. As a result, similar pixels in the reference image may experience different temporal variability during the acquisition interval. This is the reason why the temporal-based methods have a poor ability to handle these special cases. In comparison, the two introduced auxiliary images are fully exploited to capture the temporal variability of the land cover in the newly developed method. Pixels in the auxiliary images tend to provide more reliable information, since they present temporal variability that is consistent with the missing pixels. The improved idea plays a key role for scenes which have undergone major land cover changes.

The implementation considerations in the spatiotemporal fusion model also benefit the cloud removal performance. Abundant similar information always exists within an image, since

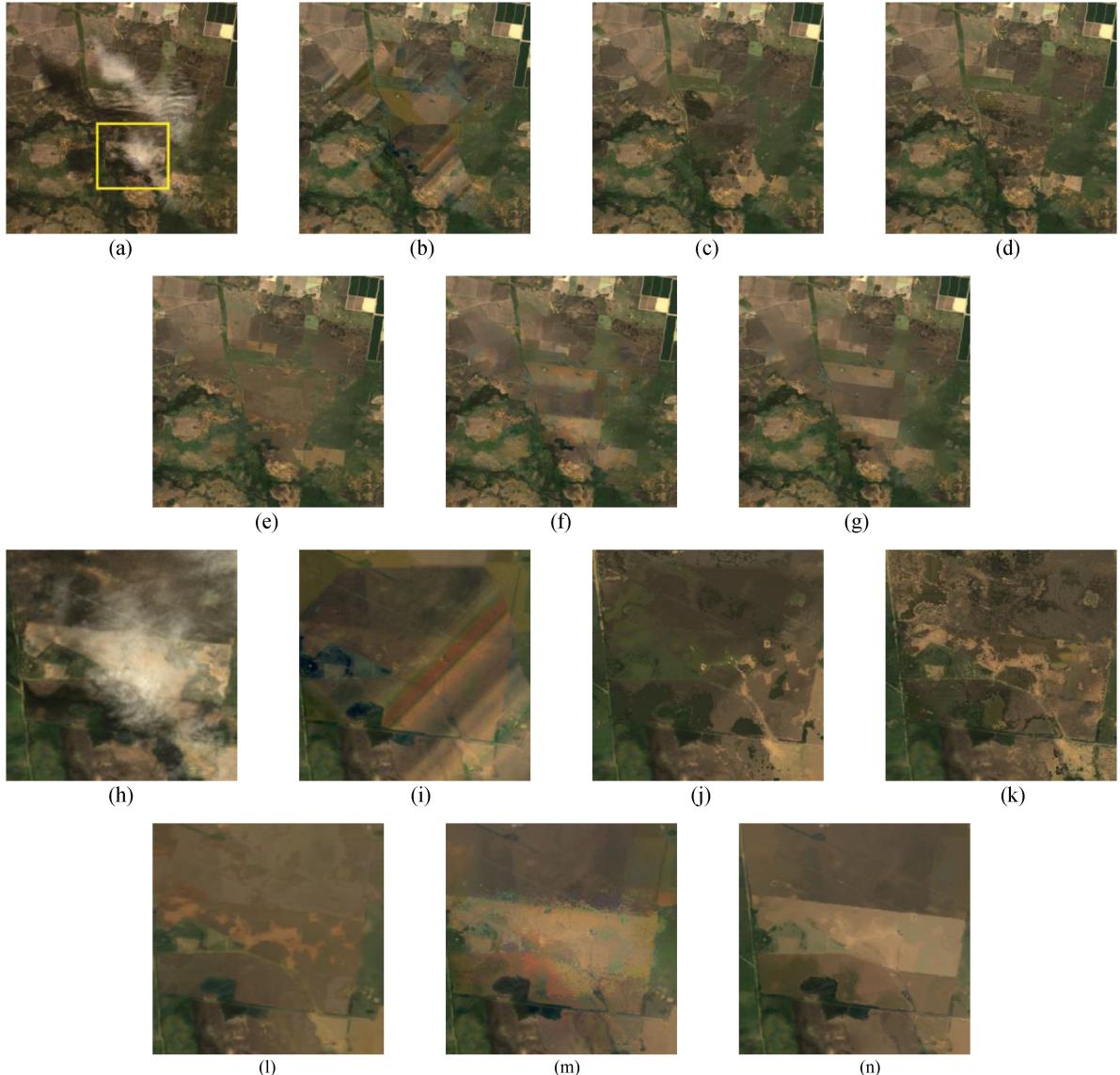


Fig. 8. Recovered results for the test with observed cloud (with the red, green, and blue bands as RGB). (a) Original observed Landsat image with cloud contamination from January 29, 2005. (b)–(g) Images recovered by LLHM, MNSPI, WLR, PM-MTGSR, STARFM, and the proposed method, respectively. (h)–(n) Zoomed-in views of the subset region marked in yellow in (a).

