https://doi.org/10.1016/j.rse.2020.112001

Thick cloud removal in Landsat images based on autoregression of Landsat time-series data

Ruyin Cao[a*], Yang Chen[a], Jin Chen[b], Xiaolin Zhu[c], Miaogen Shen[d],

a School of Resources and Environment, University of Electronic Science and Technology of China, 2006 Xiyuan Avenue, West Hi-tech Zone, Chengdu, Sichuan 611731, China

b State Key Laboratory of Earth Surface Processes and Resource Ecology, Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

c Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University

d Key Laboratory of Alpine Ecology, Institute of Tibetan Plateau Research, CAS Center for Excellence in Tibetan Plateau Earth Sciences, Chinese Academy of Sciences, 16 Lincui Road, Beijing 100101, China

* Corresponding author.

E-mail addresses: cao.ruyin@uestc.edu.cn (R.Y.)

Abstract

Thick-cloud contamination causes serious missing data in Landsat images, which substantially limits applications of these images. To remove the thick clouds in Landsat data, the most popular methods employ auxiliary data such as a cloud-free image of the same area acquired on another date (referred to as the “reference image”). However, the performances of most previous methods strongly depend on the usefulness of the specific reference image, but in some cases high-quality cloud-free reference images are rarely available. In addition, some of these methods ignore the use of partially cloud-contaminated reference images, but clear pixels in these images can be very useful. To address these issues, a new cloud-removal method (**A**uto**R**egression to **R**emove **C**louds (ARRC)) has been developed in this study. The most important improvement of ARRC was that it considered the autocorrelation of Landsat time-series data and employed multi-year Landsat images including partially cloud-contaminated images in the cloud removing process. ARRC also addressed the cases that autocorrelation of Landsat time series might be adversely affected by abrupt land cover changes over multiple years. We compared ARRC with the widely used MNSPI (modified neighborhood similar pixel interpolator) method in four testing sites, including an urban area in Beijing and three croplands in the North China Plain, northeastern Vietnam, and Iowa, USA. Results based on images with simulated clouds showed that ARRC performed better than MNSPI and achieved lower RMSE values (e.g., 0.02129 vs. 0.03005, 0.03293 vs. 0.04725, 0.02740 vs. 0.03556, and 0.03303 vs.

0.03973 in the near-infrared band for the four testing sites, respectively). Besides, the experiments suggested the improved performance when clear pixels in partially cloud-contaminated images were used by ARRC. Furthermore, cloud-free images reconstructed by ARRC are visually better than those reconstructed by MNSPI, when both approaches were applied to real cloud-contaminated Landsat images.

Keywords: Cloud removal, Gap filling, Landsat data, Landsat time-series images, Partially cloud-contaminated images

1. Introduction

Landsat images are currently the most widely used remotely-sensed data sources for land-surface mapping with medium spatial resolution at global and regional scales (Chen et al., 2015; Hansen and Loveland, 2012; Wulder et al., 2019). However, since clouds cover approximately one-third of the earth's area at any given time, one great challenge for the applications of Landsat images is serious cloud contamination (Ju and Roy, 2008; Lin et al., 2014). Clouds (e.g., thick ones) can completely block the transmission of electromagnetic waves in the optical band, which leads to missing values in Landsat data. Reconstructing cloud-induced missing values is thus very important for Landsat data and has stimulated broad research interest, which has led to the development of numerous cloud-removal methods (e.g., reviewed by Shen et al., 2015).

The methods to remove thick clouds can be generally classified into four categories according to the type of auxiliary information employed. The first is spatial-based, in which clear pixels in the neighborhood of clouds are used to reconstruct a cloud pixel by using traditional spatial interpolation techniques such as nearest-neighbor or kriging interpolation (Siravenha et al., 2011; Yu et al., 2011). Spatial-based methods are suitable for small cloud regions, but perform poorly when cloud cover expands or landscape is heterogeneous. The second type of methods is based on the use of multiple data sources, in which auxiliary information can be estimated from multisource images such as synthetic aperture radar (SAR) images, which are less affected by clouds (Hoan and Tateishi, 2009; Huang et al., 2015). The use of multisource images to assist cloud removal, however, suffers from inconsistencies in the spectral and spatial resolutions of different data. The third type is temporal-based method, which employ cloud-free images of the same region acquired on other dates (referred to as “reference images”) to fill missing values in the target image if the temporal changes between them can be quantified. Some representative methods of this type include temporal filtering (Cao et al., 2018; Vuolo et al., 2017), temporal replacement, and temporal learning using sparse representation, compressed sensing and machine learning (Li et al., 2019b; Lorenzi et al., 2013; Tahsin et al., 2017). Finally, hybrid methods belong to the fourth type, in which the respective advantages of the aforementioned three types are partly integrated (Zhang et al., 2018; Zhu et al., 2012). For example, the modified neighborhood similar pixel interpolator (MNSPI) method combines both spatial-based and temporal-based estimations to fill cloud-induced missing reflectance (Zhu et al., 2012).

Among these methods, it has been recognized that the use of temporal auxiliary information is essential for better cloud removal (Li et al., 2019a; Shen et al., 2015). As a result, reference images are extensively used in previous methods. Unfortunately, these methods have not yet fully addressed the issue in the use of temporal auxiliary information. Most previous methods use one cloud-free reference images. Thus, cloud-removal performance greatly depends on the selection of the reference image. Generally, a better performance requires that the reference image should be acquired at a date as close as possible to that of the target image so that there are not substantial landscape changes (Chen et al., 2011; Zhu et al., 2012). However, a short time interval between the reference and target images cannot be satisfied in many particular applications because of the 16-d revisit period of Landsat and temporally continuous cloud contamination. Several studies attempted to address this problem by using multi-temporal reference images to remove cloud. Zeng et al. (2013) reconstructed missing pixels by first using auxiliary multi-temporal images and then used a regularization method to recover the remaining missing pixels. Lin et al. (2013) proposed to generate a synthetic reference image in which cloud-free pixels corresponding to different cloud patches can be acquired from different reference images. They found that cloud-removed images using multitemporal reference images achieved higher accuracy than those based on a single reference image. However, it may be still difficult to find a satisfactory reference image for a large cloud patch. Chen et al. (2017) proposed another way to employ multiple reference images to remove cloud. For each cloud patch, they first sorted reference image patches according to the spectral similarity between the target image and reference images and then selected the most similar three patches of reference images to estimate missing values. The weighted average of the three estimates was used to get the final estimations of this cloud patch. Chen's method selected the three reference images by calculating patch similarity, but the selection did not consider the similarity difference of each pixel within a cloud patch. In addition, some cloud pixels cannot be reconstructed if they were contaminated by cloud in the most similar three reference images, and thus these cloud pixels needed further processing. In actuality, both Lin et al. (2013) and Chen et al. (2017) aimed to select the most similar reference image from multiple reference images for each cloud patch. However, the most similar reference image does not necessarily mean that it is a better reference image for cloud removal (see our discussion). Therefore, we expect a new cloud-removal method that reduces dependence on a specific reference image and makes full use of multiple partially cloud-contaminated reference images in a simple and effective way.

In this study, a new method that we call **A**uto**R**egression to **R**emove **C**loud (ARRC) was developed. The new method uses the time series of multi-year land-surface reflectance observations and reconstructs missing data in the time series by the autoregression of Landsat time series. In some cases, the autocorrelation of Landsat time series might be adversely affected because of abrupt land cover changes. ARRC also considers and addresses these cases. ARRC employs a large number of available Landsat images, even some with clouds in them, for cloud removal in the target image. We expect the performance of ARRC to be more stable than previous temporal-based methods because the new method is less affected by the selection of specific reference images. Although the basic idea of ARRC is simple, using Landsat time-series images and their autocorrelation for cloud removal is, as far as we know, original and new. In this paper, we first demonstrate the ARRC algorithm and then compare ARRC with a widely used existing method (MNSPI) in four testing regions.

2. The ARRC Algorithm development

Fig. 1 shows a flowchart of the ARRC algorithm. In general, ARRC reconstructs the reflectance in band b for a cloud pixel (x, y) ($R(x, y, b)$) by using the weighted sum of two estimations, expressed as

$$R(x, y, b) = R_l(x, y, b) \times W_l(x, y, b) + R_s(x, y, b) \times W_s(x, y, b) \qquad (1)$$

where $R_l(x, y, b)$ indicates an estimation based on multi-year time-series reflectance images (referred to as "long-term estimation") and $R_S(x, y, b)$ indicates an estimation based on a single reference image ("short-term estimation"). $W_l(x, y, b)$ and $W_s(x, y, b)$ are the weights of the two estimations, which sums to 1. We include both long-term and short-term estimations in ARRC because $R_l(x, y, b)$ is more suitable for near-stationary time series and $R_S(x, y, b)$ accounts for cases with abrupt land cover changes. We demonstrate how to determine $R_l(x, y, b)$ and $R_S(x, y, b)$ and their weights in sections 2.1, 2.2, and 2.3, respectively.

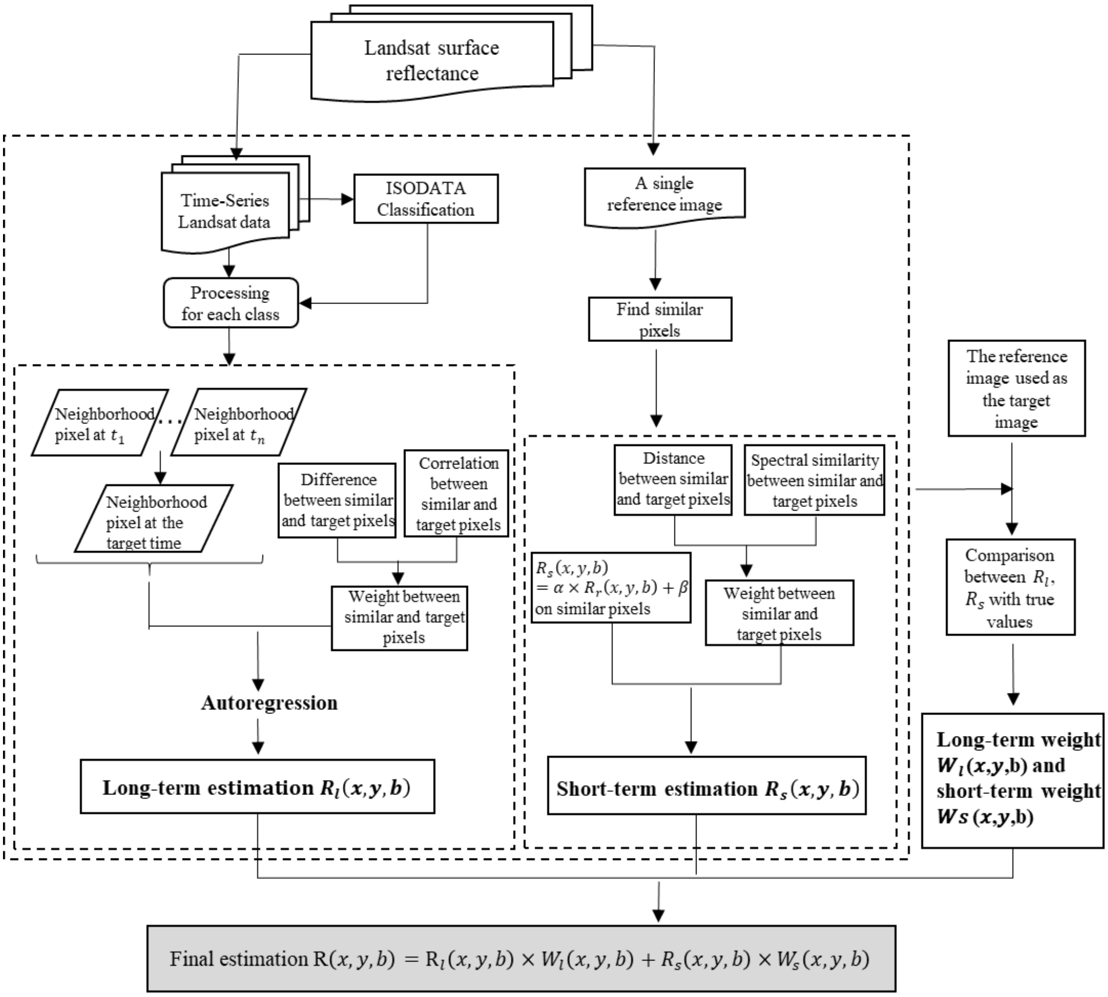

Fig. 1. Flowchart of the ARRC method.

