Science of Remote Sensing 5 (2022) 100046

Contents lists available at ScienceDirect

Science of Remote Sensing

journal homepage: www.sciencedirect.com/journal/science-of-remote-sensing

Fusing or filling: Which strategy can better reconstruct high-quality fine-resolution satellite time series?

Hongtao Shu [a,b,c], Shiguo Jiang [d], Xiaolin Zhu [c,*], Shuai Xu [c], Xiaoyue Tan [c], Jiaqi Tian [c], Yi Nam Xu [c], Jin Chen [a,b]

[a] State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing, 100875, China
[b] Beijing Engineering Research Center for Global Land Remote Sensing Products, 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, Hong Kong, China
[d] Department of Geography and Planning, University at Albany, State University of New York, Albany, NY, USA

ARTICLE INFO

Keywords:
Satellite time series
Gap filling
Spatiotemporal fusion
Image reconstruction
Cloud
Landsat
MODIS

ABSTRACT

High-quality fine-resolution satellite time series data are important for monitoring land surface dynamics in heterogeneous areas. However, the quality of raw satellite time series is affected by clouds and the revisit frequency. Currently, there are two major strategies to reconstruct high-quality fine-resolution time series: the interpolation of the missing pixels using auxiliary data from the same satellite (known as filling) and the fusion of fine-resolution and coarse-resolution images (known as fusing). These two strategies use different principles and input data to reach the same goal, but which one is superior in different scenarios is not known. Therefore, this study fills this research gap by comparing two representative methods from filling and fusing: the Neighborhood Similar Pixel Interpolator (NSPI) for filling and the Flexible Spatiotemporal DAta Fusion (FSDAF) for fusing. The potential factors affecting the accuracy of the two methods were investigated using two simulated experiments. The results show that (1) the accuracy of both methods decreases with the time interval between the image to be reconstructed and the auxiliary image; (2) NSPI is generally better than FSDAF for reconstructing images with small cloud patches but this superiority is insignificant in homogeneous areas; (3) the accuracy of NSPI significantly decreases with cloud size, and NSPI is worse than FSDAF for reconstructing images with large clouds; and (4) the performance of FSDAF is significantly affected by the scale difference between the fine- and coarse-resolution images, especially for heterogeneous areas. The findings of this study can help users select the appropriate method to reconstruct satellite time series for their specific applications.

1. Introduction

High-quality fine-resolution satellite time series data are essential to monitor and predict land surface variations, and are widely applied in land cover change detection, vegetation growth modeling, and ecosystem dynamics monitoring (Hansen et al., 2008; Johnson et al., 2016; Masek et al., 2008; Shen et al., 2011; Townshend et al., 1991; Zhu and Liu, 2014; Zhu and Woodcock, 2014). However, most satellite images acquired by a single sensor cannot concurrently have high temporal and spatial resolution. For example, the fine spatial resolution sensor (e.g., 30 m Landsat) can recognize heterogeneous spatial details but is difficult to capture abrupt land surface changes due to the low revisit frequency (e.g., 16-days for Landsat). Furthermore, due to cloud contaminations, the temporal resolution of the available observations is generally coarser than the satellite revisit frequency, resulting in large uncertainties in collecting land surface information (Ju and Roy, 2008). Therefore, the reconstruction of high-quality fine-resolution time series data is fundamental to extracting accurate and timely spatial and temporal information on the land surface (Li et al., 2017).

Previous studies have proposed two main strategies for remote sensing time series reconstruction: image filling and image fusing. These two strategies achieve image reconstruction in different ways. The filling strategy focuses on filling the contaminated pixels on the fine-resolution clouded image with fine-resolution cloud-free auxiliary images in the time series (Fig. 1 (a)), while the fusing strategy produces the whole fine-resolution image by blending fine- and coarse-resolution

* Corresponding author. Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China.
E-mail address: xiaolin.zhu@polyu.edu.hk (X. Zhu).

https://doi.org/10.1016/j.srs.2022.100046
Received 27 December 2021; Received in revised form 3 March 2022; Accepted 29 March 2022
Available online 4 April 2022

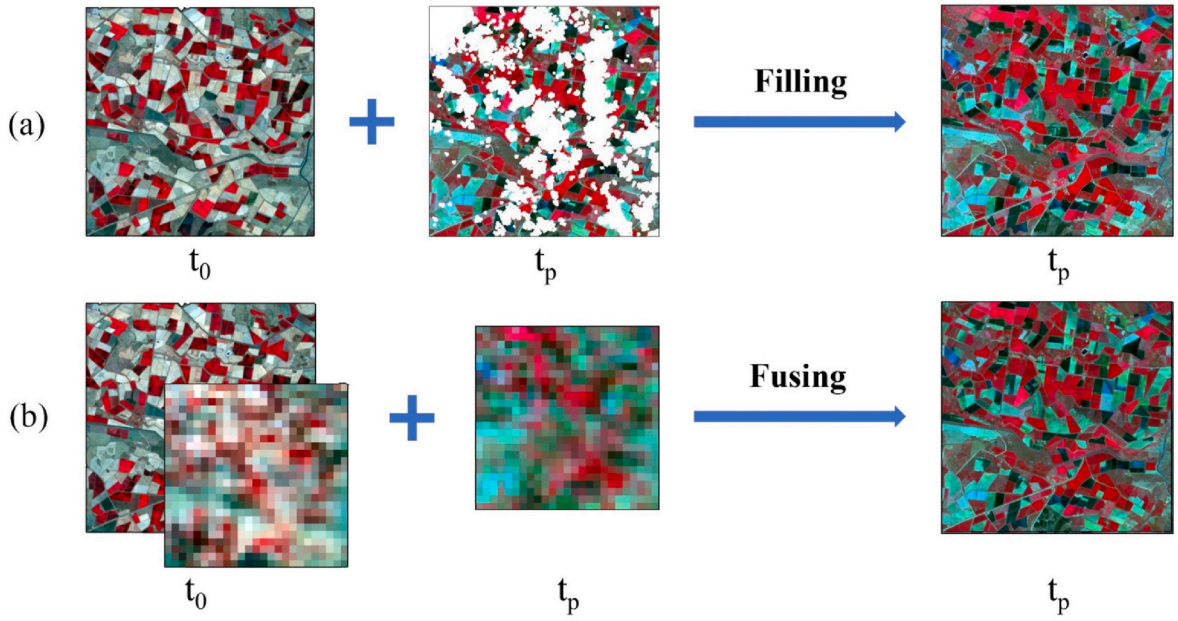

Fig. 1. Conceptual diagram of the two strategies to reconstruct a fine-resolution satellite image in the time series: (a) filling and (b) fusing.

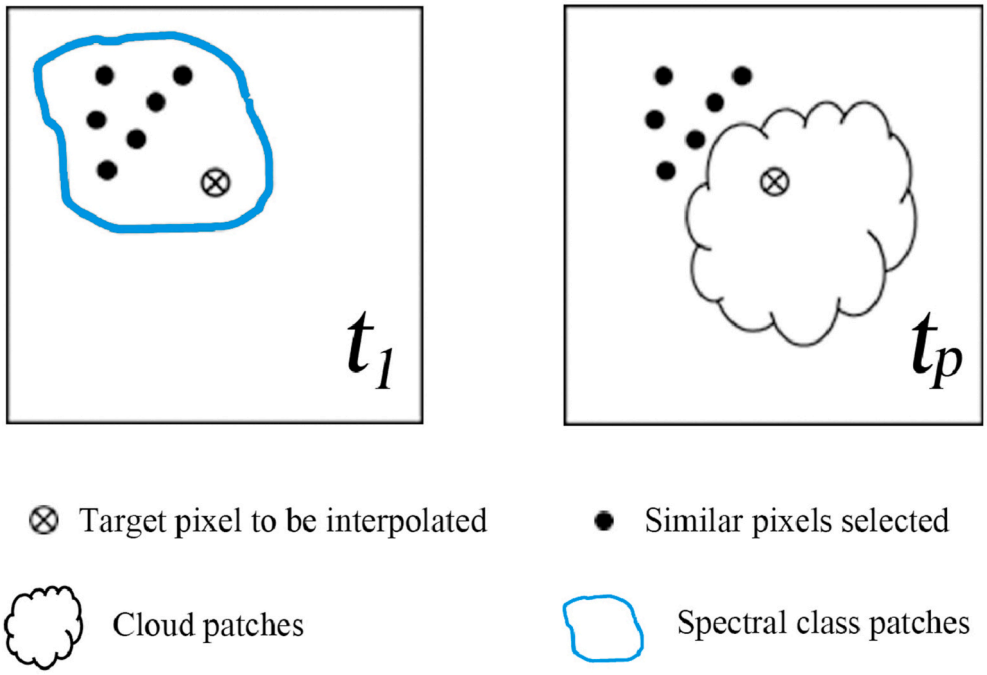

Fig. 2. Schematic diagram of input data in NSPI: a cloud-free auxiliary image acquired at t_1 and the clouded image acquired at t_p.

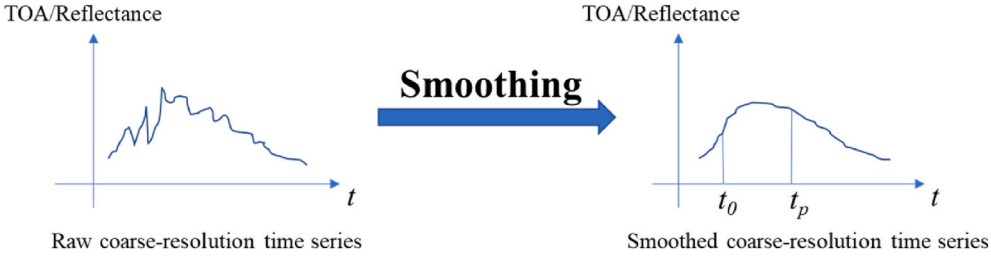

Fig. 3. Schematic diagram of smoothing coarse-resolution time series to get the coarse-resolution cloud-free images for the FSDAF fusion method.