satellite-based observations usually cover a wide spatial range. In order to enhance the accuracy and robustness of the predictions, the data fusion model takes full advantage of the similar pixels to describe the temporal variability of the land cover. In addition, the previous temporal-based methods only set a condition of spectral similarity to select similar pixels, while the other standard of temporal consistency is used to weed out those pixels with different temporal variability in this study. The stricter filtering conditions make a critical difference for scenes dominated by land cover changes.

Another innovation of the proposed method is the employment of residual correction. Given that some intrinsic differences of multisource data always exists in their radiometric characteristics, the preliminary recovered patches generated from the spatiotemporal data fusion may exhibit slight visual disruption with their cloud-free neighborhood. To eliminate the spectral inconsistency, a residual correction procedure is used

to adjust the estimated pixels in accordance with those in the remaining regions. The introduced method mathematically formulates the adjustment problem as Poisson equation, and solves it using a global optimization process. A distinct advantage is that the residual correction is conducted based on the patches rather than pixels, and thus, the computational efficiency can be expected to be improved.

B. Limitations of the Proposed Method

Undeniably, some shortcomings exist in the proposed method. First, besides the close radiometric characteristics with the target source, the auxiliary source should have a relatively high temporal resolution to match the acquisition dates of the reference and target images. However, due to the tradeoff between spatial resolution and swath width, the auxiliary data source usually has a low spatial resolution. For example, the MODIS images used in this paper, with a spatial resolution of 500 m, are insuf-

ficient to capture details of some small features, especially over heterogeneous landscapes. Under this circumstance, the modeled temporal variability may have some errors, thus, reducing the reconstruction accuracy. A recent study attempted to fuse 250-m MODIS red and NIR bands with 500-m MODIS bands, using the area-to-point regression kriging approach to produce 250-m MODIS bands and enhance the spatiotemporal fusion [46]. The generated 250-m MODIS images could take the place of the original 500-m images to improve the final accuracy.

Second, some inherent radiometric differences inevitably exist in the multisource data. As a result of these differences, the generated results from spatiotemporal data fusion may have some visual disruption between the reconstructed areas and the original cloud-free areas. In this paper, a patch-based residual correction procedure based on Poisson equation was used to ease the problem. Some radiometric normalization algorithms for multisource data could be embedded into the method to reduce the spectral inconsistency, and thus, more accurate cloud-free results may be obtained.

V. CONCLUSION

In this paper, a cloud removal method based on spatiotemporal data fusion has been proposed, specifically aiming at scenes with significant land cover changes. In addition to a cloud-free multitemporal image, we further introduce two auxiliary images from another data source into the reconstruction process. Three aspects in the proposed method contribute to the final estimation. First, the two auxiliary images are employed to capture the temporal variability of the land cover, and thus, the problem encountered by the temporal-based methods can be avoided. The similar information in the remote sensing scenes is then fully utilized to enhance the robustness of the prediction. Finally, a residual correction procedure based on Poisson equation is incorporated into the proposed method to enhance the spectral coherence between the recovered and remaining regions. Both simulated and real-data experiments demonstrated that the proposed method can achieve a satisfactory performance for scenes with significant land cover changes, compared with the existing temporal-based methods. The proposed method acts a technical supplement to the current thick cloud removal framework, making it possible to handle scenes with significant land cover changes. In our future work, the proposed method will be combined with temporal-based methods to build a complete cloud removal framework, realizing the goal of dealing with different kinds of changes.

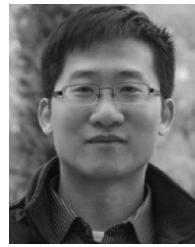
ACKNOWLEDGMENT

The authors would like to thank the editors and anonymous reviewers for their valuable suggestions.