2.1 The long-term estimation algorithm in ARRC

Based on the autocorrelation of Landsat reflectance time-series data, the long-term estimation (i.e., $R_l(x, y, b)$ in Eq. (1)) is estimated from

$$R_l(x, y, b) = \sum_{i=1}^{m} R_i(x, y, b) \times a_i(x, y, b) + a_0(x, y, b) \qquad (2)$$

where *m* is the total number of Landsat images acquired during multiple years that are cloud-free at the location of the cloud pixel (x, y). Obviously, *m* may vary for different cloud pixels. $a_i(x, y, b)$ and $a_0(x, y, b)$ are the regression parameters to be estimated. We assume that the cloud pixel (x, y) and a clear pixel in its neighborhood more likely share the same regression parameters if both pixels show similar trajectories of reflectance time-series data. Fig. 2 shows how ARRC determines the neighborhood of a cloud patch. Specifically, we first generate a local neighborhood with a window size of 30×30 pixels for each cloud pixel, and then we combine all of the local neighborhoods together. All the pixels that are in the combined neighborhood but outside the cloud patch are referred to as the neighboring pixels of this cloud patch.

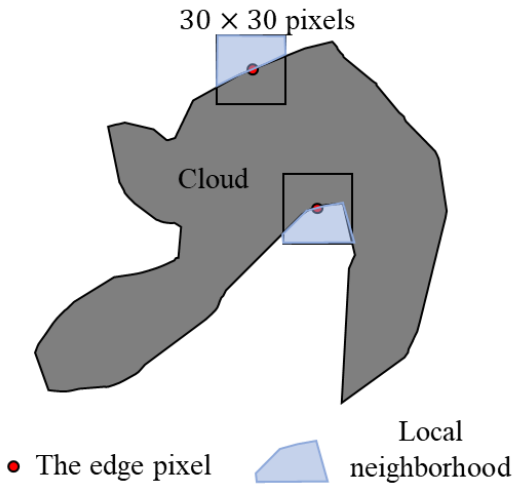

Fig. 2. Schematic showing neighborhood generation of a cloud target.

Next, we determine the similarities between the cloud pixel (x, y) and its neighboring pixels in two steps. First, we employ the unsupervised classifier ISODATA to implement clustering of all pixels in the cloud region and its neighborhood. The input data in ISODATA are new time-series images that have been formed from cloud-free images from all years (i.e. all images are cloud-free for the cloud region and its neighborhood). We set the range of the number of classes to be between 2 and 5, and the number of classes was automatically determined by ISODATA. For the number of classes for the four testing regions in this study, please refer to Table S1 in the supplementary materials. Neighboring pixels within the same class as the cloud pixel (x, y) may also have different similarities. Therefore, in the second step we further calculate the absolute difference ($DIFF$) and linear correlation (COR) of the reflectance time series between the cloud pixel (x, y) and each neighboring pixel belonging to the same class (referred to as (x_j, y_j)), as

$$DIFF_{j\_(x,y)} = \frac{1}{m}\sum_{i=1}^{m}\left|R_i(x,y,b) - R_i\left(x_j, y_j, b\right)\right| \quad (3)$$

$$COR_{j\_(x,y,b)} = Correlation\left(time\_series_{(x_j,y_j,b)}, time\_series_{(x,y,b)}\right) \quad (4)$$

Here, $time\_series_{(x,y,b)}$ consist of by all images during multiple years that are cloud-free at the location of (x, y), denoted as $[R_1(x,y,b), R_2(x,y,b), \ldots, R_m(x,y,b)]$. $time\_series_{(x_j,y_j,b)}$ are determined as the reflectance values at the corresponding m dates for (x_j, y_j), denoted as $\left[R_1\left(x_j,y_j,b\right), R_2\left(x_j,y_j,b\right), \ldots, R_m\left(x_j,y_j,b\right)\right]$. Because some reflectance values in $time\_series_{(x_j,y_j,b)}$ may be contaminated by clouds (e.g., in partially cloud-contaminated images), we perform linear interpolation on $time\_series_{(x_j,y_j,b)}$ to fill these values. By combining the absolute difference $DIFF_{j\_(x,y)}$ and the linear correlation $COR_{j\_(x,y,b)}$ together, we define the Similarity ($S_{j\_(x,y,b)}$) as

$$S_{j\_(x,y,b)} =$$

$$\frac{COR_{j\_(x,y,b)}/DIFF_{j\_(x,y)}}{\sum_{j=1}^{n}\left(COR_{j\_(x,y,b)}/DIFF_{j\_(x,y)}\right)} \tag{5}$$

where n is the total number of neighboring pixels in the same class as the cloud pixel (x, y). Using these n pixels and considering their similarities, we thus estimate the regression parameters in Eq. (2) by minimizing the following object function:

$$\underset{\boldsymbol{a_0}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{b}),\quad \boldsymbol{a_1}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{b})\quad ,\ldots,\quad \boldsymbol{a_m}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{b})}{\mathbf{arg}\,\boldsymbol{min}} \sum_{j=1}^{n} S_{j_{x,y,b}} \times [\left(\sum_{i=1}^{m} R_i\left(x_j, y_j, b\right) \times a_i(x,y,b) + a_0(x,y,b)\right) - R\left(x_j, y_j, b\right)]^2 \tag{6}$$

We solve Eq. (6) by the least squares method to estimate $a_0(x,y,b), a_1(x,y,b),\ldots,$ $a_m(x,y,b)$, which are inserted into Eq. (2) to acquire the long-term estimation $R_l(x,y,b)$.

2.2 The short-term estimation algorithm in ARRC

The long-term estimation algorithm may not be applicable to cases with abrupt land cover changes during multiple years. We thus developed a separate short-term estimation (i.e. $R_S(x,y,b)$ in Eq. 1) based on a single reference image without clouds, expressed as

$$R_S(x,y,b) = \alpha(x,y,b) \times R_r(x,y,b) + \beta(x,y,b) \tag{7}$$

where $R_r(x,y,b)$ is the reflectance in the reference image. $\alpha(x,y,b)$ and $\beta(x,y,b)$ are the slope and intercept of the linear regression, respectively, which are retrieved by minimizing the following object function:

$$\mathop{\mathbf{arg\,min}}_{\boldsymbol{\alpha}(x,y,b),\quad \boldsymbol{\beta}(x,y,b)} \sum_{is=1}^{h} w_{is} \times \left(R(x_{is}, y_{is}, b) - \alpha(x,y,b) \times R_r(x_{is}, y_{is}, b) - \beta(x,y,b)\right)^2$$

where (x_{is}, y_{is}) indicates a similar pixel in the neighborhood. $R(x_{is}, y_{is}, b)$ and $R_r(x_{is}, y_{is}, b)$ are the reflectances of the similar pixel in the target image and the reference image. Here, the "similar" pixels were determined based on the single reference image. We followed a previous study (see Eqs. 1-2 in Chen et al., 2011) and used the 20 most similar pixels (i.e. h = 20, where h is the total number of similar pixels), as suggested by Chen et al. (2011). w_{is} is the weight for each similar pixel (i.e., the higher the weight the more similar the pixel), and is determined by considering both spatial distance (D_{is}) and spectral distance (S_{is}), as follows:

$$D_{is} = \sqrt{(x_{is} - x)^2 + (y_{is} - y)^2}$$

$$S_{is} = \sqrt{\frac{\sum_{b=1}^{g} (R_r(x_{is}, y_{is}, b) - R_r(x,y,b))^2}{g}} \tag{9}$$

where g is the number of bands. Using a normalized form of D_{is} and S_{is}, we express w_{is} as

$$w_{is} = \frac{1/(nor(D_{is}) \cdot nor(S_{is}))}{\sum_{is=1}^{h} 1/(nor(D_{is}) \cdot nor(S_{is}))}, \text{ where}$$

$$nor(D_{is}) = \frac{D_{is} - \min(D_{is})}{\max(D_{is}) - \min(D_{is})} + 1, \qquad nor(S_{is}) = \frac{s_{is} - \min(s_{is})}{\max(s_{is}) - \min(s_{is})} + 1 \tag{10}$$

Therefore, by solving Eq. (8) using the least squares method we can acquire $\alpha(x,y,b)$ and $\beta(x,y,b)$, and then substitute them into Eq. (7) to obtain the short-term estimation $R_S(x,y,b)$.

The similarity between the cloud pixel and each neighboring pixel is used in both the long-term and short-term estimation algorithms. However, it should be noted that the definitions of "similarity" in the two estimations are different. In the long-term estimation, "similarity" is calculated from multi-year reflectance time-series data (Eqs. 3-5), whereas in the short-term estimation, "similarity" considers the spectral distance in multi-spectral space and the spatial distance in space (Eqs. 9-10).

2.3 Combining the long-term and short-term estimations in ARRC

The long-term ($R_l(x,y,b)$) and short-term ($R_S(x,y,b)$) estimations are combined by a weighted function (Eq. 1) in which the weights should be determined according to prediction errors. For example, a larger weight should be given to $R_l(x,y,b)$ if $R_l(x,y,b)$ has a smaller prediction error than $R_S(x,y,b)$. Unfortunately, the true reflectance values in the cloud region of the target image are unknown. Therefore, we calculate prediction errors based on the single reference image used by the short-term estimation algorithm (i.e. R_r in Eq. 7). To avoid confusion in the following demonstrations, we use the abbreviations CR_T and NCR_T to denote the cloud region and its neighborhood, respectively, in the target image, and CR_R and NCR_R to denote the corresponding areas in the reference image.

We employ the long-term estimation algorithm to predict CR_R and the prediction error for pixel (x, y) at band b (i.e., $\varepsilon_l(x,y,b)$) is as follows:

$$\varepsilon_l(x, y, b) = \left|R_{r\_l}(x, y, b) - R_r(x, y, b)\right| \quad (11)$$

where $R_{r\_l}(x, y, b)$ is the predicted reflectance from the long-term estimation algorithm for the reference image. However, the short-term estimation for location (x, y) in CR_R (referred to as $\varepsilon_s(x, y, b)$) cannot be acquired because the same location (x, y) in CR_T is cloudy. We thus estimate $\varepsilon_s(x, y, b)$ indirectly. We first randomly select half of all cloud-free pixels in NCR_T and then employ the short-term estimation algorithm to make predictions for these pixels in NCR_R. Assuming that the number of these pixels is *p*, we use the weighted average of the prediction errors to approximately represent $\varepsilon_s(x, y, b)$, expressed as

$$\varepsilon_s(x, y, b) = \sum_{inei=1}^{p} W_{inei} \times \left|R_{r\_s}(x_{inei}, y_{inei}, b) - R_r(x_{inei}, y_{inei}, b)\right| \quad (12)$$

where W_{inei} represents the weight. We also consider spatial and spectral distances and calculate W_{inei} using the same function form as Eq. (10). In actuality, Eq. (12) considers the different contributions of each neighboring pixel in the determination of $\varepsilon_s(x, y, b)$.

The weights for the long-term and short-term estimations (i.e., $W_l(x, y, b)$ and $W_s(x, y, b)$ in Eq. 1) are estimated from

$$W_l(x, y, b) = [1/\varepsilon_l(x, y, b)]/[1/\varepsilon_l(x, y, b) + 1/\varepsilon_s(x, y, b)]$$

$$W_s(x, y, b) = [1/\varepsilon_s(x, y, b)]/[1/\varepsilon_l(x, y, b) + 1/\varepsilon_s(x, y, b)] \quad (13)$$

By substituting the long-term estimation (Eq. 2), short-term estimation (Eq. 7), and their weights (Eq. 13) into Eq. (1), ARRC determines the final estimation for the missing values in the target image.

3. Data and validations

3.1 Testing regions

We tested the ARRC method in four regions. The first was the North China Plain, where double cropping is practiced in most areas. Winter wheat is normally harvested in early June and then summer maize or soybean is planted (Xiao and Tao, 2012). The second region was Thai Binh, a key paddy rice production area located in northeastern coastal Vietnam. Two paddy rice crops are grown in Thai Binh each year (i.e., mid-June to early October and mid-December to late May) (Guan et al., 2018). The third testing region was Beijing, China, where rapid urbanization has occurred during the last two decades. The fourth region was a crop rotation area in Iowa, USA, where the rotation between corn and soybean has lasted for two decades (see, for example, the rotation maps for 2001-2002 and 2011-2012 in Fig. S1 in the supplementary materials). We chose these four testing regions because they are challenging regions for the reconstruction of missing values. The cropland regions in the North China Plain and Thai Binh have at least two growing seasons in each year; thus, land-surface vegetation phenology changes quickly. In addition, Thai Binh has a typical tropical monsoon climate with serious cloud contamination especially during the rainy season (May to October). The crop rotation region in Iowa has very heterogeneous landscapes with different changes of crop types between years (Fig. S1). Beijing has experienced substantial land cover changes during the past two decades such as from vegetated surface to buildings.