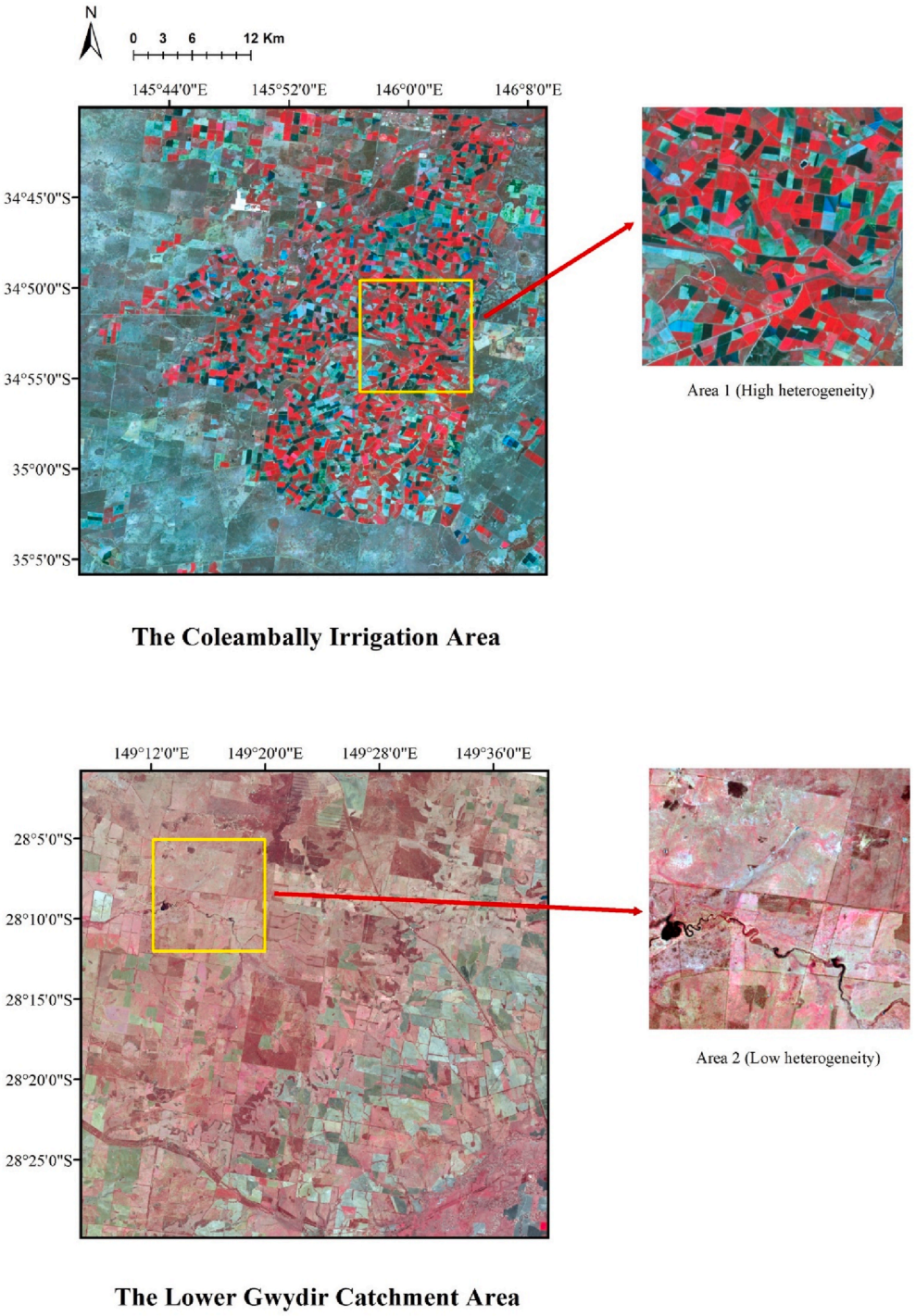

Fig. 4. The CIA and LGWY Area in NIR–red–green composites: Area 1 is a heterogeneous landscape while Area 2 is a homogeneous landscape. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

satellite images (Fig. 1 (b)).

Image filling was proposed to handle the missing pixels of the Landsat 7 ETM+ image (named as SLC-off image) due to the failure of the scan-line corrector (SLC) in 2003. The missing values were filled for the first time using the localized linear histogram matching (LLHM) method by incorporating the scanned data of the SLC-off image and another temporally close SLC-on image (USGS, 2004). With the development of the filling strategy, image filling is extended for cloud removal. Based on different perspectives of information involved, image filling strategies can be classified into three categories: spatial-based, temporal-based and hybrid methods. Spatial-based filling locates the neighboring pixels similar to the target pixel and uses spatial interpolation for image inpainting (Cheng et al., 2014), while the temporal-based algorithm relies on one or more clear temporally close images to fill the missing value (Maxwell et al., 2007; Roy et al., 2008). Hybrid filling restores the value of contaminated pixels by combining spatial information from clear pixels near the target pixel and temporal information from clear observations obtained close to the contaminated image (Chen et al., 2011; Wang et al., 2021a, 2021b; Weiss et al., 2014; Zhu et al., 2012a, 2012b). The Neighborhood Similar Pixel Interpolator (NSPI) is the first hybrid method to get a robust prediction of contaminated pixels in multi-temporal satellite images. It has been successfully

Table 1
Date information for the cloud-free Landsat-MODIS pairs (LMPs) for the two study sites. For CIA DSSD means the number of days since Oct 01, 2001, and for LGWY DSSD represents the number of days since Apr 01, 2004.

Image #	CIA		LGWY	
	DSSD	DATE	DSSD	DATE
1	017	Oct 17, 2001	16	Apr 16, 2004
2	033	Nov 02, 2001	32	May 02, 2004
3	040	Nov 09, 2001	96	Jul 05, 2004
4	056	Nov 25, 2001	128	Aug 06, 2004
5	065	Dec 04, 2001	144	Aug 22, 2004
6	097	Jan 05, 2002	208	Oct 25, 2004
7	104	Jan 12, 2002	240	Nov 26, 2004
8	136	Feb 13, 2002	256	Dec 12, 2004
9	145	Feb 22, 2002	272	Dec 28, 2004
10	168	Mar 17, 2002	288	Jan 13, 2005
11	184	Apr 02, 2002	320	Feb 14, 2005
12	193	Apr 11, 2002	336	Mar 02, 2005
13	200	Apr 18, 2002	368	Apr 03, 2005
14	209	Apr 27, 2002		
15	216	May 04, 2002		

used to remove clouds and fill the un-scanned gaps in Landsat-7 ETM+ images to reconstruct high-quality Landsat time series (Chen et al., 2011; Zhu et al., 2012a, 2012b). Besides the above methods, Zhu et al. (2015) proposed a time-series modeling method: using clear observations to fit time series models for each Landsat pixel and detect abrupt surface change, they generate synthetic Landsat images at any given time. For the scenario of thick and frequent cloud coverage, the reliability of the filling strategy is questionable due to the lack of available observations.

Different from image filling which uses data from the single sensor, image fusing aims to produce fine spatial and temporal resolution images from two sensors, one with high revisit frequency but low spatial resolution (e.g., MODIS) and the other with high spatial resolution but low revisit frequency (e.g., Landsat). Previous studies categorized the spatiotemporal fusing strategy into five types: unmixing-based, weight function-based, learning-based, Bayesian-based, and hybrid methods (Zhu et al., 2018a). For convenience, we refer to the pixels in images with low spatial resolution as "coarse pixels" and those with high spatial resolution as "fine pixels". The unmixing-based fusing methods use a linear unmixing process to unmix coarse pixels on the prediction date and then obtain the predicted image with a high spatial resolution (Rao et al., 2015; Wu et al., 2012; Zhukov et al., 1999; Zurita-Milla et al., 2008). The weight function-based methods such as the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) (Gao et al., 2006) and the Enhanced STARFM (ESTARFM) (Zhu et al., 2010) assume that all the coarse pixels represent homogeneous land cover and that reflectance changes between fine- and coarse-resolution images are consistent or proportional so that the changes derived from coarse-resolution images can be added to the fine pixels with a weighted function. The learning-based methods establish the relationship between coarse-fine image pairs using machine learning algorithms and apply this relationship to predict the fine-resolution image on a given prediction date (Huang and Song, 2012; Liu et al., 2016). The Bayesian-based methods consider spatiotemporal fusion as a Maximum A Posterior (MAP) problem and model the relationships between the observed and predicted images (Liao et al., 2016; Shen et al., 2016). Lastly, the hybrid methods combine two or more methods mentioned above to improve the effectiveness of spatiotemporal fusing, e.g., the Flexible Spatiotemporal DAta Fusion, FSDAF (Zhu et al., 2016) and its derivations: the Improved FSDAF (Liu et al., 2019), the Enhanced FSDAF (Shi et al., 2019), SFSDAF (Li et al., 2020) and FSDAF 2.0 (Guo et al., 2020).

Although filling and fusing are both proposed to reconstruct remote sensing time series, their principles and data inputs are different, implying a difference in the scope of application of the two strategies. However, the knowledge about the selection of strategies for real-world applications is still limited. It is necessary to recognize the potential influential factors of each strategy before the investigation of the

Table 2
Main influential factors of NSPI and FSDAF (✓ means influence exists, × means no impact).

Influential factors	NSPI	FSDAF
Time interval between the auxiliary images and the target image to be reconstructed	✓	✓
Spatial heterogeneity	✓	✓
Cloud size	✓	×
Scale ratio between fine- and coarse-resolution images	×	✓

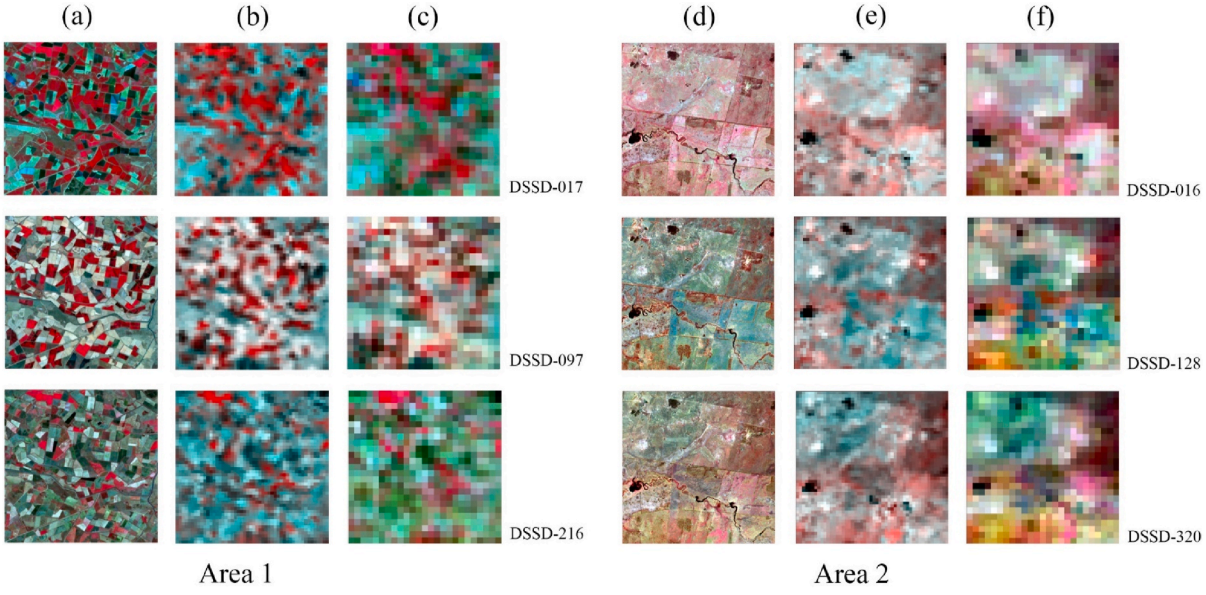

Fig. 5. Landsat (30 m) and MODIS (240 m/480 m) preprocessed images on different dates in Area 1 and Area 2 in the same spectral stretch, (a) and (d) are Landsat images, (b) and (e) are MODIS-240 m images, (c) and (f) are MODIS-480 m images. (a), (c), (d) and (f) are in NIR-red-green composites, (b) and (e) are in NIR-red-red composites. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