REFERENCES

- [1] M. A. Wulder, J. C. White, S. N. Goward, J. G. Masek, J. R. Irons, and M. Herold, "Landsat continuity: Issues and opportunities for land cover monitoring," *Remote Sens. Environ.*, vol. 112, pp. 955–969, Mar. 2008.
- [2] M. A. Wulder, J. C. White, T. R. Loveland, C. E. Woodcock, A. S. Belward, and W. B. Cohen, "The global Landsat archive: Status, consolidation, and direction," *Remote Sens. Environ.*, vol. 185, pp. 271–283, Nov. 2016.
- [3] Q. Cheng, H. Shen, L. Zhang, Q. Yuan, and C. Zeng, "Cloud removal for remotely sensed images by similar pixel replacement guided with a spatio-temporal MRF model," *ISPRS J. Photogramm. Remote Sens.*, vol. 92, pp. 54–68, Jun. 2014.
- [4] J. Ju and D. P. Roy, "The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally," *Remote Sens. Environ.*, vol. 112, pp. 1196–1211, Mar. 2008.
- [5] H. Shen, X. Li, Q. Cheng, C. Zeng, G. Yang, and H. Li, "Missing information reconstruction of remote sensing data: A technical review," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 61–85, Sep. 2015.
- [6] C. Zhang, W. Li, and D. Travis, "Gaps–fill of SLC-off landsat ETM+ satellite image using a geostatistical approach," *Int. J. Remote Sens.*, vol. 28, pp. 5103–5122, Nov. 2007.
- [7] F. Van der Meer, "Remote-sensing image analysis and geostatistics," *Int. J. Remote Sens.*, vol. 33, pp. 5644–5676, Mar. 2012.
- [8] H. Shen and L. Zhang, "A MAP-based algorithm for desriping and inpainting of remotely sensed images," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 5, pp. 1492–1502, May 2009.
- [9] A. Maalouf, P. Carre, B. Augereau, and C. Fernandez-Maloigne, "A bandlet-based inpainting technique for clouds removal from remotely sensed images," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2363–2371, Jul. 2009.
- [10] Q. Cheng, H. Shen, L. Zhang, and P. Li, "Inpainting for remotely sensed images with a multichannel nonlocal total variation model," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 175–187, Jan. 2014.
- [11] D. Pathak, P. Krähnäbühl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2536–2544.
- [12] C. H. Lin, P. H. Tsai, K. H. Lai, and J. Y. Chen, "Cloud removal from multitemporal satellite images using information cloning," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 232–241, Jan. 2013.
- [13] C. H. Lin, K. H. Lai, Z. B. Chen, and J. Y. Chen, "Patch-based information reconstruction of cloud-contaminated multitemporal images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 163–174, Jan. 2014.
- [14] Y. Zhang, B. Guindon, and J. Cihlar, "An image transform to characterize and compensate for spatial variations in thin cloud contamination of Landsat images," *Remote Sens. Environ.*, vol. 82, pp. 173–187, Oct. 2002.
- [15] I. Gladkova, M. D. Grossberg, F. Shahriar, G. Bonev, and P. Romanov, "Quantitative restoration for MODIS band 6 on aqua," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2409–2416, Jun. 2012.
- [16] P. Rakwatin, W. Takeuchi, and Y. Yasuoka, "Restoration of aqua MODIS band 6 using histogram matching and local least squares fitting," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 2, pp. 613–627, Feb. 2009.
- [17] H. Shen, C. Zeng, and L. Zhang, "Recovering reflectance of AQUA MODIS band 6 based on within-class local fitting," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 1, pp. 185–192, Mar. 2011.
- [18] P. S. A. Beck, C. Atzberger, K. A. Høgda, B. Johansen, and A. K. Skidmore, "Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI," *Remote Sens. Environ.*, vol. 100, pp. 321–334, Feb. 2006.
- [19] L. Malambo and C. D. Heatwole, "A Multitemporal profile-based interpolation method for gap filling nonstationary data," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 252–261, Jan. 2016.
- [20] G. Yang, H. Shen, L. Zhang, Z. He, and X. Li, "A moving weighted harmonic analysis method for reconstructing high-quality SPOT VEGETATION NDVI time-series data," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 6008–6021, Nov. 2015.
- [21] L. Min, L. Soo Chin, and K. Leong Keong, "Producing cloud free and cloud-shadow free mosaic from cloudy IKONOS images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2003, pp. 3946–3948.
- [22] J. Storey, P. Scaramuzza, G. Schmidt, and J. Barsi, "Landsat 7 scan line corrector-off gap-filled product development," in *Proc. Pecora 16 Global Priorities Land Remote Sens.*, 2005, pp. 23–27.
- [23] X. Zhu, F. Gao, D. Liu, and J. Chen, "A modified neighborhood similar pixel interpolator approach for removing thick clouds in landsat images," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 3, pp. 521–525, May 2012.
- [24] C. Zeng, H. Shen, and L. Zhang, "Recovering missing pixels for Landsat ETM+ SLC-off imagery using multi-temporal regression analysis and a regularization method," *Remote Sens. Environ.*, vol. 131, pp. 182–194, Apr. 2013.
- [25] Q. Meng, B. E. Borders, C. J. Ciesewski, and M. Madden, "Closest spectral fit for removing clouds and cloud shadows," *Photogram. Eng. Remote Sens.*, vol. 75, pp. 569–576, 2009.