3.2 Landsat data

We collected all available Landsat surface reflectance (SR) images (Tier 1) during 1990-2016 for the four testing regions from the platform of Google Earth Engine. Owing to a failure of the scan-line corrector (referred to as SLC-off), there are missing strips in Landsat 7 ETM+ images after May 2003. We excluded these SLC-off images to avoid the impact of the missing strips on cloud removal. Atmospheric corrections have been performed on these SR images (Masek et al., 2006; Vermote et al., 2016). The Fmask method were used to automatically detect cloud and cloud shadow pixels in each Landsat image (Zhu and Woodcock, 2012; Zhu et al., 2015) and these pixels were regarded as missing pixels. For more information regarding the Landsat images for the four testing regions, please refer to Table S2 in the supplementary materials.

3.3 Validation

We used three statistical indices for quantitative assessments. The first is the root mean square error (*RMSE*) which is defined as

$$\mathrm{RMSE} = \frac{1}{n}\sqrt{\sum_{i=1}^{n}[R_{predict}(x_i, y_i, b) - R_{true}(x_i, y_i, b)]^2} \qquad (14)$$

where *n* represents the total number of pixels in the cloud region. The second evaluation index is the correlation coefficient (*COR*) of the linear regression between the predicted and true reflectance values. *COR* was employed to quantify spatial consistency between the predicted and true images. The third evaluation index is the Structure SIMilarity index (*SSIM*) (Wang et al., 2004), which has been widely used to assess the overall image structure similarity between the predicted and true images (e.g., Zhao et al., 2018). *SSIM* is between 0 and 1. The more similar the predicted image is to the true image, the closer the *SSIM* value is to 1.

We performed quantitative evaluations at six bands, including the blue, green, red, near-infrared (NIR), and two short-wave infrared (SWIR) bands. The wavelengths of the two SWIR bands (expressed as SWIR1 and SWIR2) are approximately 1.55-1.75 μm (i.e., band 5 in Landsat 5 and band 6 in Landsat 8) and 2.08-2.35 μm (i.e., band 7 in Landsat 5 and band 7 in Landsat 8).

4. Experiments and Results

We compared ARRC with the previously widely used method MNSPI (Zhu et al., 2012). Obviously, the performance of both MNSPI and the short-term estimation algorithm in ARRC can be affected by the selection of the single cloud-free reference image. We thus investigated the percentages of different time intervals between the cloud image and the clear reference image in each testing region. We found that the average time intervals in the North China Plain, Beijing, Thai Binh, and Iowa were 40, 104, 101, and 125 days, respectively (Fig. S2). Therefore, in our experimental design, the time intervals cannot deviate very much from these average values, which suggests a fair comparison between the methods. We designed six groups of experiments.

4.1 Experiment I: Simulating clouds in the four testing regions

Experiment I Design: As in previous studies, quantitative assessments were performed by simulating clouds on clear Landsat images. We first randomly selected one clear image in each testing region, and then simulated a cloud region in the clear image (referred to as a “cloud-simulated image”). Here, the shapes of the simulated cloud were taken from those of true clouds. For cloud removal, the temporal reference image was the clear image with an imaging date closest to that of the cloud-simulated image. This reference image was used by MNSPI and the short-term estimation algorithm in ARRC. For North China Plain, the cloud-simulated and reference images were Landsat 5 images acquired on 26 July and 8 June, 2011 (Fig. 3A), respectively; for Thai Binh, they were Landsat 8 images on 30 December and 11 October, 2014 (Fig. 3B); for Beijing, they were Landsat 5 images on 8 August and 11 October, 2010 (Fig. 3C); and for Iowa, they were Landsat 5 images on 2 October and 16 September, 2004 (Fig. 3D).

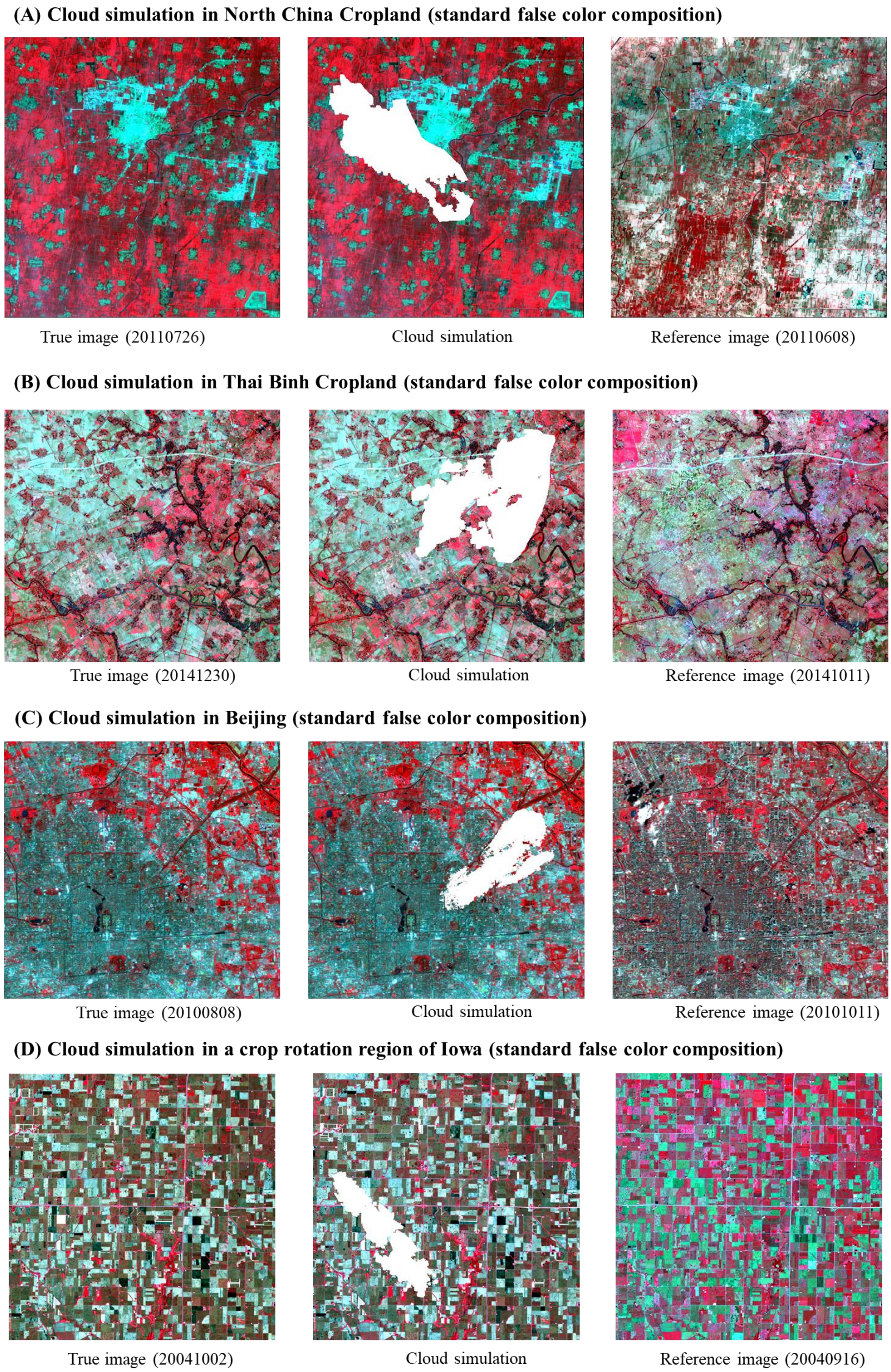

Fig. 3. The test images acquired for (A) North China cropland (Landsat 5), (B) Thai Binh cropland in Vietnam (Landsat 8), (C) Beijing (Landsat 5), and (D) a crop rotation region in Iowa, USA (Landsat 5). Each row shows from left to right the true image with the imaging date, cloud

simulation on this true image, and the image acquired on the nearest date without clouds ("reference image"). The reference image was used in the MNSPI method and for the short-term estimation algorithm in ARRC. The image sizes (pixel×pixel) for at four sites (A-D) are 643×654, 637×583, 882×822, and 800×800, respectively. The percentages of cloud pixels in the cloud-simulated images (A-D) are 10.45%, 14.99%, 10.45%, and 5.61%, respectively.

Experiment I Results: The performances of MNSPI and ARRC on cloud-simulated images in the North China Plain, Thai Binh, Beijing, and Iowa are shown in Figures. 4, 5, 6, and 7, respectively.

For the North China Plain, generally the ARRC-derived image is more similar to the true image than the MNSPI-derived images (Fig. 4). MNSPI exhibited obvious errors in some local areas (see the enlarged view panels in Fig. 4). Quantitative assessments confirmed the observations and showed that ARRC achieved lower *RMSE* and higher *COR* and *SSIM* in all six bands (Table 1). For example, the *RMSE* values for ARRC and MNSPI were 0.00503 vs. 0.00695, 0.00599 vs. 0.00864, 0.00956 vs. 0.01541, 0.03293 vs. 0.04725, 0.01114 vs. 0.01420, and 0.01410 vs. 0.02476 in the blue, green, red, NIR, SWIR1, and SWIR2 bands, respectively.

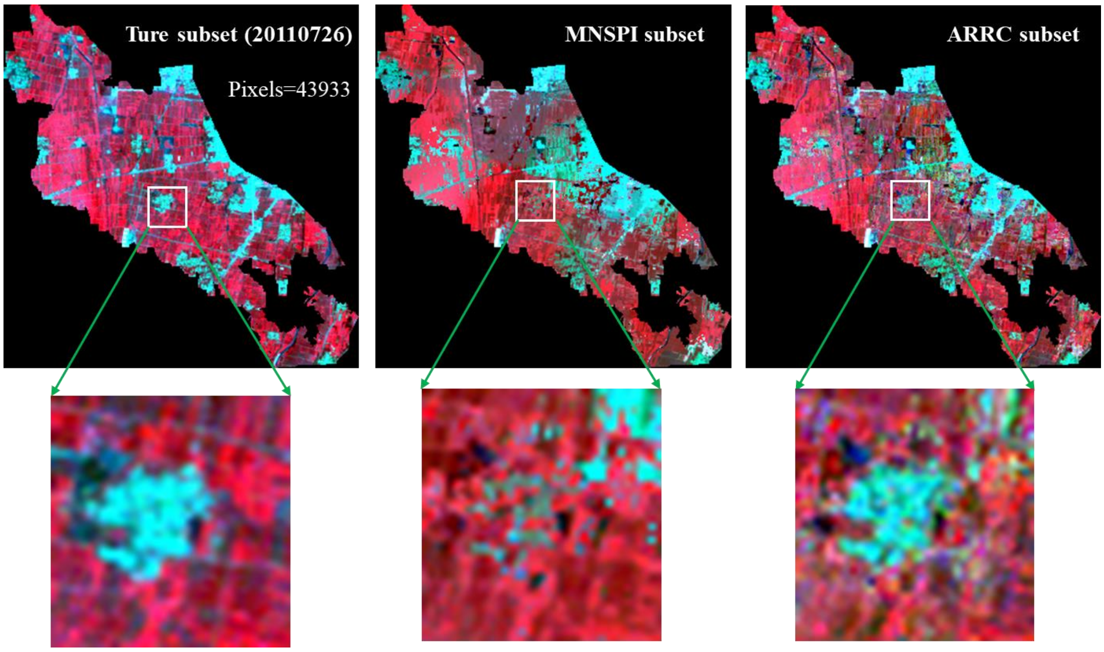

Fig. 4. Visual comparisons of the performance of MNSPI with that of ARRC for the image with cloud simulation (26 July 2011) of North China plain. These images are shown by standard false color.

Table 1. Performance of MNSPI and ARRC (both long-term and short-term estimations) for the cloud-simulated image (26 July 2011) of North China plain. The number of simulated cloud pixels is 43933. The weights to combine the long-term estimation and short-term estimation (i.e. Eq. 1) are shown in the brackets behind the RMSE values.

		MNSPI	**ARRC**	Long-term estimation (ARRC)	Short-term estimation (ARRC)
RMSE	Blue	0.00695	0.00503	0.00492 (0.62)	0.00894 (0.38)
	Green	0.00864	0.00599	0.00598 (0.49)	0.00919 (0.51)
	Red	0.01541	0.00956	0.00839 (0.60)	0.01590 (0.40)
	NIR	0.04725	0.03293	0.03323 (0.64)	0.04799 (0.36)
	SWIR1	0.01420	0.01114	0.01283 (0.62)	0.01584 (0.38)
	SWIR2	0.02476	0.01410	0.01444 (0.64)	0.02405 (0.36)
COR	Blue	0.734	0.832	0.845	0.608
	Green	0.695	0.831	0.845	0.667
	Red	0.633	0.842	0.882	0.613
	NIR	0.674	0.844	0.841	0.655
	SWIR1	0.762	0.843	0.809	0.722
	SWIR2	0.663	0.868	0.883	0.682
SSIM	Blue	0.7017	0.7833	0.7991	0.6196
	Green	0.6046	0.7346	0.7547	0.6716
	Red	0.4381	0.5332	0.5536	0.4953
	NIR	0.6799	0.8262	0.7405	0.6157

SWIR1	0.7967	0.8643	0.8337	0.7411
SWIR2	0.6564	0.8132	0.8819	0.6703

For Thai Binh, some linear features such as roads and rivers were not well reconstructed by MNSPI (see the enlarged view panels in Fig. 5). ARRC generally performed better than MNSPI with *RMSE* values of 0.00526 vs. 0.00570, 0.00758vs. 0.00856, 0.01091 vs. 0.01440, 0.02748 vs. 0.03556, 0.02656 vs. 0.03081, and 0.02036 vs. 0.02712 in the blue, green, red, NIR, SWIR1 and SWIR2 bands, respectively (Table 2).