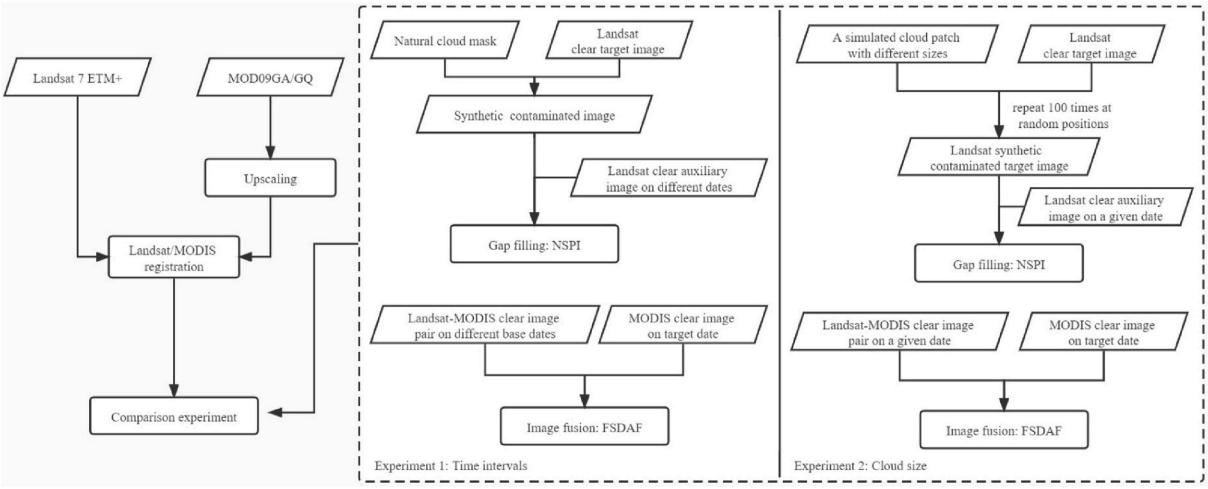

Fig. 6. Flowchart of the comparison experiment of NSPI and FSDAF with different indicators.

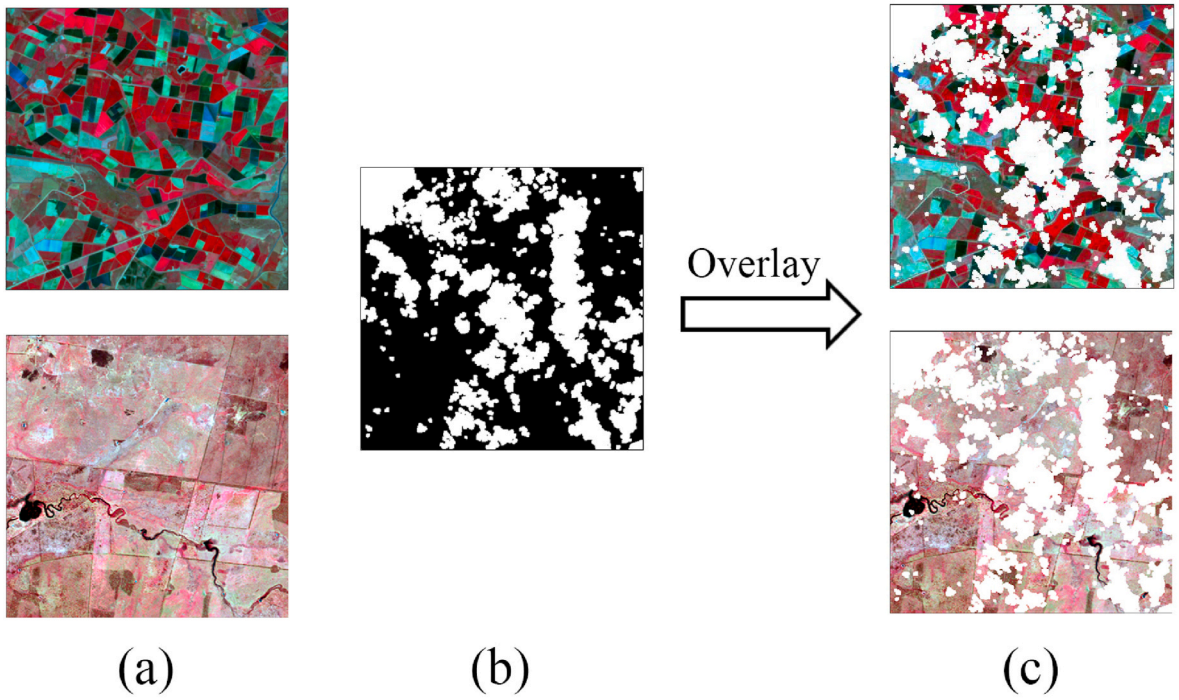

Fig. 7. Schematic diagram of generating the clouded target image. (a) Clear Landsat image on the first date in Area 1 and Area 2; (b) Natural cloud mask (cloud coverage: 37.19%); (c) Synthetic contaminated target image on the first date in Area 1 and Area 2.

capacity of each strategy for image reconstruction under different scenarios. For example, spatial heterogeneity is one of the influencing factors for both filling and fusing. It affects the searching of similar pixels in filling and the relationship modeling between coarse-fine image pairs in fusing. In addition, quantification of the sensitivity of each strategy to the influential factors is crucial for the selection of method when the input data has diverse quality (e.g., different cloud coverage).

To deepen our understanding of the superiority of filling and fusing, this study assessed the performance of both strategies in different scenarios. Two typical methods, NSPI representing the filling strategy and FSDAF denoting the fusing strategy, were selected for the performance comparison. NSPI and FSDAF are two classic methods widely tested in the literature. Three properties make them especially appropriate for validating our ideas: (1) they have simple principle and are milestones in the field of image reconstruction; (2) they are the base models for many new methods developed recently; and (3) they are reliable and robust for various sites and have been widely employed by practical studies. The objectives of this study are to determine: 1) what are the potential factors influencing the performance of image filling and fusing, respectively and 2) how the accuracy of reconstructed images changes with the influential factors. The results of this study provide guidelines on the selection of strategies for image reconstruction to support real-world applications.

2. Brief overview of NSPI and FSDAF

2.1. NSPI

NSPI is a hybrid method that integrates "spatial-filling" and

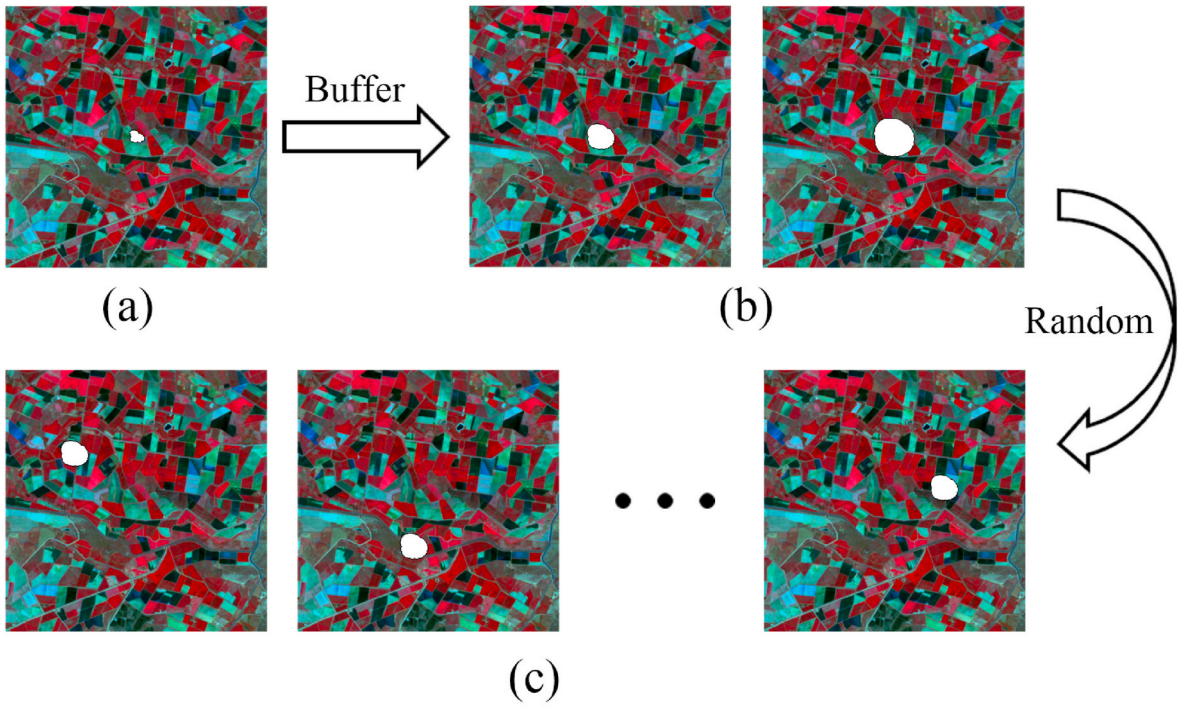

Fig. 8. Schematic diagram of buffering and simulating cloud patches. (a) One initial small cloud patch placed at the center; (b) Different sizes of cloud mask after buffering; (c) Different positions of cloud mask in each cloud size from 100 times of simulation.

"temporal-filling" into the robust prediction of contaminated pixels in fine-resolution images. It has been successfully used to remove clouds and fill the un-scanned gaps in Landsat-7 images to reconstruct high-quality Landsat time series and widely used in practical studies (Chen et al., 2011; Wang et al., 2020). NSPI needs at least one cloud-free auxiliary image acquired from the same satellite as the clouded image as input (Fig. 2).