- [26] S. Jin, C. Homer, L. Yang, G. Xian, J. Fry, and P. Danielson, "Automated cloud and shadow detection and filling using two-date Landsat imagery in the USA," *Int. J. Remote Sens.*, vol. 34, pp. 1540–1560, Mar. 2013.
- [27] X. Li, H. Shen, L. Zhang, H. Zhang, Q. Yuan, and G. Yang, "Recovering quantitative remote sensing products contaminated by thick clouds and shadows using multitemporal dictionary learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7086–7098, Nov. 2014.
- [28] X. Li, H. Shen, L. Zhang, and H. Li, "Sparse-based reconstruction of missing information in remote sensing images from spectral/temporal complementary information," *ISPRS J. Photogramm. Remote Sens.*, vol. 106, pp. 1–15, Aug. 2015.
- [29] M. Xu, X. Jia, M. Pickering, and A. Plaza, "Cloud removal based on sparse representation via multitemporal dictionary learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2998–3006, May 2016.
- [30] X. H. Li, H. F. Shen; H. F. Li, and L. P. Zhang, "Patch matching-based multitemporal group sparse representation for the missing information reconstruction of remote-sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 8, pp. 3629–3641, Aug. 2016.
- [31] S. Malek, F. Melgani, Y. Bazi, and N. Alajlan, "Reconstructing cloud-contaminated multispectral images with contextualized autoencoder neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2270–2282, Apr. 2018.
- [32] Q. Zhang, Q. Yuan, C. Zeng, X. Li, and Y. Wei, "Missing data reconstruction in remote sensing image with a unified spatial-temporal-spectral deep convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4274–4288, Aug. 2018.
- [33] F. Chen, L. Tang, and Q. Qiu, "Exploitation of CBERS-02B as auxiliary data in recovering the Landsat7 ETM+ SLC-off image," in *Proc. Int. Conf. Geoinform.*, 2010, pp. 1–6.
- [34] C. Li, Y. Zheng, and Y. Wu, "Recovering missing pixels for landsat ETM+ SLC-off imagery using HJ-1A /1B as auxiliary data," *Int. J. Remote Sens.*, vol. 38, pp. 3430–3444, Jun. 2017.
- [35] R. Eckardt, C. Berger, C. Thiel, and C. Schmullius, "Removal of optically thick clouds from multi-spectral satellite images using multi-frequency SAR data," *Remote Sens.*, vol. 5, pp. 2973–3006, May 2013.
- [36] B. Huang, Y. Li, X. Han, Y. Cui, W. Li, and R. Li, "Cloud removal from optical satellite imagery with SAR imagery using sparse representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 1046–1050, May 2015.
- [37] Y. Li, W. Li, and C. Shen, "Removal of optically thick clouds from high-resolution satellite imagery using dictionary group learning and interdictionary nonlocal joint sparse coding," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 1870–1882, May 2017.
- [38] D. P. Roy, J. Ju, P. Lewis, C. Schaaf, F. Gao, and M. Hansen, "Multitemporal MODIS–Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data," *Remote Sens. Environ.*, vol. 112, pp. 3112–3130, Jun. 2008.
- [39] F. Gao, J. Masek, M. Schwaller, and F. Hall, "On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2207–2218, Aug. 2006.
- [40] Q. Cheng, H. Liu, H. Shen, P. Wu, and L. Zhang, "A Spatial and temporal nonlocal filter-based data fusion method," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4476–4488, May 2017.
- [41] P. Pérez, M. Gangnet, and A. Blake, "Poisson image editing," *ACM Trans. Graph.*, vol. 22, pp. 313–318, 2003.
- [42] X. Zhu, E. H. Helmer, F. Gao, D. Liu, J. Chen, and M. A. Lefsky, "A flexible spatiotemporal method for fusing satellite images with different resolutions," *Remote Sens. Environ.*, vol. 172, pp. 165–177, Jan. 2016.
- [43] I. V. Emelyanova, T. R. McVicar, T. G. Van Niel, L. T. Li, and A. I. J. M. van Dijk, "Assessing the accuracy of blending Landsat–MODIS surface reflectances in two landscapes with contrasting spatial and temporal dynamics: A framework for algorithm selection," *Remote Sens. Environ.*, vol. 133, pp. 193–209, Jun. 2013.
- [44] F. Li, D. L. B. Jupp, S. Reddy, L. Lymburner, N. Mueller, and P. Tan, "An evaluation of the use of atmospheric and BRDF correction to standardize landsat data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 3, no. 3, pp. 257–270, Sep. 2010.
- [45] Q. Wang and P. M. Atkinson, "Spatio-temporal fusion for daily Sentinel-2 images," *Remote Sens. Environ.*, vol. 204, pp. 31–42, Jan. 2018.
- [46] Q. Wang, Y. Zhang, A. O. Onojeghuo, X. Zhu, and P. M. Atkinson, "Enhancing spatio-temporal fusion of MODIS and landsat data by incorporating 250 m MODIS Data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 9, pp. 4116–4123, Sep. 2017.