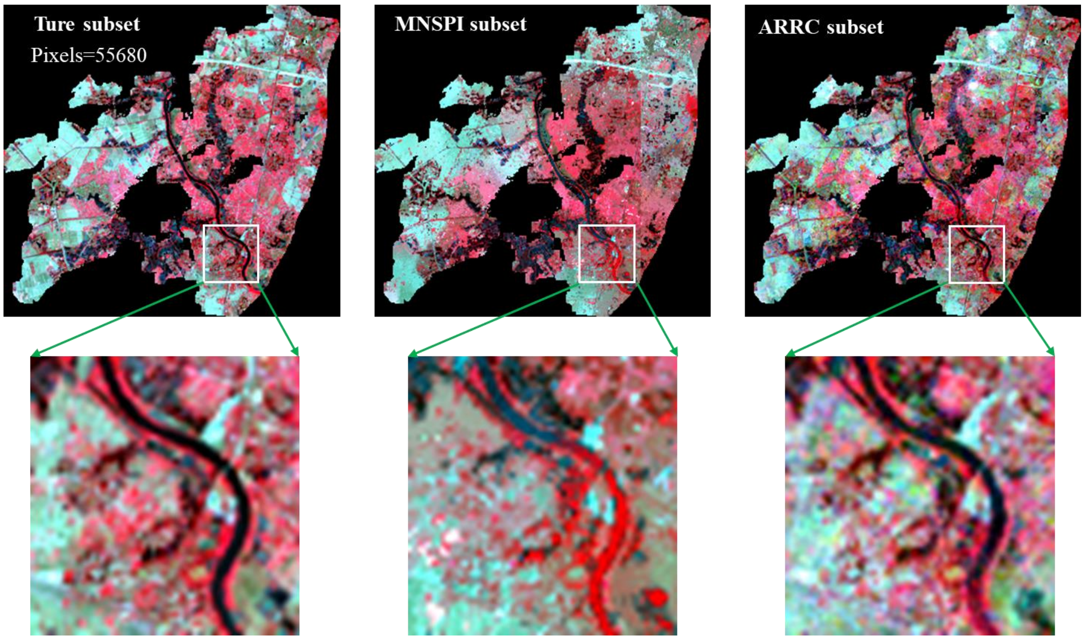

Fig. 5. Visual comparisons of the performance of MNSPI with that of ARRC for the image with cloud simulation (30 Devember 2014) of Thai Binh cropland. These images are shown by standard false color.

Table 2. Performance of MNSPI and ARRC (both long-term and short-term estimations) for the cloud-simulated image (30 Devember 2014) of Thai Binh cropland. The number of simulated cloud pixels is 55680. The weights to combine the long-term estimation and short-term estimation (i.e. Eq. 1) are shown in the brackets behind the RMSE values.

	MNSPI	**ARRC**	Long-term	Short-term

				estimation (ARRC)	estimation (ARRC)
RMSE	Blue	0.00570	0.00526	0.00564 (0.53)	0.00666 (0.47)
	Green	0.00856	0.00758	0.00836 (0.53)	0.01044 (0.47)
	Red	0.01440	0.01091	0.01266 (0.53)	0.01301 (0.47)
	NIR	0.03556	0.02748	0.03466 (0.53)	0.03243 (0.47)
	SWIR1	0.03081	0.02656	0.03191 (0.57)	0.03109 (0.43)
	SWIR2	0.02712	0.02036	0.02533 (0.57)	0.02784 (0.43)
COR	Blue	0.733	0.736	0.709	0.614
	Green	0.739	0.803	0.779	0.675
	Red	0.779	0.878	0.847	0.831
	NIR	0.739	0.849	0.794	0.785
	SWIR1	0.829	0.880	0.843	0.834
	SWIR2	0.818	0.882	0.864	0.816
SSIM	Blue	0.7092	0.6792	0.3331	0.4177
	Green	0.6907	0.7732	0.6346	0.4038
	Red	0.8083	0.8610	0.5283	0.5309
	NIR	0.8082	0.8450	0.4404	0.5976
	SWIR1	0.7890	0.8247	0.4033	0.4776
	SWIR2	0.4029	0.5288	0.3090	0.4199

For Beijing, some spatial details were also better preserved by ARRC (Fig. 6). Quantitative evaluation indices confirmed better performance of ARRC compared with that of MNSPI (Table 3). For the crop rotation region in Iowa, the reconstructed images by both ARRC and MNSPI seemed to be less satisfactory (Fig. 7). There were some differences in spatial details between the truth image and the reconstructed images, highlighting the challenge to remove cloud in these very heterogeneous areas. However, compared with MNSPI, ARRC achieved lower *RMSE* and higher *COR* and *SSIM* in all six bands (Table 4).

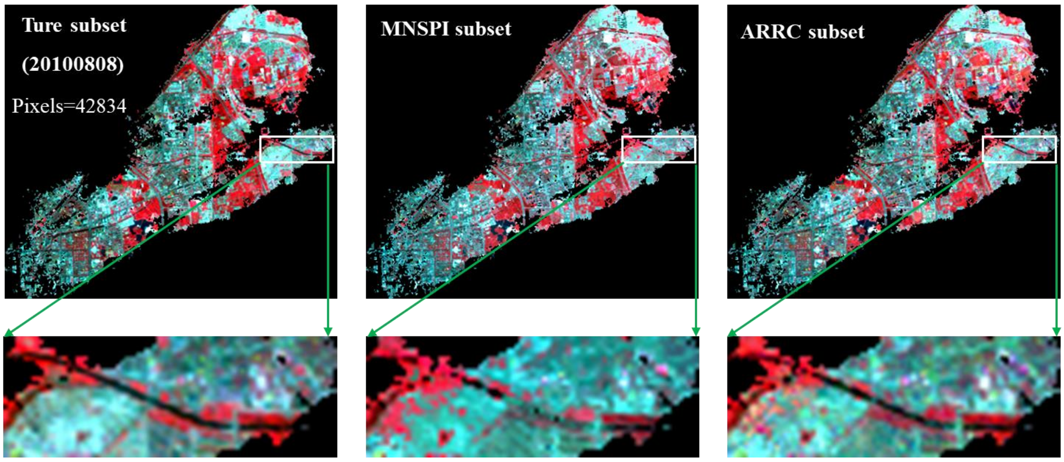

Fig. 6. Visual comparisons of the performance of MNSPI with that of ARRC for the image with cloud simulation (8 August 2010) for Beijing. These images are shown by standard false color.

Table 3. Performance of MNSPI and ARRC (both long-term and short-term estimations) for the image with cloud simulation (8 August 2010) for Beijing. The number of simulated cloud pixels is 42834. The weights to combine the long-term estimation and short-term estimation (i.e. Eq. 1) are shown in the brackets behind the RMSE values.

		MNSPI	**ARRC**	Long-term estimation (ARRC)	Short-term estimation (ARRC)
RMSE	Blue	0.01054	0.00868	0.00542 (0.58)	0.01942 (0.42)
	Green	0.01249	0.01015	0.00703 (0.59)	0.01442 (0.41)
	Red	0.01553	0.01135	0.01004 (0.59)	0.01711 (0.41)
	NIR	0.03005	0.02129	0.01986 (0.64)	0.03156 (0.36)
	SWIR1	0.02496	0.01682	0.01617 (0.60)	0.02481 (0.40)
	SWIR2	0.02913	0.02025	0.01701 (0.58)	0.02707 (0.42)
COR	Blue	0.847	0.895	0.941	0.729
	Green	0.817	0.872	0.926	0.785
	Red	0.834	0.892	0.924	0.815
	NIR	0.749	0.878	0.898	0.732
	SWIR1	0.813	0.921	0.936	0.849
	SWIR2	0.835	0.915	0.941	0.846
SSIM	Blue	0.7354	0.7497	0.7315	0.5342
	Green	0.7976	0.8719	0.8232	0.7763
	Red	0.6751	0.7814	0.7249	0.6879
	NIR	0.8051	0.9068	0.8848	0.8059
	SWIR1	0.8363	0.9141	0.9072	0.8469
	SWIR2	0.8048	0.8921	0.9210	0.8239

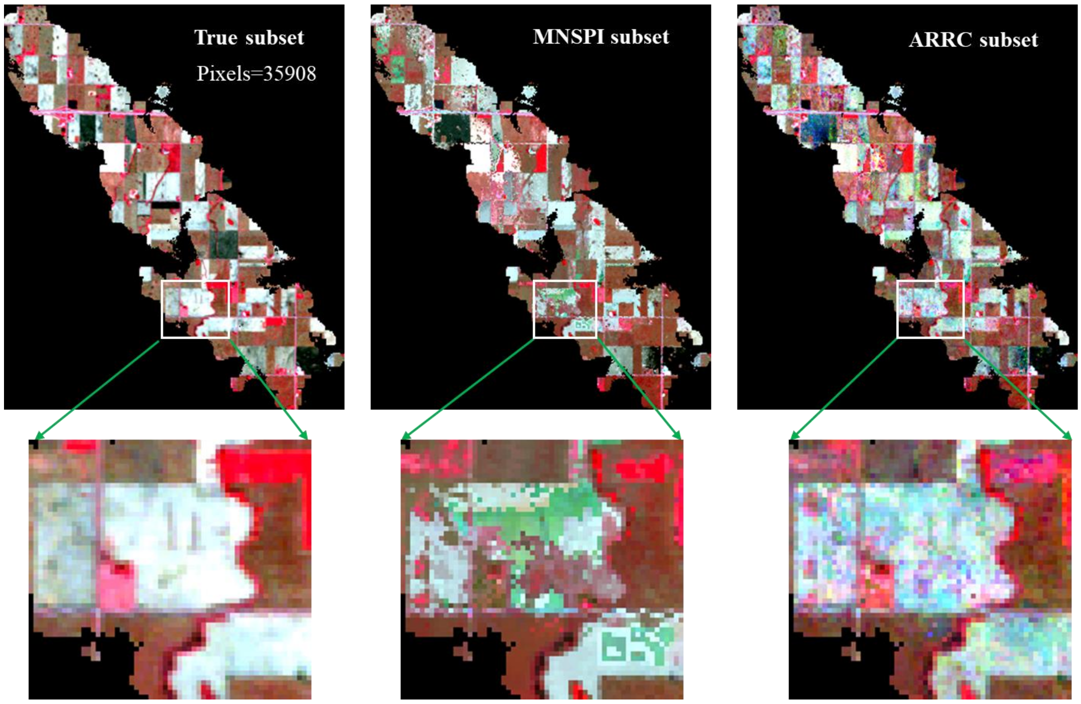

Fig. 7. Visula comparisons of the performance of MNSPI with that of ARRC for the image with cloud simulation (2 October 2004) for a crop rotation area in Iowa. These images are shown by standard false color.

Table 4. Performance of MNSPI and ARRC (both long-term and short-term estimations) for the image with cloud simulation (2 October 2004) for a crop rotation area in Iowa. The number of simulated cloud pixels is 35908. The weights to combine the long-term estimation and short-term estimation (i.e. Eq. 1) are shown in the brackets behind the RMSE values.

		MNSPI	**ARRC**	Long-term estimation (ARRC)	Short-term estimation (ARRC)
RMSE	Blue	0.00814	0.00730	0.00779 (0.60)	0.00848 (0.40)
	Green	0.01388	0.01167	0.01288 (0.62)	0.01375 (0.38)
	Red	0.02271	0.01884	0.02053 (0.59)	0.02232 (0.41)
	NIR	0.03973	0.03376	0.03389 (0.56)	0.03997 (0.44)
	SWIR1	0.02696	0.02360	0.02556 (0.61)	0.02722 (0.39)
	SWIR2	0.01937	0.01592	0.01699 (0.59)	0.01960 (0.41)
COR	Blue	0.753	0.801	0.774	0.735
	Green	0.698	0.783	0.756	0.701
	Red	0.675	0.788	0.755	0.699
	NIR	0.571	0.668	0.647	0.566
	SWIR1	0.756	0.816	0.791	0.758
	SWIR2	0.811	0.875	0.868	0.808
SSIM	Blue	0.6736	0.7431	0.7326	0.7103
	Green	0.6854	0.7321	0.7137	0.5461

Red	0.5843	0.7375	0.6555	0.6045
NIR	0.5880	0.6671	0.6467	0.5820
SWIR1	0.7348	0.7956	0.7867	0.7387
SWIR2	0.7507	0.8434	0.8553	0.7482

Because ARRC achieved its final results through a weighted combination of the long-term and short-term estimations, we investigated the two estimations in the four testing regions to gain a better understanding of the performance of ARRC (see Tables 1-4). We found that the long-term estimations had smaller *RMSE* values than the short-term estimations in all bands and testing regions. By combining the two estimations, the *RMSE* values for ARRC were further reduced in some cases, such as the NIR and two SWIR bands for the North China Plain (Table 1), and all six bands for Thai Binh (Table 2) and Iowa (Table 4). In other cases, the ARRC *RMSE* values tended to be closer to, albeit somewhat larger than, the *RMSE* values of the long-term estimations.