NSPI has two assumptions: (1) neighboring similar pixels have a high degree of spectral similarity; and (2) this spectral similarity can be kept in the time series. It includes four major steps. First, it selects a certain number (e.g., 20) of pixels that are spectrally similar to the target contaminated pixel under clouds or within the gaps using the auxiliary image (named as similar pixels, see Fig. 2). The similarity between these selected pixels and the target pixel can be quantified by RMSE:

$$RMSE_i = \sqrt{\frac{\sum_{i=1}^{n}(F(x_i,y_i,t_1)-F(x,y,t_1))^2}{n}} \quad (1)$$

where $F(x,y,t_1)$ is the target pixel value at location (x, y) at t_1, and $F(x_i, y_i, t_1)$ is the ith similar pixel value at location (x_i, y_i) at t_1. Smaller RMSE means higher similarity.

Second, NSPI predicts the target pixel at t_p by a weighted average of pixel values of similar pixels in the clouded image at t_p:

$$F_1(x,y,t_p) = \sum_{i=1}^{n} W_i \cdot F(x_i,y_i,t_p) \quad (2)$$

where $F(x_i,y_i,t_p)$ is the ith similar pixel value at t_p. The prediction $F_1(x,y,t_p)$ is named as the spatial prediction because it is based on spatial dependence characteristics.

Third, NSPI predicts the target pixel at t_p using temporal change information from similar pixels. This prediction is named as the temporal prediction $F_2(x,y,t_p)$, which is obtained by a weighted average of the temporal change provided by similar pixels:

$$F_2(x,y,t_p) = F(x,y,t_1) + \sum_{i=1}^{n} W_i \cdot \left(F(x_i,y_i,t_p) - F(x_i,y_i,t_1)\right) \quad (3)$$

Last, NSPI combines spatial prediction and temporal prediction by introducing weight to obtain the final prediction of the target pixel at t_p:

$$F(x,y,t_p) = T_1 \cdot F_1(x,y,t_p) + T_2 \cdot F_2(x,y,t_p) \quad (4)$$

where T_1 and T_2 are the weights considering the reliability of spatial and temporal prediction.

2.2. FSDAF

FSDAF is the first spatiotemporal fusion method specifically designed for solving the challenge of abrupt land cover changes (e.g., flooding, wildfire) recorded in the input data and has been widely used in fusing remote sensing images (Zhang et al., 2017, 2022; Zhao et al., 2021). FSDAF assumed that (1) same-class pixels have same temporal changes; (2) land cover changes can be downscaled by spatial interpolation if it is observable at coarse images; and (3) temporal changes of similar pixels have high spatial dependence. It requires only a pair of fine- and coarse-resolution cloud-free images acquired on the same day (e.g., t_0) and a coarse-resolution cloud-free image on any other day (e.g., t_p) to predict the fine-resolution image, so given cloud-free coarse time series, fine-resolution time series can be constructed. However, the coarse-resolution time series is affected by clouds, but this can be mitigated by the smoothing techniques, as shown in Fig. 3. The steps of FSDAF are summarized as follows:

First, it predicts the fine-resolution images at t_p (the temporal prediction $F_{t_p}^{temp}$) by unmixing the temporal changes derived from the coarse-resolution images from t_0 to t_p, and then calculate the residual of temporal prediction as:

$$R(x,y) = \Delta C(x,y) - \frac{1}{m}\left[\sum_{i=1}^{m} F_{t_p}{}^{temp}(x_i,y_i) - \sum_{i=1}^{m} F_{t_0}(x_i,y_i)\right] \quad (5)$$

where $R(x,y)$ is the residual of a coarse pixel at location (x, y), $\Delta C(x,y)$ represents the real temporal change derived from the coarse pixel, m means the number of fine pixels (x_i, y_i) within the coarse pixel (x, y), F_{t_0} is the value of fine pixels at t_0. The residual is mainly caused by abrupt

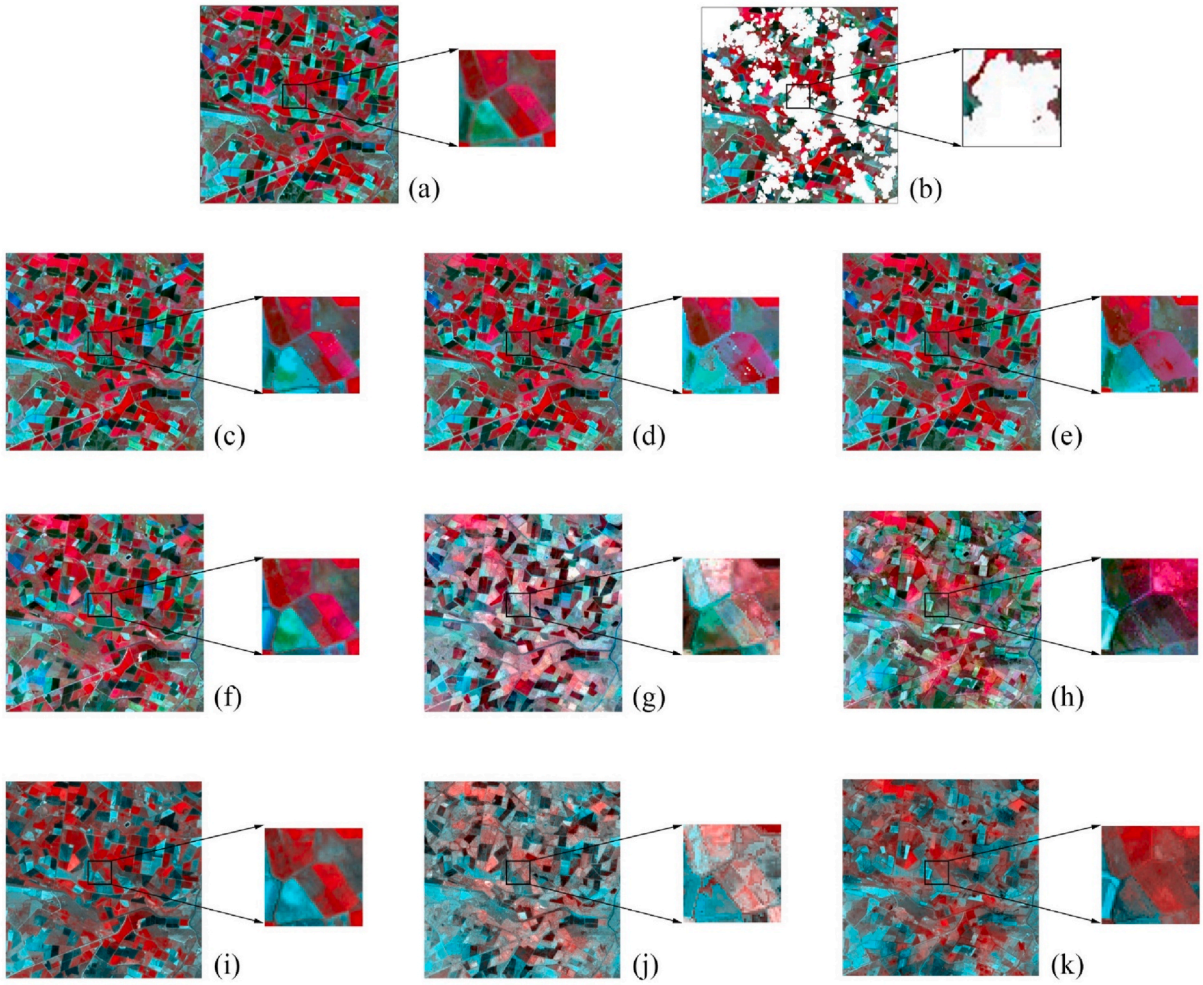

Fig. 9. Spatial patterns of the reconstructed images of Area 1 derived from the available images with different time intervals using NSPI and FSDAF strategies: (a) reference image on DSSD-017, (b) simulated cloudy sub-image on DSSD-017 with cloud coverage of 80.04%, (c)–(e) NSPI-30 m filled images derived on DSSD-033, -097, −193, respectively, (f)–(h) FSDAF-480 m fused images derived on DSSD-033, -097, −193, respectively, (i)–(k) FSDAF-240 m fused images derived on DSSD-033, -097, −193, respectively. (a)–(h) are false-color images (NIR, red and green as R-G-B) and (i)–(k) are false-color images (NIR, red and red as R-G-B) since only NIR and red at 240 m are available for MODIS. The smaller images on the right of each image are the zoom-in sub-image. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

land cover changes.

Then, a Thin Plate Spline (TPS) interpolator is adopted to predict the fine-resolution image from the coarse-resolution image at t_p (the spatial prediction $F_{t_p}^{spat}$). Since this spatial prediction can capture abrupt land cover changes, it was used to guide the redistribution of the residual obtained in Eq. (5), i.e., it is assumed that $F_{t_p}^{spat}$ best represents true values of the fine pixels at t_p and the residual of the homogenous area can be adjusted by the spatial prediction as:

$$R_{ho}(x,y) = F_{t_p}^{spat}(x,y) - F_{t_p}^{temp}(x,y) \tag{6}$$

Finally, as the above operations are implemented within coarse pixels, the prediction of the final pixel at t_p may have block effects. Thus, the similar pixel strategy used in the classic STARFM was used to get the final prediction as:

$$F(x,y,t_p) = F(x,y,t_0) + \sum_{i=1}^{n} W_i\left(F_{t_p}^{r}(x_i,y_i)\right) \tag{7}$$

where $F_{t_p}^{r}$ represents the prediction of the fine pixel after redistributing the residuals, W_i is the weight for the ith similar pixel that is determined by the spectral similarity and spatial distance.