Huanfeng Shen (M'10–SM'13) received the B.S. degree in surveying and mapping engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2002 and 2007, respectively.

In 2007, he joined the School of Resource and Environmental Sciences, Wuhan University, where he is currently a Luojia Distinguished Professor. He has authored more than 100 research papers. His research interests include image quality improvement, remote sensing mapping and application, data fusion and assimilation, and regional and global environmental change. He has been supported by several talent programs, such as the Youth Talent Support Program of China in 2015, China National Science Fund for Excellent Young Scholars in 2014, and the New Century Excellent Talents by the Ministry of Education of China in 2011.

Dr. Shen is currently a member of the editorial board of the *Journal of Applied Remote Sensing*.

Jingjin Wu received the B.S. degree in geographic information system from Anhui University, Hefei, China, in 2015. He is currently working toward the Ph.D. degree in cartography and geographic information engineering at the School of Resource and Environmental Sciences, Wuhan University, Wuhan, China.

His research interests include spatiotemporal data fusion and missing information reconstruction of remote sensing images.



Qing Cheng received the B.S. degree in the geographic information system and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2010 and 2015, respectively.

She is currently working with the School of Urban Design, Wuhan University, as a Lecturer. Her research interests include remote sensing data reconstruction, data fusion, and urban remote sensing.



Mahemujiang Aihemaiti received the B.S. degree in water conservancy and hydropower engineering and the Ph.D. degree in water conservancy engineering from Xinjiang Agricultural University, Ürümqi, China, in 2012 and 2015, respectively.

He is currently a Lecturer at the College of Water Conservancy and Civil Engineering, Xinjiang Agricultural University.

Chengyue Zhang received the B.S. degree in geographic information system and the M.S. degree in geomatics engineering from Wuhan University, Wuhan, China, in 2016 and 2018, respectively. He is currently working toward the Ph.D. degree at the School of Resource and Environmental Sciences, Wuhan University, Wuhan, China.

His research interests include cloud removal, data fusion, and PM2.5 reconstruction.



Zhiwei Li received the B.S. degree in geoinformation science and technology from China University of Geosciences, Wuhan, China, in 2015. He is currently working toward the Ph.D. degree with the School of Resource and Environmental Sciences, Wuhan University, Wuhan, China.

His research interests include cloud detection and removal, deep learning, and remote sensing applications. He is currently working on projects for cloud detection and removal for multisource satellite images in land resources investigation.