4.2 Experiment II: Different temporal reference images

Experiment II Design: We conducted an experiment to investigate whether and to what extent the performances of MNSPI and ARRC are affected by the selection of different temporal reference images. To illustrate this issue, this experiment was performed on the image of the North China Plain as the example. To be exact, we first simulated clouds on a clear image (6 May 2005) of the North China Plain (Fig. 8A) and then we performed cloud removal on this simulated cloud image (Fig. 8B) based on different temporal reference images. Two clear images with the closest dates to the cloud-simulated image were acquired on 22 May and 23 June 2005 and used as reference images (Figs. 8C and D). When using the 23 June 2005 reference image, we further considered two scenarios in which the 22 May 2005 image was assumed to be unavailable or partially covered by clouds (Fig. 8E).

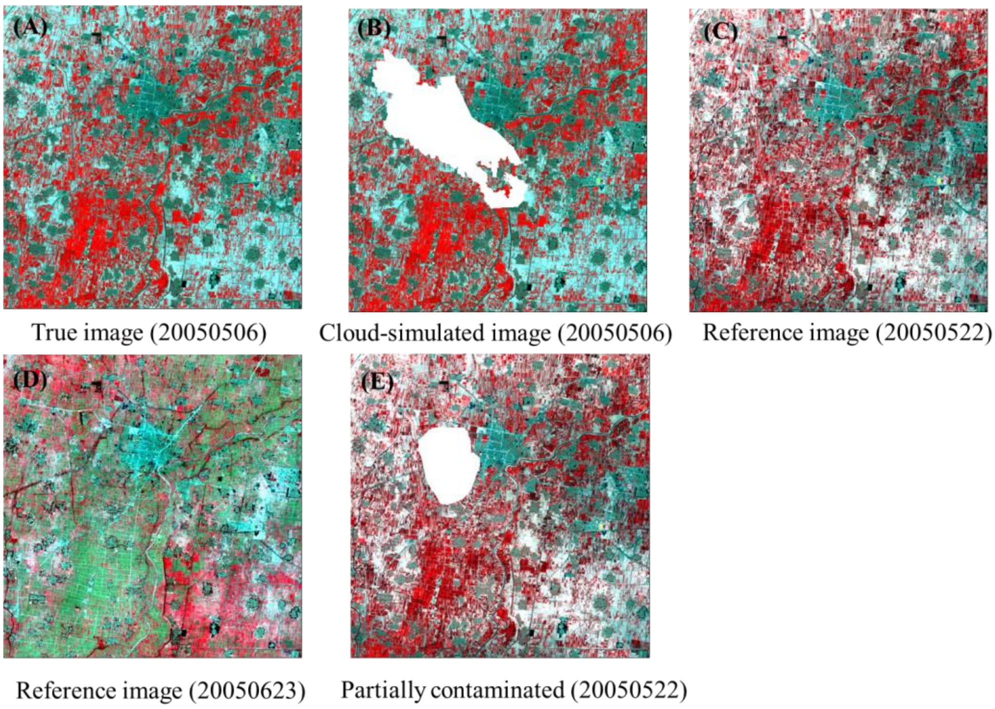

Fig. 8. (A-C) The 6 May 2005 target image, cloud simulation on this image, and the 22 May 2005 reference image. (D) The 23 June 2005 reference image. (E) Simulated partial cloud contamination on the image in (C) at 2005-05-22.

Experiment II Results: When using the reference image (22 May 2005) to remove clouds in the target image (6 May 2005) for the North China Plain, we found that both ARRC and MNSPI performed rather well, with comparable values of *RMSE* in the green and red bands. In the other bands, ARRC showed lower *RMSE* values than MNSPI (see *RMSE* in Table 5 and *COR* and *SSIM* in Table S3). The good performance of MNSPI in this case is not surprising because the time interval between the reference and target images was only 16 days. If this reference image had not been available (e.g., cloudy), the next available reference image was from one month later (23 June 2005). In the North China Plain, winter wheat is harvested in early June (Xiao and Tao, 2012); thus, the land surface in the new reference image would have been substantially different from that of the target image (see Fig. 8). As a result, changing the reference image would obviously decrease the accuracy of MNSPI (e.g., *RMSE*: 0.00968 vs. 0.01882, 0.01204 vs. 0.02757, and 0.02402 vs. 0.03725 in the green, red and NIR bands; Table 5). Changing the reference image affected the performance of ARRC to a smaller extent. In particular, ARRC could effectively make use of the partially cloud-contaminated image (22 May 2005) to further improve its performance (*RMSE* for use vs. non-use: 0.01246 vs. 0.01381, 0.01549 vs. 0.01799, 0.01811 vs. 0.02129 in the green, red and NIR bands; Table 5). This group of experiments suggests two advantages of ARRC. First, ARRC achieves more stable performance when the reference image is less satisfactory. In these cases, ARRC greatly reduces dependence on a specific reference image. Second, ARRC employs clear pixels in partially cloud-contaminated images for better cloud removal overall.

Table 5. Performance of MNSPI and ARRC (both long-term and short-term estimations) based on different temporal reference images. We considered three scenarios: (1) reference image from 22 May 2005, (2) reference image from 23 June 2005 (without the image 22 May 2005), (3) reference image from 23 June 2005 (the 22 May 2005 image is partially covered by clouds). Noted: *RMSE*

values were shown here, and *COR* and *SSIM* were shown in Table S3 due to page limitation. The number of cloud pixels in Fig. 8B is 43933. The weights to combine the long-term estimation and short-term estimation (i.e. Eq. 1) are shown in the brackets behind the RMSE values.

Cloud date: 6 May 2005		**MNSPI**	**ARRC**	Long-term estimation (ARRC)	Short-term estimation (ARRC)
Reference date 22 May 2005 (*RMSE*)	Blue	0.00833	0.00733	0.00824 (0.43)	0.00817 (0.57)
	Green	0.00968	0.00951	0.01110 (0.44)	0.01030 (0.56)
	Red	0.01204	0.01166	0.01452 (0.45)	0.01272 (0.55)
	NIR	0.02402	0.01827	0.01847 (0.50)	0.02557 (0.50)
	SWIR1	0.01711	0.01451	0.01610 (0.47)	0.01683 (0.53)
	SWIR2	0.01921	0.01614	0.01884 (0.48)	0.01917 (0.52)
Reference date 23 June 2005 (*RMSE*)	**The 22 May 2005 image is unavailable**				
	Blue	0.01853	0.01255	0.01230 (0.55)	0.01815 (0.45)
	Green	0.01882	0.01381	0.01536 (0.64)	0.01945 (0.36)
	Red	0.02757	0.01799	0.01940 (0.64)	0.02871 (0.36)
	NIR	0.03725	0.02129	0.02132 (0.60)	0.03879 (0.40)
	SWIR1	0.03762	0.02363	0.02148 (0.56)	0.03756 (0.44)
	SWIR2	0.04134	0.02716	0.02396 (0.54)	0.04336 (0.46)
Reference date 23 June 2005 (*RMSE*)	**The 22 May 2005 image is partially covered by clouds**				
	Blue	0.01853	0.01102	0.01004 (0.61)	0.01815 (0.39)
	Green	0.01882	0.01246	0.01335 (0.58)	0.01945 (0.42)
	Red	0.02757	0.01549	0.01671 (0.59)	0.02871 (0.41)
	NIR	0.03725	0.01811	0.02011 (0.60)	0.03879 (0.40)
	SWIR1	0.03762	0.02161	0.01833 (0.61)	0.03756 (0.39)
	SWIR2	0.04134	0.02437	0.02064 (0.61)	0.04336 (0.39)

4.3 Experiment III: Abrupt land cover changes

Experiment III Design: Since the long-term estimation algorithm in ARRC may be problematic for cases with abrupt land cover changes, we included the short-term estimation algorithm in ARRC to address such cases. We conducted an additional experiment to simulate abrupt land cover changes and tested the performance of ARRC under this scenario by using the same cloud-simulated images as in Fig.3. For example, to simulate abrupt land cover changes in the cloud-simulated image in Thai

Binh (acquired on 30 Dec. 2014; Fig. 9A), for each image before 2014 we replaced pixels in the cloud region (the white polygon in Fig. 9B) by pixels from another region with the same shape (the red polygon in Fig. 9B). Simulations of abrupt land cover changes in other three testing regions were shown in Fig. S3 in the supplementary materials. We compared ARRC with MNSPI to investigate whether ARRC is applicable to this challenging scenario.

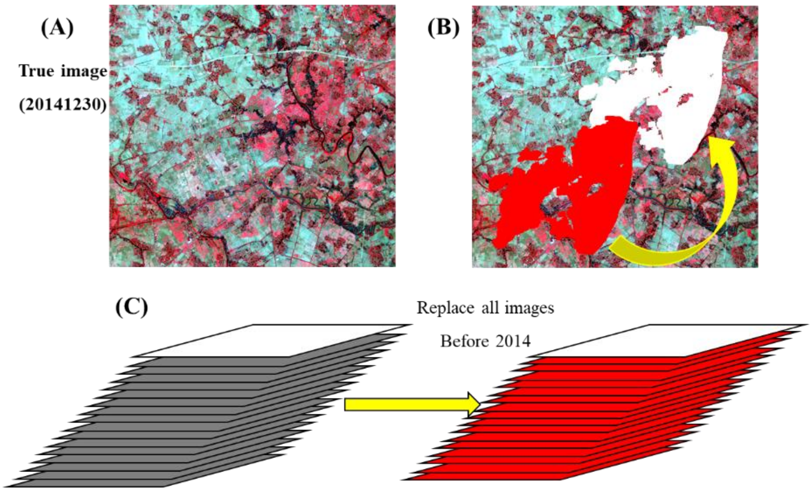

Fig. 9. (A) The true image acquired on 30 Dec. 2014 for Thai Binh cropland. (B) The cloud area (white polygon) and another area with the identical shape (red polygon). (C) We performed a simulation experiment by replacing the subset area (white polygon in panel B) by another area (red polygon in panel B) in all images before 2014.

Experiment III Results: As we expected, the long-term estimation in ARRC performed poorly under this scenario (Table 6), possible because of the destruction of the autocorrelation in the time-series images. However, benefiting from the short-term estimation, ARRC continued to achieve comparable accuracy to that of MNSPI (e.g., *RMSE*: 0.01005 vs. 0.00856, 0.01472 vs. 0.01440, and 0.03342 vs. 0.03556 in the green, red and NIR bands, respectively). Similar results were found for other testing regions (Table S4). This experiment suggests the necessity of including both long-term and short-term estimation algorithms in ARRC, which makes the new method robust for various regions, even for an area with abrupt land cover changes.

Table 6. Performance of MNSPI and ARRC (both long-term and short-term estimations) for the simulated scenario with abrupt land cover changes. Noted: the results for Thai Binh were shown here, and the results for other three testing regions were shown in Table S4 due to page limitation. The number of cloud pixels is 55680. The weights to combine the long-term estimation and short-term estimation (i.e. Eq. 1) are shown in the brackets behind the RMSE values.