3. Methodology

3.1. Materials

Since it is more challenging to reconstruct cloud-free satellite time series for regions with complex landscapes and fast temporal changes, we implemented comparison studies in these regions to better differentiate the performance of fusing and filling strategies. Two typical sites with different landscapes were selected for experiment: The Coleambally Irrigation Area (CIA) in southern New South Wales (34.87°S, 145.90°E) and the Lower Gwydir Catchment (LGWY) in northern New South Wales (28.27°S, 149.40°E). CIA has a heterogeneous and fragmented landscape where paddy fields are mixed with other croplands and natural vegetation. Besides, the rice croplands are irrigated in October, which leads to large temporal changes in the time series. LGWY is a very temporally dynamic, less spatially heterogeneous site, dominated by relatively large parcels of winter crop fields and natural vegetation. Therefore both CIA and LGWY are ideal sites for studying the performance of fusing and filling strategies and have been proved to be appropriate test sites in former researches (Emelyanova et al., 2013; Rao et al., 2015; Zhou et al., 2021; Zhu et al., 2016, 2018b). To investigate the performance of fusing and filling under different landscape

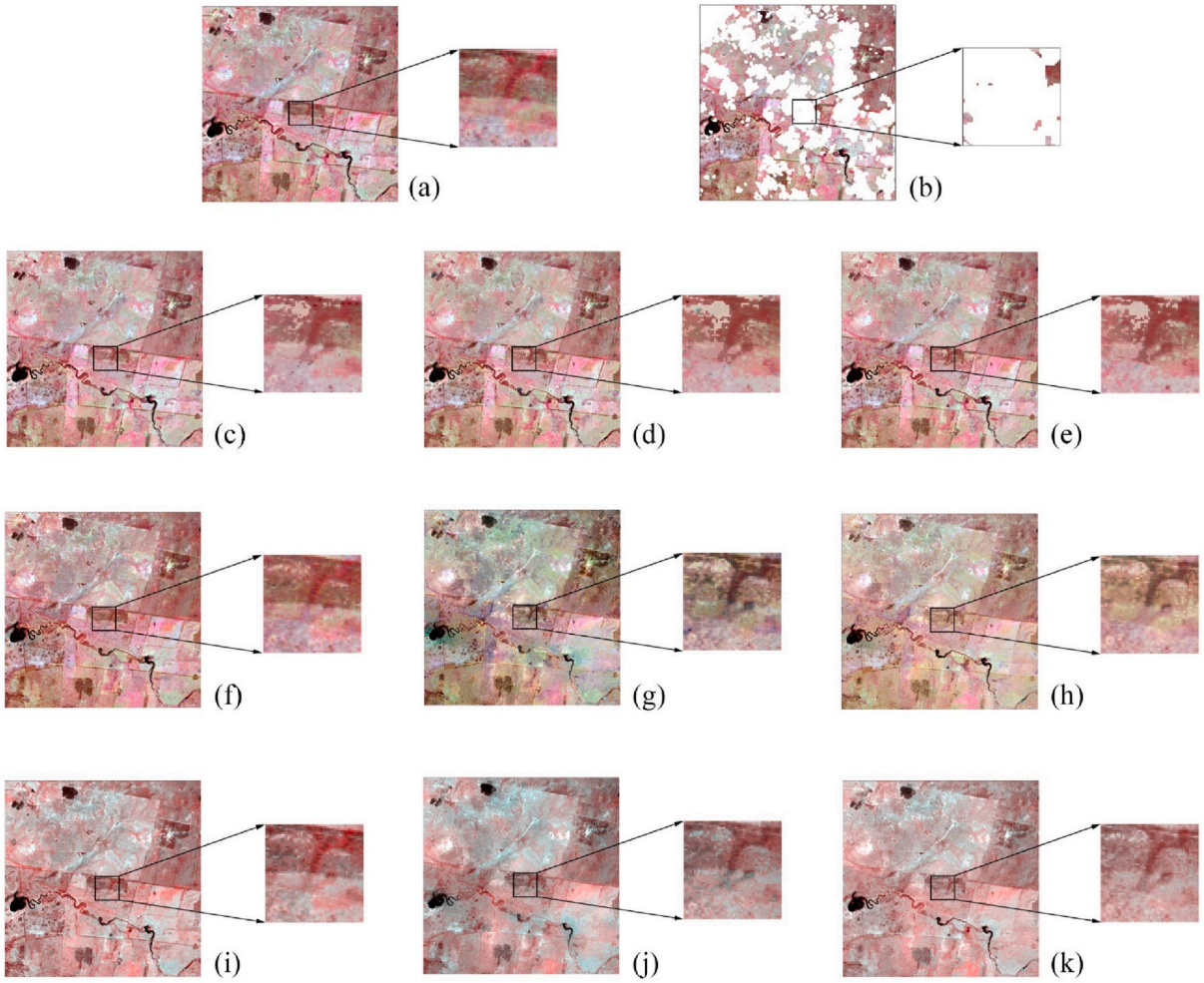

Fig. 10. Spatial patterns of the reconstructed images in Area 2 derived from the available images with different time intervals using NSPI and FSDAF strategies: (a) reference image on DSSD-016, (b) cloudy image with cloud coverage of 93.32%, (c)–(e) NSPI-30 m filled images derived on DSSD-032, -208, −288, respectively, (f)–(h) FSDAF-480 m fused images derived on DSSD-032, -208, −288, respectively, (i)–(k) FSDAF-240 m fused images derived on DSSD-032, -208, −288, respectively. (a)–(h) are false-color images (NIR, red and green as R-G-B) and (i)–(k) are false-color images (NIR, red and red as R-G-B). The smaller images on the right of each image are the zoom-in sub-image. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

heterogeneity, we select one sub-region in CIA (Area 1) dominated by fragmented croplands (high heterogeneity) and one sub-region in LGWY (Area 2) dominated by natural vegetation (low heterogeneity) (Fig. 4).

Level-2 products of Landsat images were employed as the fine-resolution images. MOD09GA/MOD09GQ images were used as coarse-resolution images and were registered with Landsat images. The 250 m and 500 m bands of MODIS images were resampled to 240 m or 480 m to meet the requirement of FSDAF. A total of 15 cloud-free pairs of Landsat-7 ETM+ (400 × 400 pixels at 30 m spatial resolution) data are available in CIA during the austral summer growing season (2001 October – 2002 May), and a total of 13 cloud-free pairs of Landsat-5 TM (400 × 400 pixels at 30 m spatial resolution) data are collected in LGWY from April 2004 to April 2005. Table 1 shows the date information for the cloud-free Landsat-MODIS pairs (LMPs) for the two study sites and Fig. 5 shows the examples of Landsat (30 m) and MODIS (240 m or 480 m) preprocessed images.

3.2. Experimental design

To design the experiments for the comparison of fusing and filling, we first investigated the potential influential factors of each strategy, especially for the two representative methods, NSPI and FSADF. First, as both methods use the spatial and temporal information in the input data, their performance will be affected by the spatial heterogeneity and the time interval between the auxiliary data and the image to be reconstructed. In general, the high spatial heterogeneity makes the spatial prediction in both NSPI and FSDAF (Eqs. (2) and (6)) less reliable and the selection of similar pixels more difficult. Similarly, images with longer time intervals have large temporal changes and even abrupt land cover changes, bringing more challenges to the temporal prediction in NSPI and FSDAF. Second, NSPI may be affected by the cloud size since large clouds may lead to weaker spatial dependence between similar pixels and the target pixel. By contrast, FSDAF is less affected by cloud sizes since it predicts the whole image. Third, FSADF may be affected by the scale ratio between fine- and coarse-resolution images because larger scale differences between them will lead to larger uncertainties in spatial interpolation. By comparison, NSPI does not have this effect as it only uses fine-resolution images. Table 2 summarizes the above influential factors for both NSPI and FSDAF. Accordingly, we designed two experiments to compare the performance of NSPI and FSDAF based on these factors (Fig. 6). Experiment 1 investigates the performance of the two methods using auxiliary data with different time intervals, and Experiment 2 examines the performance of the two methods to reconstruct fine-resolution images with different cloud sizes. In both experiments, we selected two sites with contrasting landscape heterogeneity, and we used coarse-resolution images with two different spatial

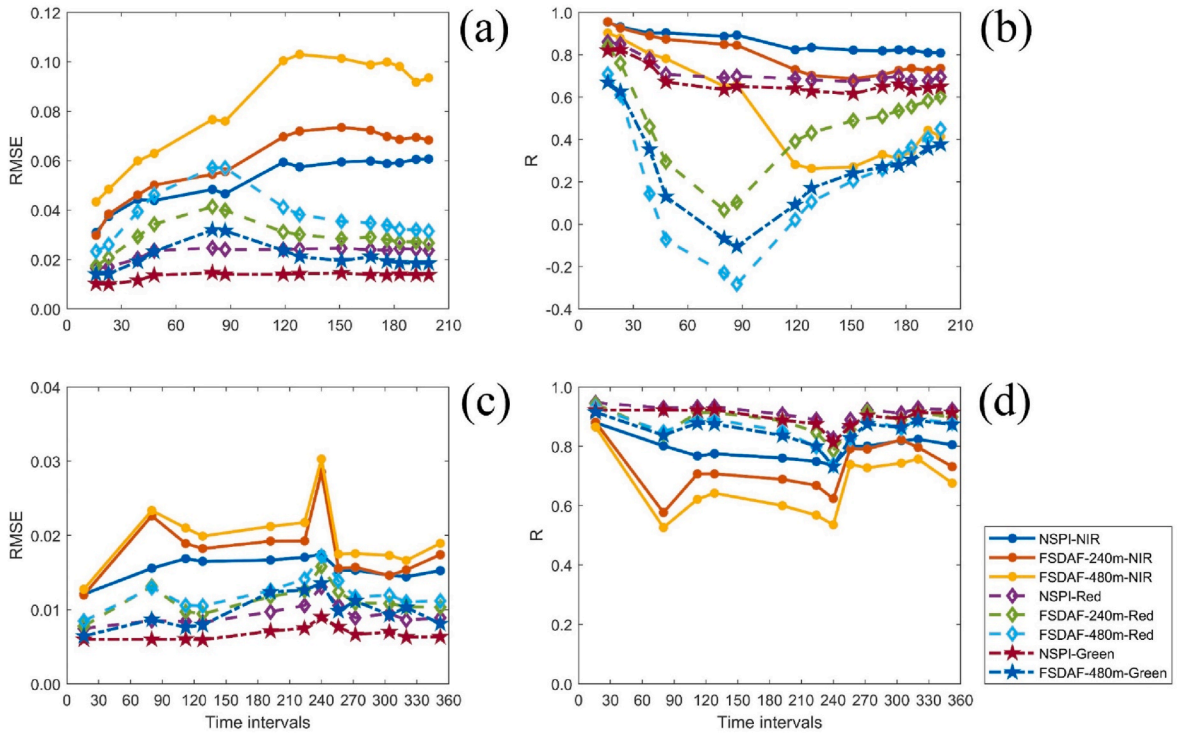

Fig. 11. Comparison of accuracies (i.e., RMSE and R) of the reconstructed NIR, red and green bands derived from NSPI-30 m, FSDAF-240 m and FSDAF-480 m, respectively: (a) and (b) are results of Area 1 (i.e., highly heterogeneous area) and (c) and (d) are results of Area 2 (i.e., low heterogeneous area). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

resolutions (240 m and 480 m) for FSDAF.

Experiment 1: ***different time intervals between the auxiliary images and the target image on two sites***

To examine the impact of time intervals on filling and fusing, we employed the input data with different time intervals in NSPI and FSDAF to predict the Landsat-like image on the first date (i.e., the target image), respectively. For the purpose of evaluating the accuracy of reconstructed images in a quantitative way, for the NSPI method, we applied a real cloud mask extracted from Hong Kong (Zhu and Helmer, 2018) to the target image to simulate a clouded image and grouped the auxiliary and target images into multiple image pairs with different time intervals (Fig. 7). Then these image pairs were used to predict the target image using the NSPI method. For the FSDAF method, we used the base Landsat-MODIS pairs with different time intervals as input to predict the target image. To compare the impact of spatial heterogeneity on the performance of fusing and filling strategies, the experiment was conducted on two sites with distinct spatial heterogeneity.

Experiment 2: ***different sizes of cloud patches on two sites***

To assess the impact of cloud size on filling and fusing strategies, we predicted the target image with NSPI in cases of different cloud sizes and compared it with the results of FSDAF (Experiment 2 in Fig. 6). An initial small cloud patch was selected from a real cloud mask in Fig. 7 (b) and was placed in the image center (Fig. 8 (a)). Then cloud patches with different buffer radii ranging from 1 to 100 were created to simulate the scenario with different cloud sizes (Fig. 8 (b)). For example, the cloud size of 30 means the initial cloud patch is buffered with a distance of 30 pixels. To make the results more robust, for each cloud size, we simulated the cloud mask 100 times and placed it at random positions of the image (Fig. 8 (c)). NSPI was employed in each scenario, and the average performance of the prediction accuracy was calculated for each cloud size (Fig. 8).

3.3. Accuracy assessment

Two classic accuracy indices i.e., the root mean square error (RMSE) and correlation coefficient (R), were selected to evaluate the performance of the filling and fusing methods. They were calculated using all the contaminated pixels of the predicted and true images. The results with different influential factors of the filling and fusion were assessed. RMSE shows the average pixel-wise prediction errors of the contaminated pixels for quantitative analysis while R shows the pixel-wise similarity between the predicted and the true image.

4. Results

4.1. Experiment of time interval

Fig. 9 shows the reconstructed images in Area 1 (i.e., highly heterogeneous area) derived from the available images with different time intervals using NSPI and FSDAF strategies. The results were named as NSPI-30 m, FSDAF-240 m and FSDAF-480 m, based on their methods and spatial resolutions. They reveal that compared with the reference image (i.e., truth values) (Fig. 9 (a)), both NSPI and FSDAF capture more spatial details in highly heterogeneous areas when the available images have shorter time intervals (e.g., one week) (Fig. 9 (c)–(k)). However, NSPI generally performs better than FSADF when the available images have the same time interval. For example, for the time interval of DSSD-097 (i.e., one and a half months), the reconstructed image derived from NSPI (Fig. 9 (d)) has better visual performance than that derived from FSDAF (Fig. 9 (g) and (j)). In addition, the results of FSDAF-240 m and FSDAF-480 m are slightly different when the available images have the same time interval, as shown in Fig. 9 (f)–(k).

Fig. 10 shows the reconstructed images in Area 2 (i.e., low heterogeneous area) derived from the available images with different time

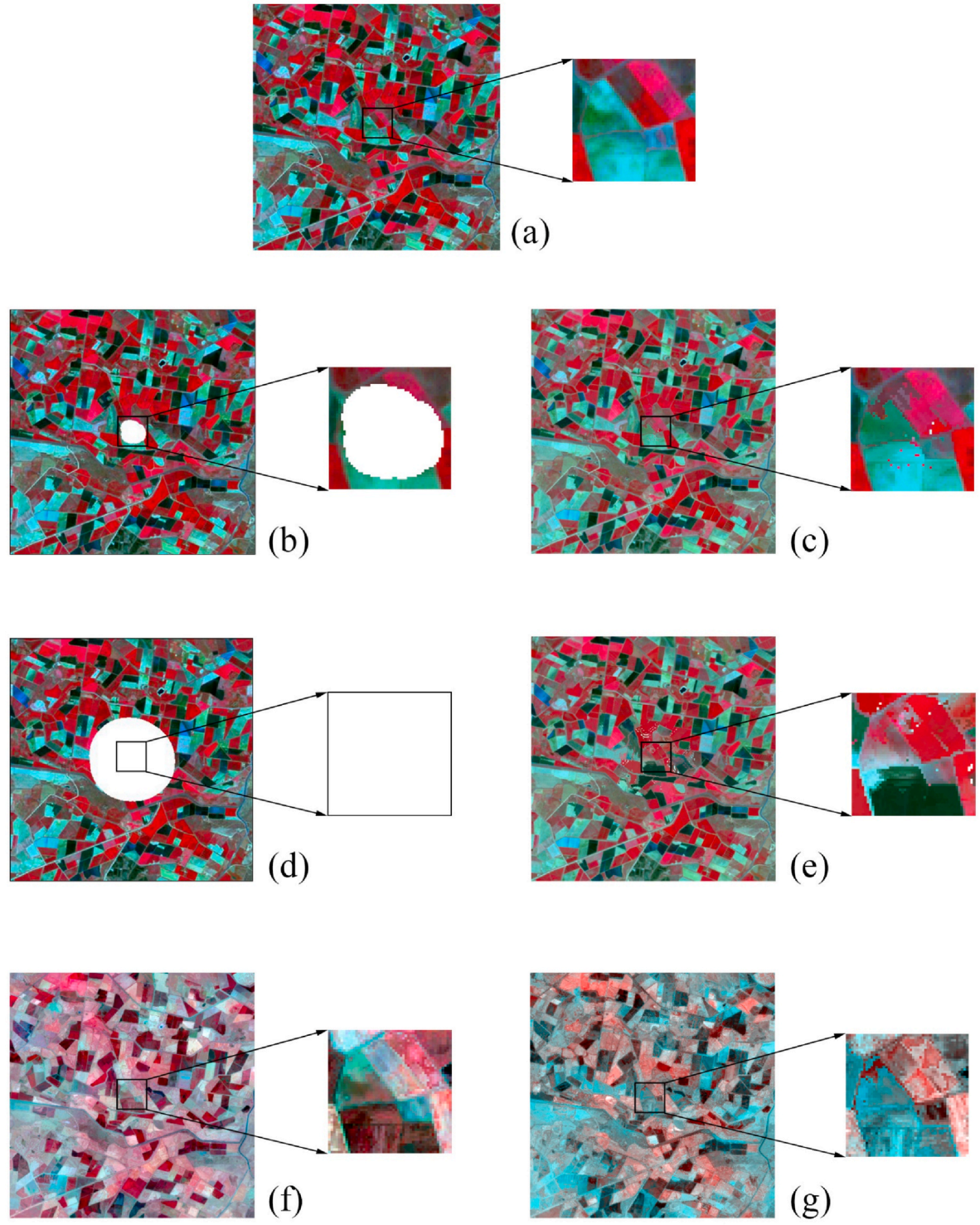

Fig. 12. Spatial patterns of reconstructed images in Area 1 derived from the available images with different cloud sizes using NSPI and FSDAF strategies: (a) reference image on DSSD-017; (b) and (d) cloudy images with cloud coverage of 51.76% and 100%, respectively; (c) and (e) NSPI-30 m filled images derived on DSSD-097 at cloud size of 10 and 60; (f) FSDAF-480 m fused images derived on DSSD-097; (g) FSDAF-240 m fused images derived on DSSD-097. (a)–(f) are false-color images (NIR, Red and Green as R-G-B) and (g) is a false-color image (NIR, Red and Red as R-G-B). The smaller images on the right of each image are the zoom-in sub-image. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

intervals using NSPI and FSDAF strategies. Unlike the highly heterogeneous area (Fig. 9), both filled and fused images in the low heterogeneous area generally show high consistency no matter which time intervals are used (Fig. 10). In other words, for low heterogeneous areas, the two strategies are less affected by the time intervals. However, despite the adaptability of FSDAF for solving the challenge of abrupt land cover changes, the accuracies of fused images decrease with the increase of time intervals and frequent temporal change. As NSPI searched the similar neighborhood pixel by pixel, some small and unexpected patches were filled, resulting in the salt and pepper noise shown in the zoom-in sub-images in Fig. 10 (d) and (e). Compared with the filled results, the fused results derived from FSDAF are smoother,