Thai Binh (30 December 2014)		**MNSPI**	**ARRC**	Long-term estimation (ARRC)	Short-term estimation (ARRC)
RMSE	Blue	0.00570	0.00698	0.02437 (0.42)	0.00666 (0.58)
	Green	0.00856	0.01005	0.01823 (0.33)	0.01044 (0.67)
	Red	0.01440	0.01472	0.02692 (0.32)	0.01301 (0.68)
	NIR	0.03556	0.03342	0.05206 (0.39)	0.03243 (0.61)
	SWIR1	0.03081	0.04087	0.08361 (0.41)	0.03109 (0.59)
	SWIR2	0.02712	0.03166	0.06687 (0.41)	0.02784 (0.59)
COR	Blue	0.733	0.594	0.136	0.614
	Green	0.739	0.617	0.189	0.675
	Red	0.779	0.747	0.331	0.831
	NIR	0.739	0.765	0.457	0.785
	SWIR1	0.829	0.731	0.281	0.834
	SWIR2	0.818	0.737	0.284	0.816
SSIM	Blue	0.7092	0.5009	0.1394	0.4177
	Green	0.6907	0.4937	0.1768	0.4038
	Red	0.8083	0.6101	0.2887	0.5309
	NIR	0.8082	0.7384	0.4686	0.5976
	SWIR1	0.7890	0.5864	0.1881	0.4776
	SWIR2	0.4029	0.3865	0.1444	0.4199

4.4 Experiment IV: Performances on real cloud images

Experiment IV Design: We compared ARRC with MNSPI for cloud removal on real cloud-contaminated images in one year for each testing region. In this experiment, we first determined the years in which there are the maximum number of images and found the years were 2004, 2015, 2016, and 2002 for North China, Thai Binh, Beijing, and Iowa, respectively. We then reconstructed all cloud images if the percentage of cloud-contaminated pixels in a cloud image is less than 80%. As a result, we performed cloud removal on 4, 4, 6, and 5 cloud images in North China, Thai Binh, Beijing, and Iowa, respectively.

Experiment IV Results: Fig. 10 shows the performances of MNSPI and ARRC on one real cloud image for each testing region. The cloud-removed images reconstructed by ARRC are visually better than those reconstructed by MNSPI. Spatial details can be well restored by ARRC, even for the cases with large clouds. Similar results were observed for cloud-removed images at other dates (see Fig. S4 in the supplementary materials).

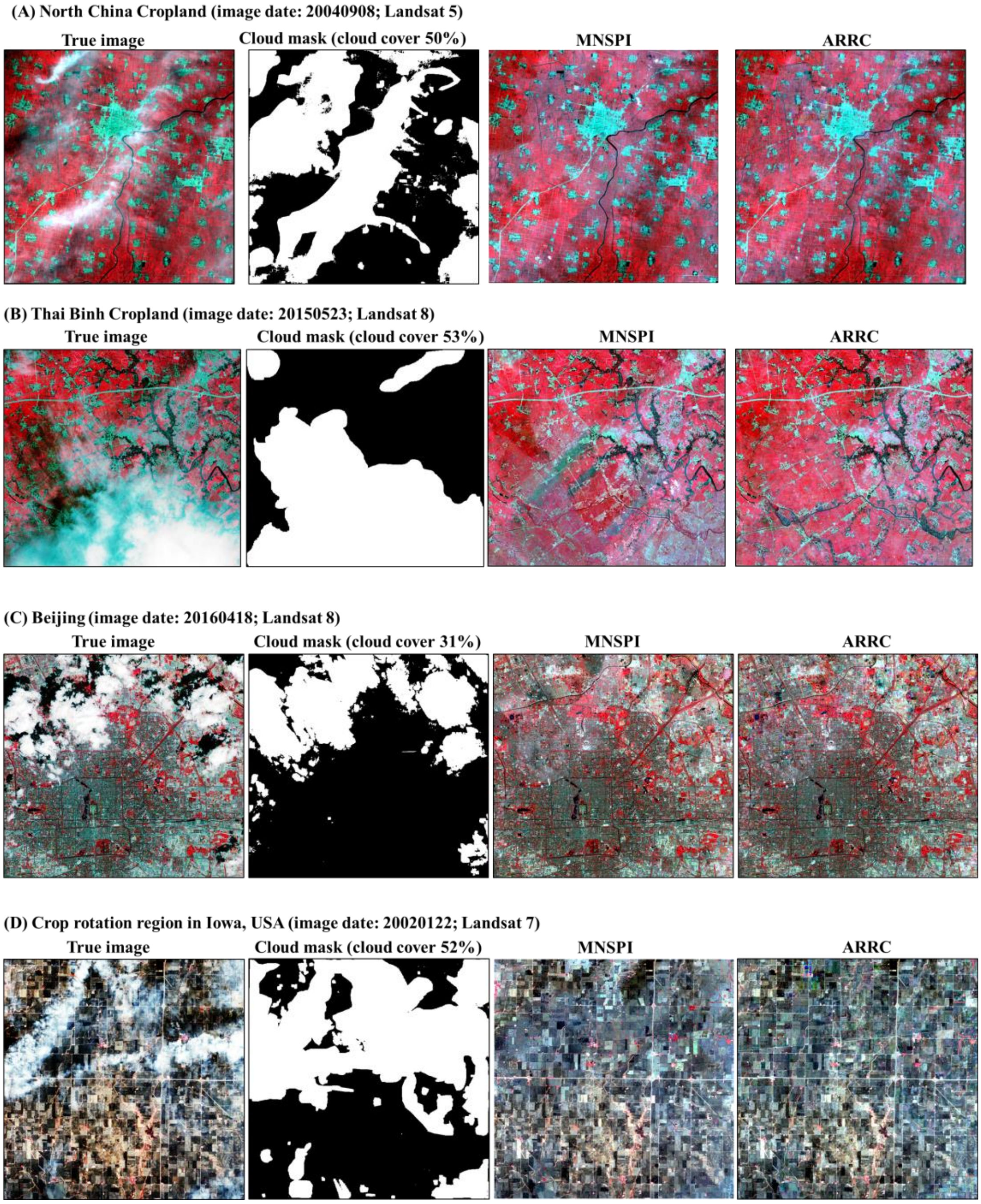

Fig. 10. (A-D) The performances of MNSPI and ARRC on one real cloud image for each testing region. Noted: for the results at all dates, please refer to Fig. S4 in the supplementary materials.

4.5 Experiment V: The length of time-series images in ARRC

Experiment V Design: The long-term estimation algorithm in ARRC employed the time-series images. In this experiment, we tested how the long-term estimation algorithm was affected by the duration of the Landsat time-series images. Using the same cloud-simulated and reference images as in Fig. 3, we investigated the performance of the long-term estimation algorithm when employing Landsat time-series images of different durations (from 1 to 5 years). These investigations can inform users the minimum number of time-series images required by ARRC.

Experiment V Results: Fig. 11 shows the relationship between the performance of the long-term estimation in ARRC and the duration of time-series images from different bands and testing regions. We found that the relationship is somewhat different in different regions. In the North China Plain, *RMSE* initially decreased rapidly as the duration of time-series images increased, and fluctuations of *RMSE* were relatively small when duration exceeded 3-4 years. In Beijing, the performance varied little when more than 2 years of data were employed. These results can be explained as follows: in the North China Plain, double cropping produces relatively complex interannual growth curves, so a minimum of 3-4 years of Landsat images are required for autocorrelation of the time-series images to be learned by ARRC. In contrast, interannual variations of reflectance from the urban surface are relatively stable, so just 1-2 years of Landsat images are required by ARRC. In Thai Binh, however, we found more dramatic fluctuations of *RMSE*, which may be due to the limited number of clear images within a year in this tropical region. In Iowa where crop rotation occurs between different years, more Landsat images were required by ARRC to achieve better performance. Overall, these investigation results suggest that in general time-series images from 3-4 years can fulfill the requirements of ARRC in most cases, which greatly reduces the data burden when applying the new method.

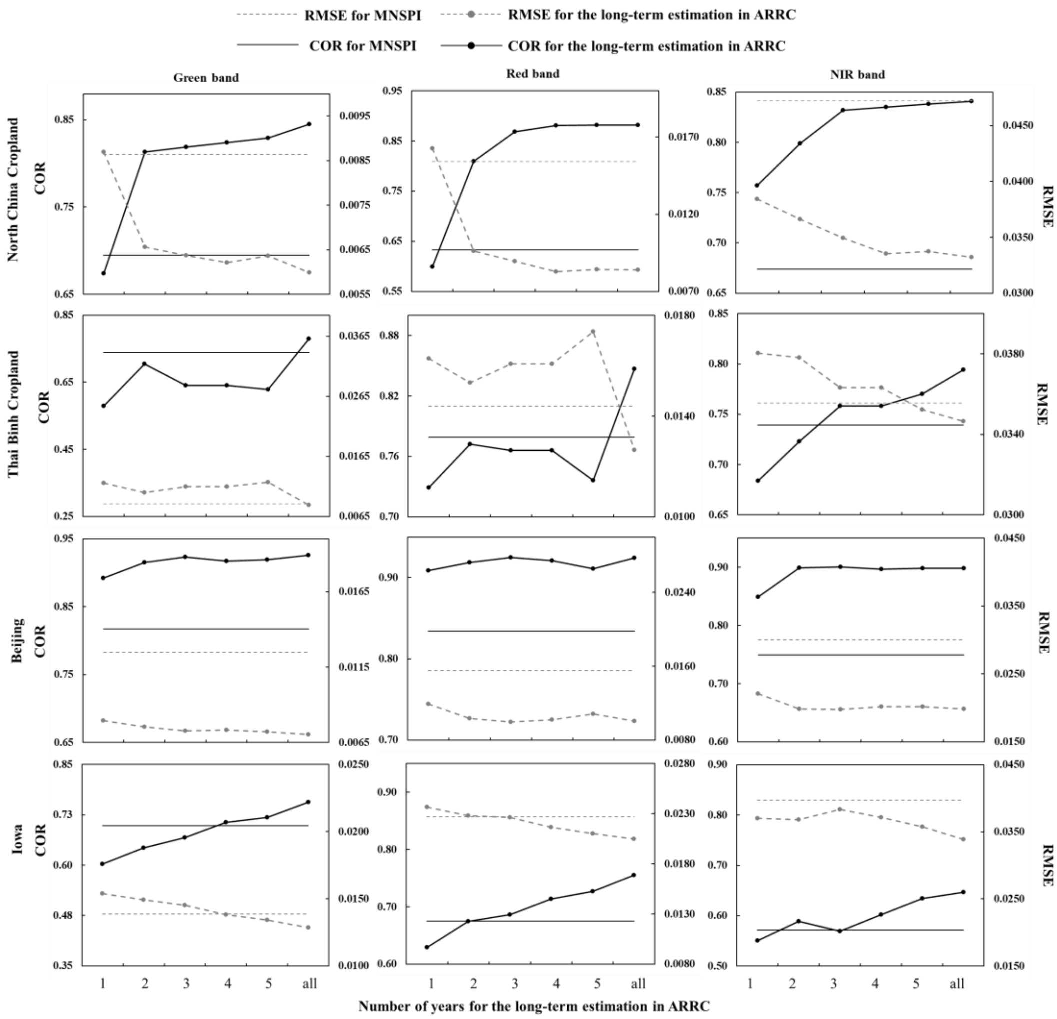

Fig. 11. Relationship between the performance of the long-term estimation in ARRC and the number of years of time-series images employed by ARRC for different bands and testing areas.

4.6 Experiment VI: Computation efficiency and algorithm scalability

Experiment VI Design: Computation efficiency and algorithm scalability should be considered for the practical applications of cloud removal. Here, computation efficiency is described as the time required for reconstructing one cloud pixel (i.e., (total time)/(total cloud pixels)). More processing time per cloud pixel indicates lower computation efficiency. Because both MNSPI and ARRC employed neighboring pixels to assist cloud removal, one concern may be the lower computation efficiency for larger clouds with more neighboring pixels. Therefore, we investigated computation efficiency for clouds with different sizes. If computation efficiency does not obviously decrease with the increase of cloud size, this can be regarded as good “computation efficiency scalability”. In addition, we considered “accuracy scalability”, which is calculated as the reconstruction accuracy for different sizes of cloud. A cloud removal method with good scalability is expected to have stable computation efficiency and accuracy for different sizes of cloud.

To address the issue mentioned above, we performed a simulation experiment as follows: we first simulated cloud with different sizes (50×50, 100×100,…, and 400×400 pixels at an interval of 50) in one clear image in each testing region (see Fig. 12 and Fig. S5). These clear images are the same as those used in Experiment I. We then investigated the changes in processing time per cloud pixel against cloud sizes. We evaluated cloud-removal accuracy by comparing the reconstructed pixels with their true values. Here, only reconstructed pixels within the minimum size of cloud (i.e., 50×50 pixels) were employed for evaluations, which guaranteed that the same cloud pixels were used for evaluations at different cloud sizes. This experiment is taken on a personal computer (CPU: Inter Core i7-8700).

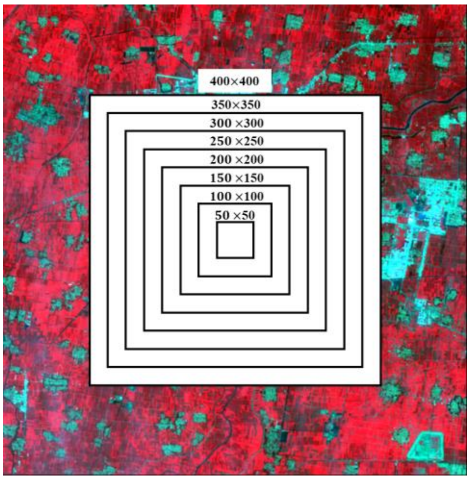

Fig. 12. The different sizes of simulated cloud in a clear image (20110726) in North China Plain. The cloud sizes are 50×50, 100×100, 150×150, 200×200, 250×250, 300×300, 350×350, and 400×400 pixels. For the simulated clouds in the other three testing regions, please refer to Fig. S5 in the supplementary materials.