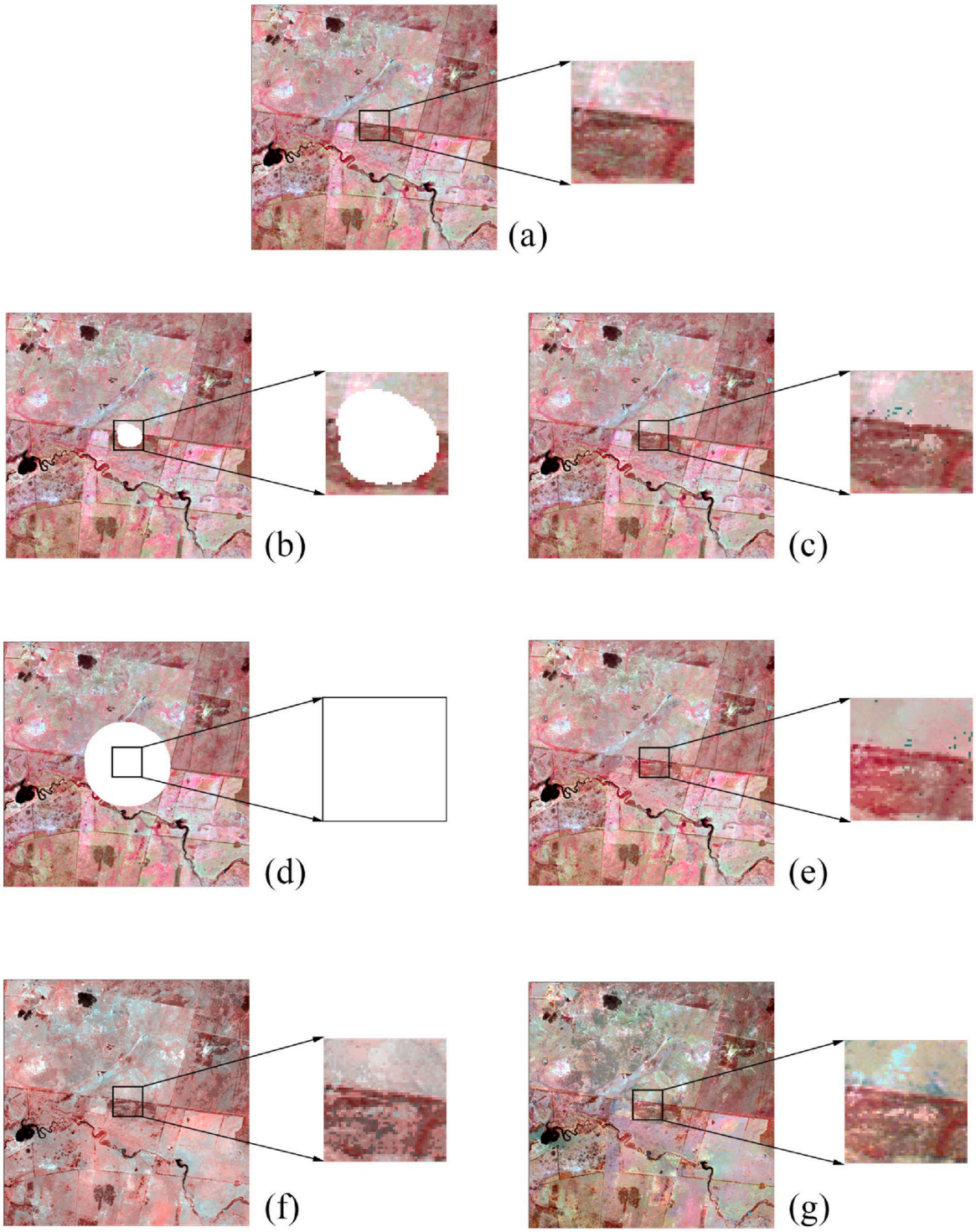

Fig. 13. Spatial patterns of the reconstructed images in Area 2 derived from the available images with different cloud sizes using NSPI and FSDAF strategies: (a) reference image on DSSD-016; (b) and (d) cloudy images with cloud coverage of 51.76% and 100%, respectively; (c) and (e) NSPI-30 m filled images derived from DSSD-128 at cloud size 10 and 60; (f) FSDAF-480 m fused images derived on DSSD-128; (g) FSDAF-240 m fused images derived on DSSD-128. (a)–(f) are false-color images (NIR, red and green as R-G-B) and (g) is a false-color image (NIR, red and red as R-G-B). The smaller images on the right side of each image are the zoom-in sub-image. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

which can contribute to better visual performance (Fig. 10 (f)–(k)).

We further quantified the accuracies (i.e., RMSE and R) of NIR, red and green bands reconstructed by NSPI and FSDAF strategies, respectively, as shown in Fig. 11. For highly heterogeneous areas, the accuracy of the NIR band derived from the two strategies generally decreases with longer time intervals (i.e., larger RMSE and smaller R), as shown in Fig. 11 (a) and (b). Among the NIR bands, the NSPI-30 m NIR band has the highest accuracy while the FSDAF-480 m NIR band has the lowest accuracy whichever time interval is used. In addition, the accuracies of the reconstructed NIR bands of NSPI-30 m and FSDAF-240 m have high similarity for shorter time intervals (e.g., within one month), as shown in Fig. 11 (a) and (b). Unlike the NIR band, the accuracies of NSPI-30 m red

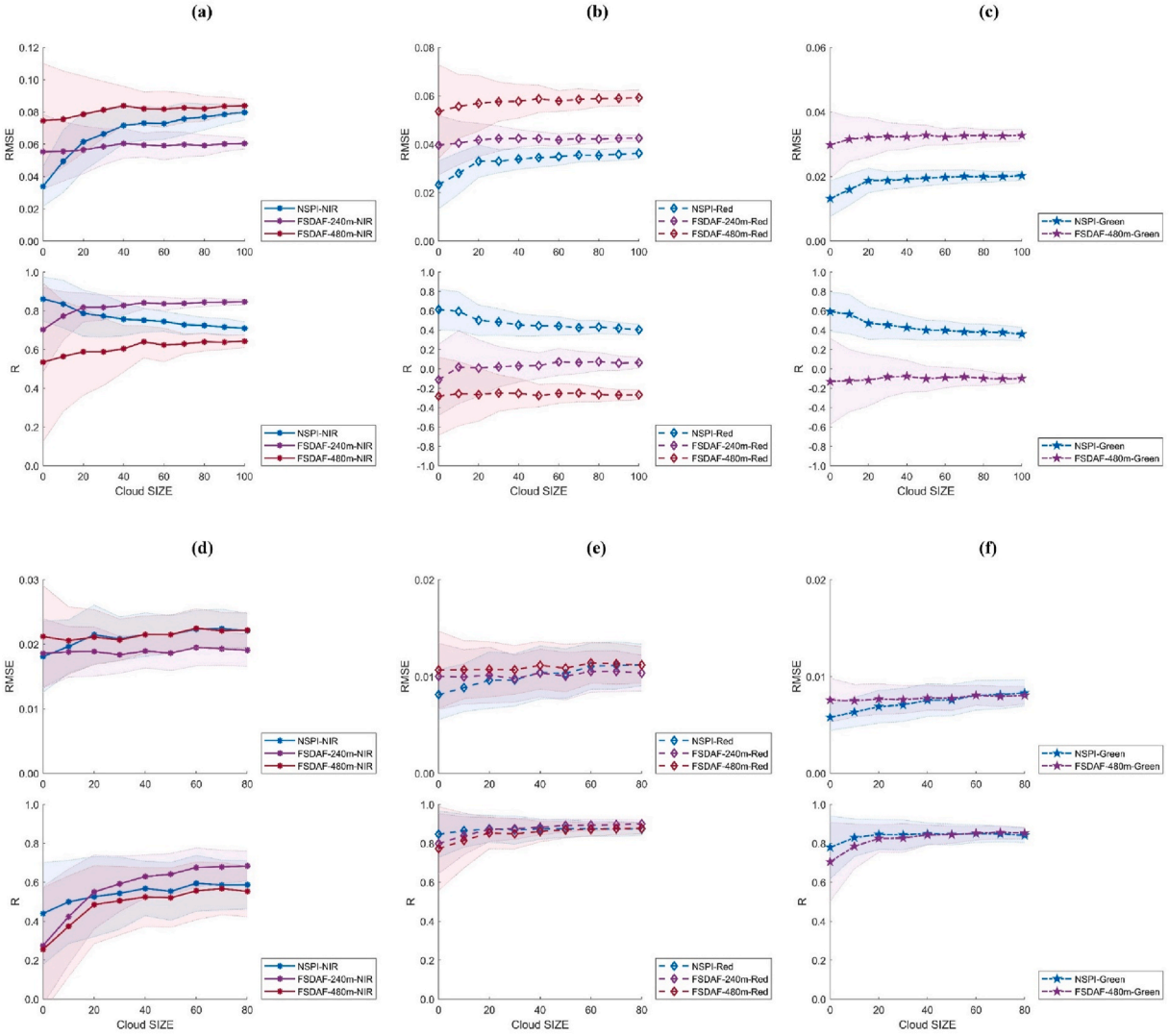

Fig. 14. Comparison of accuracies (i.e., RMSE and R) with shaded error (per SD) of reconstructed NIR, red and green bands derived from NSPI-30 m, FSDAF-240 m and FSDAF-480 m, respectively. (a), (b) and (c) are the results of Area 1 (i.e., highly heterogeneous area); (d), (e) and (f) are the results of Area 2 (i.e., low heterogeneous area). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

and green bands keep stable in different time intervals, but those of FSDAF-derived red and green bands show a distinct declining trend followed by a rise after the inflection point at around 90 days (Fig. 11 (a) and (b)). The accuracy of FSDAF-240 m is generally higher than that of FSDAF-480 m. For low heterogeneous areas, the NSPI method generally has a better performance than FSDAF (Fig. 11 (c) and (d)). The abnormal sudden increase of FSDAF-derived NIR band in Fig. 11 (c) at 240 days interval can be attributed to the occurrence of local flooding events. Also, all NSPI-30 m bands have similar performance regardless of time interval and the accuracies of FSDAF-derived bands generally have slight fluctuations due to temporal dynamic change, as shown in Fig. 11 (c) and (d).

4.2. Experiment of cloud size

Figs. 12 and 13 show the reconstructed images in Area 1 (i.e., highly heterogeneous area) and Area 2 (i.e., low heterogeneous area), respectively, with different simulated cloud sizes using NSPI and FSDAF strategies and the auxiliary image on DSSD-097 (Area 1) and DSSD-128 (Area 2) (i.e., an auxiliary image can get moderate accuracy shown in Experiment 1). As the FSDAF method does not use cloudy images, the cloud sizes only affect the NSPI method. The results were named as NSPI-cloud10 (10-pixel buffered cloud which is a relatively small cloud), NSPI-cloud60 (60-pixel buffered cloud which is a relatively large cloud), FSDAF-240 m and FSDAF-480 m. For highly heterogeneous areas, NSPI-cloud10 has better performance than NSPI-cloud60, thereby suggesting that cloud size is a key factor for NSPI filling results (Fig. 12 (c) and (e)). In addition, the NSPI method generally captures more spatial details than the FSDAF method (Fig. 12 (c), (e), (f) and (g)). For low heterogeneous areas, the reconstructed images have high similarity whichever method and cloud size were used (Fig. 13). However, some small and unexpected salt and pepper noises are filled by NSPI, as shown in Fig. 12 (c), (e) and Fig. 13 (c), (e).