Experiment VI Results: Fig. 13 shows the performances (*RMSE*) of ARRC and MNSPI at the NIR band for different cloud sizes. ARRC achieved lower RMSE values than MNSPI under all cloud sizes. *RMSE* for ARRC does not increase obviously with the increase of cloud sizes in North China Plain, Thai Binh, and Beijing. In Iowa, however, *RMSE* values for MNSPI and ARRC increase with cloud sizes when cloud is below the size of 200×200, which may be due to the very heterogeneous landscape in this crop rotation area. Similar observations were also found at other five bands (Fig. S6). These investigations suggest an acceptable "accuracy scalability" of ARRC.

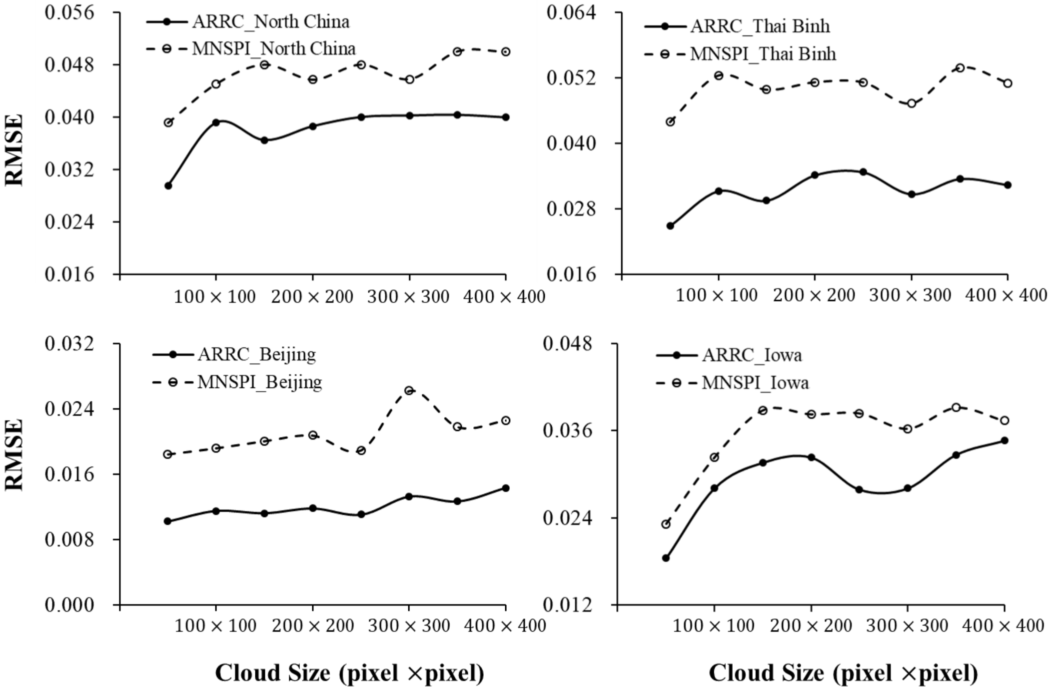

Fig. 13. The performances (RMSE) of ARRC and MNSPI at the NIR band against cloud sizes. Noted: for fair comparisons, RMSE values were calculated for only those pixels within the minimum size of cloud (i.e., 50×50 pixels).

In addition, we found that the time to reconstruct a cloud pixel is relatively stable and does not increase for larger cloud for both ARRC and MNSPI (Fig. 14). These results suggest the good scalability of both methods in terms of computation efficiency, which can be explained as: MNSPI finishes the search of similar neighboring pixels once 20 similar pixels have been found. Thus, computation efficiency does not decrease with the increase of cloud sizes although larger cloud has more neighboring pixels. However, ARRC calculated the similarity between a cloud pixel and all neighboring pixels in the same class as this cloud pixel, which is affected by cloud sizes. We thus used matrix operations to address this problem. For example, the correlation coefficients between a cloud pixel and neighboring pixels (i.e., Eq. 4) were estimated by matrix operations, which was sped up hundreds of times (see Fig. S7 in the supplementary materials). We also noted that ARRC took more processing time per cloud pixel than MNSPI (approximately 0.02s vs. 0.0006s; Fig. 14). Fortunately, ARRC has good scalability in terms of computation efficiency. Therefore, when using ARRC to process large clouds, it is possible to greatly reduce processing time by parallel computing with multiple CPU cores or Graphics Processing Unit (GPU).

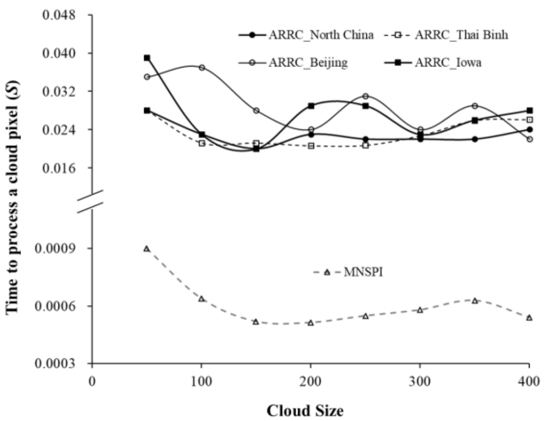

Fig. 14. The time required to reconstruct a cloud pixel for different cloud sizes. Because the time of MNSPI for different testing regions is similar, we showed the average time for MNSPI.

5. Discussion

5.1 Improvements in ARRC for cloud removal

Employing temporal auxiliary information for cloud removal is not new and has been widely adopted in previous studies (Shen et al., 2015). Most previous cloud-removal methods choose one or several reference images. In this study, we developed the new ARRC method that acquires temporal auxiliary information based on autocorrelation of long-term time-series Landsat data. The new method incorporated the following three improvements.

First, ARRC avoids the dependence on a specific reference image for those cases when the reference image is not satisfactory. Our experiments confirmed that compared with MNSPI, ARRC is less affected by the lack of a satisfactory reference image (Table 5). This result is easy to understand because the long-term estimation algorithm included in ARRC employs all available images as references and shows better performance than the short-term estimation algorithm in all four testing regions (Tables 1-4). In practice, it is very likely that some unsatisfactory reference images will be chosen because the time intervals between the reference and target images are typically several months (Fig. S2). Therefore, the employment of Landsat time-series images is important for accurate and robust cloud removal in practical applications.

The second improvement in ARRC is that it can effectively use temporal auxiliary information provided by clear pixels in partially cloud-contaminated images. Our simulation experiments showed that the cloud removal performance of ARRC on 6 May 2005 improved when the partially cloud-contaminated image from 22 May 2005 were used (Fig. 8 and Table 5). Unfortunately, most previous cloud-removal methods such as MNSPI do not use images with partial cloud cover. One solution for this problem may be to employ partially cloud-contaminated images one by one and at each time to use only clear pixels for cloud removal (e.g., Chen et al., 2017). However, such treatment is somewhat complicated especially when dealing with a large number of target images.

Third, ARRC can be widely applied to various landscapes. Logically, the most challenging scenario for the application of ARRC is changes of land cover between years because the long-term estimation algorithm in ARRC employed the time-series Landsat data. We tested ARRC in Beijing, where rapid urbanization is occurring, and in cropland regions of China, Vietnam, and Iowa, where human activities are intensive. The results showed that the long-term estimation algorithm in ARRC also performed well in these regions (see Tables 1-4). This good performance may be possible because (1) landscapes heterogeneous in space can be characterized by the use of similar pixels, and (2) gradual changes of land cover over time have less of an effect on the long-term estimation algorithm, which by using 3-4 years of data is able to capture temporal change patterns of reflectance in most cases (Fig. 11). To test this explanation, we further investigated the absolute value of the parameter a_i in Eq. 2 (i.e. the regression parameter for each image in the long-term estimation algorithm). The long-term estimation is more determined by images with larger absolute values of a_i. Taking the testing region Beijing (i.e. Fig. 3C) as the example, we calculated the absolute value of a_i averaged over each image in the long-term estimation algorithm. We found that some images during 2010-2016 have obvious larger absolute values of a_i (Fig. S8), suggesting that these images (e.g., images on 20 May, 2010 and 26 July, 2011) contributed more to the long-term estimation of the cloud-simulated image on 8 August, 2010. Those images between 1990 and 2004 have relatively small values.

Therefore, the long-term estimation algorithm can be applied to the regions with gradual changes of land cover over time.

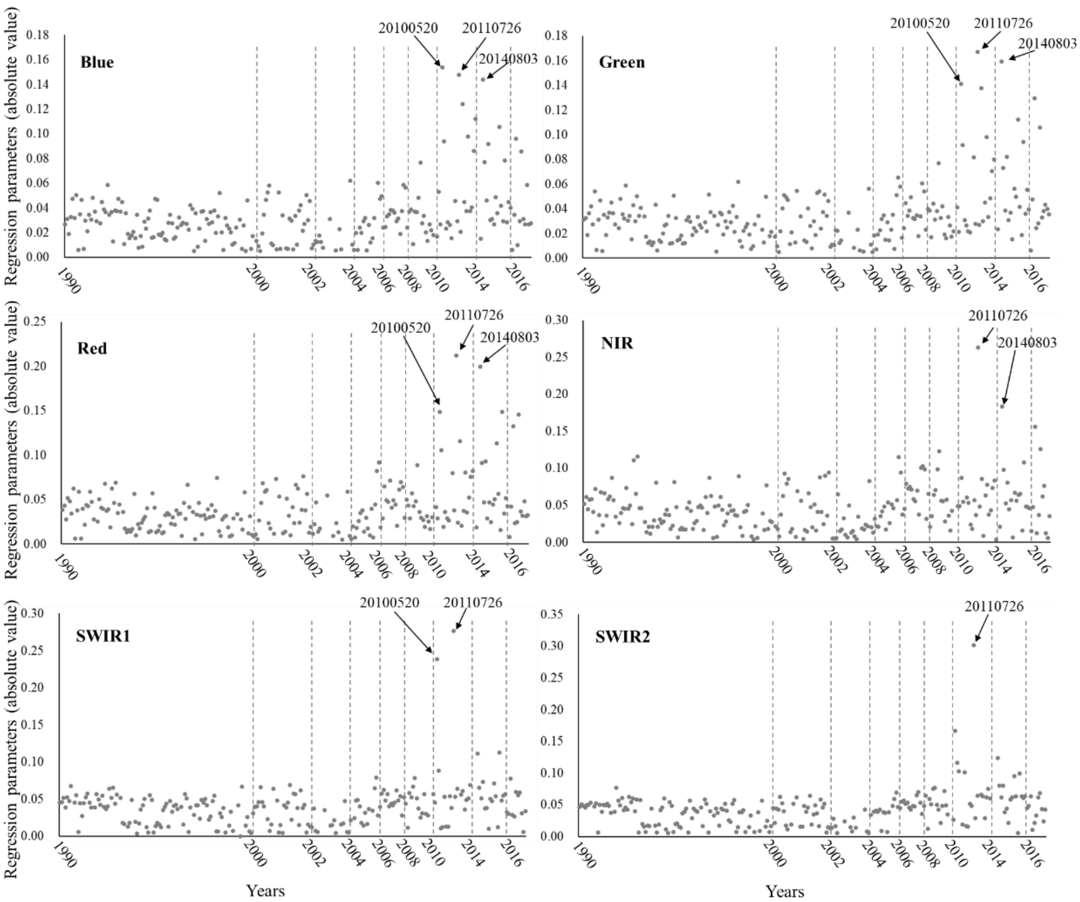

Fig. S8. The absolute value of the parameter a_i in Eq. 2 (i.e. the regression parameter for each image in the long-term estimation algorithm) for the testing region Beijing. The long-term estimation is more determined by an image if this image has a larger a_i. Noted: the absolute value of a_i is averaged over each image.

ARRC combines the long-term and short-term estimation into one framework considering the respective strengths of both components and the effectiveness of the combination method. On one hand, the short-term component is preferred if the cloud-free reference image is acquired at a date close to that of the clouded image. The experiment II suggested that the short-term estimations were better than the long-term estimations at some bands when the time interval between the cloud-free reference image and the clouded image was only 16 days (Table 5). We investigated more cases with the shortest time interval (16-d) and also found better performance of the short-term component in cases where surface change was minimal during a 16-day period (see the additional experiment in the supplementary materials). In addition, the short-term component performed better in the extreme case of abrupt land cover changes (Fig. 9 and Table 6). One the other hand, the long-term component is more appropriate for the case when a satisfactory reference image (e.g., reference image with16-d time interval) is not available. This situation is very common because the average time intervals between the reference and cloud images are typically from one to several months in many regions (Fig. S2). Our experiments confirmed that the long-term component performed better than the short-term component in these common cases (see Tables 1-4). Therefore, we included both the long-term and short-term components in ARRC to make it flexible to handle both cases, i.e. a satisfactory reference cloud-free image is available or not. We noted that combining the long-term and short-term components does not further improve the final estimates in some cases (Tables 1-4 and Table 6). However, in these cases, the performances of ARRC were more determined by the component with better performances (e.g., the long-term component in Tables 1-4 and the short-term component in Table 6 and the additional experiment in the supplementary materials). These results suggest the effectiveness of the combination method of ARRC. Because we cannot test ARRC in all scenarios, the combination of the two components can make ARRC more robust for various scenarios.

5.2 Uncertainties in ARRC

We recognize that some uncertainties regarding the application of ARRC remain. First, the neighborhood of a cloud pixel was empirically determined by using a local window with a size of 30 × 30 (Fig. 2). We also tested a larger local window (50 × 50) and found similar ARRC performance (Table S5 in the supplementary materials). This result may be because the different weights for each neighboring pixel were considered in terms of temporal consistence, spatial distance, and spectral similarity. Thus, additional pixels outside a window of 30 × 30 may provide only limited auxiliary information. On the basis of current investigation results, we recommend a window size of 30 × 30 in ARRC for the computational efficiency.

Second, a temporal reference image is necessary for MNSPI and the short-term estimation algorithm in ARRC. In our experiments, this reference image was determined to be the clear image in the same year with an imaging date closest to that of the cloud-contaminated image (called “the closest date strategy”). We also noted that some previous studies determined the reference image as the clear image that is most similar to the cloud-contaminated image (called “the most similar strategy”). Here, we used the most similar strategy to determine the reference image and performed the quantitative evaluation experiment (i.e., Experiment I) again. In specific, we followed Lin et al. (2013) to calculate the *SSIM* index between the cloud-simulated image and reference images. To accurately estimate image similarity, *SSIM* was calculated for only the cloud-free pixels in the neighborhood of the simulated cloud patch. The reference image was determined to be the one with the highest *SSIM* value (Fig. S9). The experimental results showed that ARRC also performed better than MNSPI (see Fig. S10 and Table S6 in the supplementary materials). Interestingly, we found half quantitative results of MNSPI were not improved when the reference images were determined by using the most similar strategy (comparing Table S6 with Tables 1-4), suggesting that it is still difficult to find a better reference image for cloud removal. In the future, more efforts may be necessary to quantify the relationship between the reference image and the cloud-contaminated image to improve cloud-removal performances.

Third, we used all available Landsat images to achieve the long-term estimation in ARRC. One concern may be the difference in spectra between the datasets from TM, ETM+, and OLI, which may affect the performances. To address this concern, we further investigated the performances of the long-term algorithm in ARRC by using the images from an identical sensor. For example, we used TM data to remove clouds in the TM data only. This additional experiment was performed on the same cloud-simulated images as those used in Experiment I (i.e., Fig. 3). Results showed that the long-term estimations based on an identical sensor perform worse than the estimations using all sensors in the testing regions North China and Thai Binh (Table S7). In Iowa, however, the long-term estimations using an identical sensor are better. These investigations suggest that ARRC does not necessarily to use the datasets from an identical sensor, possible because of the small difference in spectra for TM/ETM+/OLI. For the very heterogeneous areas such as crop rotation areas in Iowa, using the dataset from an identical sensor may be better choice to further improve the performances of ARRC.

Fourth, compared with some previous methods that use only one reference image, ARRC requires time to collect and preprocess at least 3-4 years of Landsat images. However, it is worth noting that more and more applications are based on a large amount of Landsat images (e.g., multi-year data) since Landsat data became freely available in 2008. ARRC may be preferred when dealing with many cloud-contaminated images because of its simple operation and robust performance.

Last, in this study we tested ARRC in four challenging landscapes (North China Plain, cropland in Vietnam, the city of Beijing, and a crop rotation area in Iowa, USA). More tests in various regions will be necessary in future studies.

6. Conclusions

We developed a new method (called ARRC) to remove thick cloud in Landsat images. The new method reconstructs missing values by the weighted sum of two estimations. One estimation is based on autocorrelation of Landsat time-series data and employs multi-year Landsat images for cloud removal (referred to the long-term estimation), and the other estimation is to remove cloud based on a single reference image (referred to the short-term estimation). We evaluated ARRC in four testing regions by using both simulated and real cloud images, including an urban area in Beijing and three croplands in the North China Plain, northeastern Vietnam, and Iowa, USA. We found that the new method performed better than the widely used MNSPI method. ARRC achieved lower RMSE values (e.g., ARRC vs. MNSPI: 0.02129 vs. 0.03005, 0.03293 vs. 0.04725, 0.02740 vs. 0.03556, and 0.03303 vs. 0.03973 in the NIR band for the four testing regions, respectively) and higher SSIM values (e.g., ARRC vs. MNSPI: 0.8262 vs. 0.6799, 0.8450 vs. 0.8082, 0.9068 vs. 0.8051, and 0.6671 vs. 0.5880 in the NIR band for the four testing regions, respectively). By applying both methods to real cloud-contaminated Landsat images, cloud-removed images generated by ARRC are visually better than those generated by MNSPI. Our experiments suggested three advantages of ARRC. First, ARRC eliminates dependence on a specific reference image for those cases when the reference image is less satisfactory. Second, ARRC uses temporal auxiliary information provided by clear pixels in partially cloud-contaminated images in a simple and effective way. Third, the performances of ARRC are robust for various landscapes and the image with large clouds.

Acknowledgments

We thank Dr. Zhu to provide the source code to run MNSPI. This work was funded by the 2nd Scientific Expedition to the Qinghai-Tibet Plateau (No. 2019QZKK0106 and 2019QZKK0307), the National Natural Science Foundation of China (Grants No. 41601381, 41571103, and 41861134038), top-Notch Young Talents Program of China (to Shen), the Key Research Program of Frontier Sciences of the Chinese Academy of Sciences (Grant No. QYZDB-SSW-DQC025).

References

Cao, R.Y., Chen, Y., Shen, M.G., et al., 2018. A simple method to improve the quality of NDVI time-series data by integrating spatiotemporal information with the Savitzky-Golay filter. Remote Sensing of Environment, 217, 244-257.

Chen, B., Huang, B., Chen, L.F., Xu, B., 2017. Spatially and Temporally Weighted Regression: A Novel Method to Produce Continuous Cloud-Free Landsat Imagery. IEEE Transactions on Geoscience and Remote Sensing, 55, 27-37.

Chen, J., Chen, J., Liao, A., et al., 2015. Global land cover mapping at 30 m resolution: a POK-based operational approach. ISPRS Journal of Photogrammetry and Remote Sensing, 103, 7-27.

Chen, J., Zhu, X.L., Vogelmann, J.E., et al., 2011. A simple and effective method for filling gaps in Landsat ETM+ SLC-off images. Remote Sensing of Environment, 115, 1053-1064.

Guan, K.Y., Hien, N.T., Li, Z., et al., 2018. Measuring Rice Yield from Space: The Case of Thai Binh Province, Vietnam. Social Science Electronic Publishing, DOI: 10.2139/ssrn.3188560.

Hansen, M.C., Loveland, T.R., 2012. A review of large area monitoring of land cover change using Landsat data. Remote Sensing of Environment, 122, 66-74.

Hoan, N.T., Tateishi, R., 2009. Cloud removal of optical image using SAR data for ALOS application: experimenting on simulated ALOS data. Journal of The Remote Sensing Society of Japan, 29(2), 410-417.

Huang, B., Li, Y., Han, X., et al., 2015. Cloud removal from optical satellite imagery with SAR imagery using sparse representation. IEEE Geoscience and Remote Sensing Letters, 12(5), 1046-1050.

Ju, J.C., Roy, D.P., 2008. The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally. Remote Sensing of Environment, 112, 1196-1211.

Li, X.H., Wang, L.Y., Cheng, Q., et al., 2019a. Cloud removal in remote sensing images using nonnegative matrix factorization and error correction. ISPRS Journal of Photogrammetry and Remote Sensing, 148, 103-113.

Li, Z.W., Shen, H.F., Cheng, Q., et al., 2019b. Thick Cloud Removal in High-Resolution Satellite Images Using Stepwise Radiometric Adjustment and Residual Correction. Remote Sensing, 11(16):1925.

Lin, C.H., Tsai, P.H., Lai, K.H., Chen, J.Y., 2013. Cloud Removal From Multitemporal Satellite Images Using Information Cloning. IEEE Transactions on Geoscience and Remote Sensing, 51(1), 232-241.

Lin, C.H., Lai, K.H., Chen, Z.B., et al., 2014. Patch-Based information reconstruction of cloud-contaminated multitemporal images. IEEE Transactions on Geoscience and Remote Sensing, 52(1), 163-174.

Lorenzi, L., Melgani, F., Mercier, G., 2013. Missing-area reconstruction in multispectral images under a compressive sensing perspective. IEEE Transactions on Geoscience and Remote Sensing, 51(7), 3998-4008.

Malek, S., Melgani, F., Bazi, Y., et al., 2018. Reconstructing Cloud-Contaminated Multispectral Images with Contextualized Autoencoder Neural Networks. IEEE Transactions on Geoscience and Remote Sensing, 56, 2270-2282.

Masek, J.G., Vermote, E.F., Saleous, N.E., et al., 2006. A Landsat surface reflectance dataset for North America 1990-2000. IEEE Geoscience and Remote Sensing Letters, 3, 568-572.

Shen, H.F., Li, X.H., Cheng, Q., et al., 2015. Missing information reconstruction of remote sensing data: a technical review. IEEE Geoscience and Remote Sensing Magazine, 3(3), 61–85.

Siravenha, A.C., Sousa, D., Bispo, A., et al., 2011. Evaluating inpainting methods to the satellite images clouds and shadows removing. In: International Conference on Signal Processing. Image Processing and Pattern Recognition (SPPR), Jeju Island, Korea, pp. 56–65.

Tahsin, S., Medeiros, C.S., Hooshyar, M., et al., 2017. Optical cloud pixel recovery via machine learning. Remote Sensing, 9, 1-19.

Vermote, E., Justice, C., Claverie, M., et al., 2016. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. Remote Sensing of Environment, 185, 46-56.

Vuolo, F., Ng, W.T., Atzberger, C., 2017. Smoothing and gap-flling of high resolution multi-spectral time series: Example of Landsat data. International Journal of Applied Earth Observation and Geoinformation, 57, 202-213.

Wang, Z., Bovik, A. C., Sheikh, H. R., et al., 2004. Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4), 600-612.

Wulder, M.A., Loveland, T.R., Roy, D.P., et al., 2019. Current status of Landsat program, science, and applications. Remote Sensing of Environment, 225, 127-147.

Xiao, D.P., and Tao, F.L., 2012. Impact of climate change in 1981-2009 on winter wheat phenology in the North China Plain. Chinese Journal of Eco-Agriculture, 20, 1539-1545.

Yu, C., Chen, L., Su, L., et al., 2011. Kriging interpolation method and its application in retrieval of MODIS aerosol optical depth. In: The 19th International Conference on Geoinformatics (ICG), Shanghai, China, pp. 1-6.

Zeng, C., Shen, H.F., Zhang, L.P., 2013. Recovering missing pixels for Landsat ETM+ SLC-off imagery using multi-temporal regression analysis and a regularization method. Remote Sensing of Environment, 131, 182-194.

Zhang, Q., Yuan, Q., Zeng, C., Li, X., Wei, Y., 2018. Missing data reconstruction in remote sensing image with a unifed spatial-temporal-spectral deep convolutional neural network. IEEE Transactions on Geoscience and Remote Sensing, 56 (8), 4274-4288.

Zhao, Y.Q., Huang, B., Song, H.H., 2018. A robust adaptive spatial and temporal image fusion model for complex land surface changes. Remote Sensing of Environment, 208, 42-62.

Zhu, X.L., Gao, F., Liu, D., et al., 2012. A Modifed Neighborhood Similar Pixel Interpolator Approach for Removing Thick Clouds in Landsat Images. IEEE Geoscience and Remote Sensing Letters, 9(3), 521-525.

Zhu, Z., Woodcock, C.E. 2012. Object-based cloud and cloud shadow detection in Landsat imagery. Remote Sensing of Environment, 118, 83-94.

Zhu, Z., Wang, S., Woodcock, C.E., 2015. Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4-7, 8, and Sentinel 2 images. Remote Sensing of Environment, 159, 269-277.