The accuracies of NIR, red and green bands reconstructed by NSPI and FSDAF strategies was further quantified, as shown in Fig. 14. In highly heterogeneous areas, the NIR band derived from NSPI has a lower accuracy (i.e., larger RMSE and smaller R) with a larger size of cloud patch (Fig. 14 (a)). By comparison, it has higher accuracy than the NIR band derived from FSDAF-240 m and FSDAF-480 m when the cloud patch is small. Then FSDAF-240 m NIR band gradually performs better than the NSPI NIR band with the increase in cloud size. Similar to the NIR band, the accuracies of NSPI-30 m red and green bands are also affected by the size of cloud patches, but the accuracies of NSPI-30 m red and green bands are generally higher than those of FSDAF-derived red

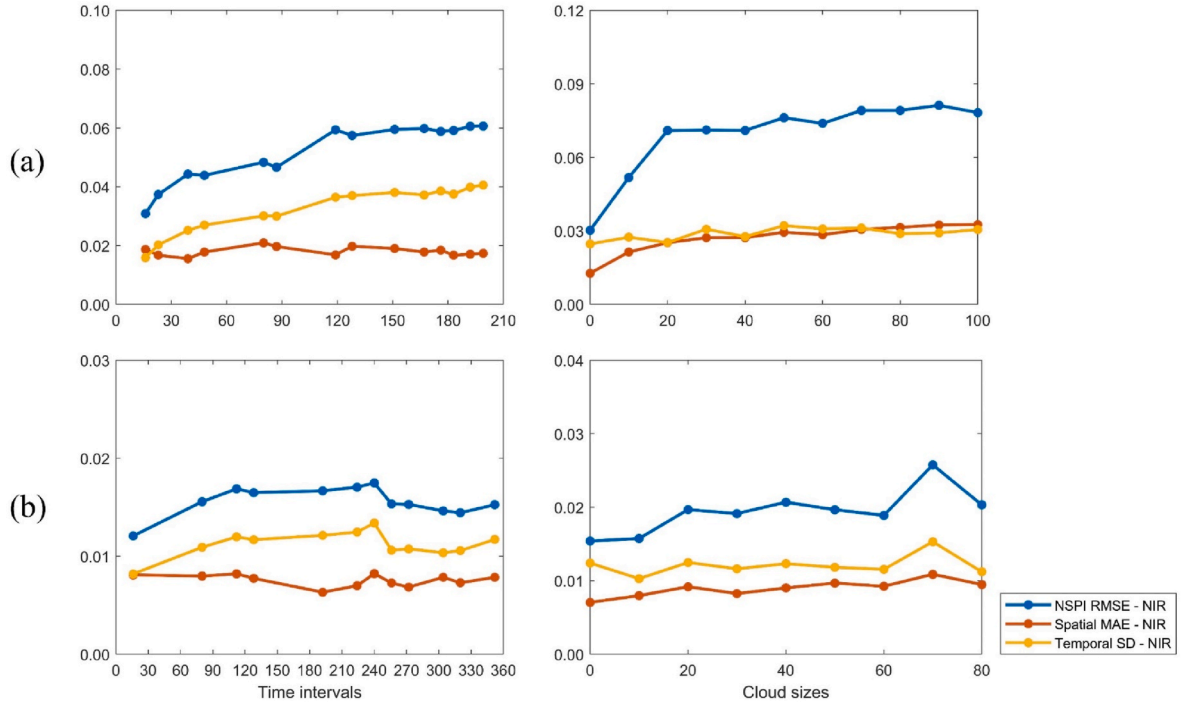

Fig. 15. Error analysis of the NSPI in different time intervals (the subplots on the left) and different cloud sizes (the two subplots on the right). (a) is Area 1 (highly heterogeneous area) and (b) is Area 2 (low heterogeneous area).

Table 3
The correlation between the accuracy of the NSPI method, the spatial MAE and the temporal SD at different time intervals and cloud sizes.

	Time intervals	Cloud sizes
Area 1	R (NSPI/Spatial MAE) = −0.041	R (NSPI/Spatial MAE) = 0.978
	R (NSPI/Temporal SD) = 0.991	R (NSPI/Temporal SD) = 0.688
Area 2	R (NSPI/Spatial MAE) = −0.162	R (NSPI/Spatial MAE) = 0.913
	R (NSPI/Temporal SD) = 0.960	R (NSPI/Temporal SD) = 0.771

and green bands (Fig. 14 (b) and (c)). In low heterogeneous areas, the accuracies of NSPI-30 m (all bands) have similar performance with Area 1 (Fig. 14 (d), (e) and (f)), while the difference between NSPI-30 m and FSDAF is smaller than Area 1. It is obvious that the performance of NSPI-30 m is significantly affected by the size of cloud patches for all bands, and in that case FSDAF remains a stable performance.

5. Discussion

The performance of NSPI and FSDAF is impacted to different degrees by various factors, including the landscape characteristics, land surface dynamics, cloud sizes, and the availability of could-free images. In general, we cannot simply claim which of the two methods is superior for reconstructing fine-resolution time series. In this section, we further discuss the mechanism of how different factors affect the performance of NSPI, FSDAF, and the combination of both strategies. For simplicity, only the NIR band was analyzed in error analysis. We also give suggestions to end users for selecting the appropriate strategy in applications.

5.1. Error analysis of NSPI

The experimental results show that NSPI is sensitive to time intervals and cloud sizes. Based on the introduction of NSPI in Section 2, NSPI has two assumptions: (1) neighboring similar pixels have a high degree of spectral similarity; and (2) this spectral similarity can be kept in the time series. We inferred that the low accuracy of NSPI in some scenarios might be caused by the violation of these two assumptions. To verify this, two variables S_i and T_i were calculated to represent these two assumptions respectively:

$$S_i = MEAN(ABS(L(x_i, y_i, t_1) - L(x, y, t_1))) \quad (8)$$

$$T_i = STD\left(L\left(x_i, y_i, t_p\right) - L(x_i, y_i, t_1)\right) \quad (9)$$

where S_i is the spatial mean absolute error (MAE) between all similar pixels and a target pixel which corresponds to assumption 1, and T_i is the temporal standard deviation (SD) of temporal changes among similar pixels which corresponds to assumption 2.

Comparisons of the spatial MAE and temporal SD with the RMSE of NSPI in two different areas reveal some interesting patterns (Fig. 15). We further calculated the correlation coefficient (R) between the accuracy of NSPI and the spatial MAE and the correlation coefficient between the accuracy of NSPI and the temporal SD (Table 3). It shows that the RMSE of NSPI has a higher correlation with the temporal SD ($R_{Area\ 1}$ = 0.991 and R_{Area2} = 0.960) in the experiment of time interval, and the accuracy of NSPI has a higher correlation with the spatial error ($R_{Area\ 1}$ = 0.978 and $R_{Area\ 2}$ = 0.913) in the experiment of cloud size. It suggests that the violation of assumption 2 in NSPI and abrupt land cover changes are the main reasons for the low accuracy of NSPI when using the input image with long time intervals, while the violation of assumption 1 and the lack of neighborhood similar pixels are the main reasons for the low accuracy of NSPI for removing large cloud patches. These effects are more significant in heterogeneous areas (Area 1).

5.2. Error analysis of FSDAF

According to the principle of FSDAF reviewed in Section 2, its error mainly comes from the temporal change derived by the unmixing method and the residual redistribution guided by the TPS interpolation. To determine whether unmixing or TPS interpolation dominates the errors in FSDAF, two indicators were calculated: one is the absolute relative difference index (ARDI) (Zhou et al., 2021) which quantifies the

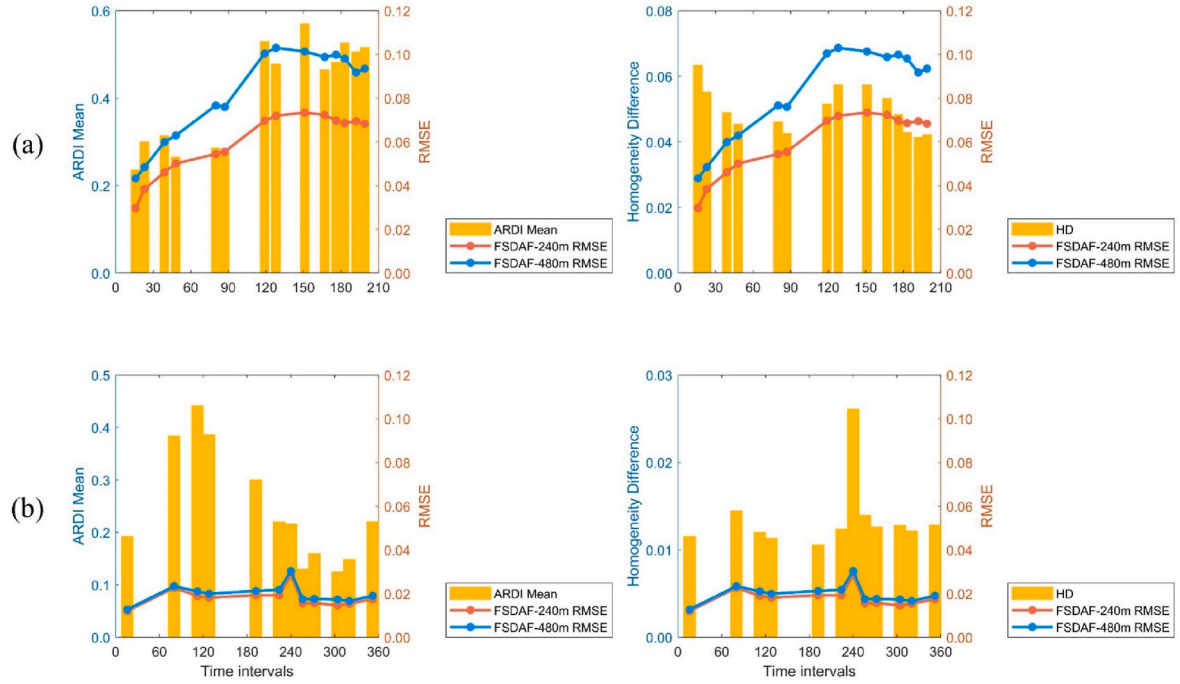

Fig. 16. Error analysis of the FSDAF with ARDI (the left side) and HD (the right side) in different time intervals. (a) is Area 1 (high heterogeneous area) and (b) is Area 2 (low heterogeneous area).

Table 4
The correlation between the accuracy of the FSDAF method, ARDI and HD in different time intervals.

	FSDAF-240 m	FSDAF-480 m
Area 1	R (ARDI/FSDAF-240 m) = 0.889	R (ARDI/FSDAF-480 m) = 0.891
	R (HD/FSDAF-240 m) = −0.304	R (HD/FSDAF-480 m) = −0.241
Area 2	R (ARDI/FSDAF-240 m) = 0.356	R (ARDI/FSDAF-480 m) = 0.318
	R (HD/FSDAF-240 m) = 0.774	R (HD/FSDAF-480 m) = 0.773

temporal change between the auxiliary and prediction time, corresponding to the uncertainty in unmixing, and the other is the homogeneity difference (HD) which calculates the difference between MODIS and Landsat data, corresponding to the uncertainty in TPS interpolation.

$$ARDI = |F_{tp} - F_{t0}| \,/\, F_{t0} \tag{10}$$

$$HD = |F_{t0} - C_{t0}| \tag{11}$$

where F_{t0} and F_{tp} are the reflectance of the fine-resolution images at the auxiliary and prediction time respectively, and C_{t0} is the coarse-resolution image at the auxiliary time.

Fig. 16 shows the error analysis between the RMSE of FSDAF, ARDI and HD. The correlation coefficient (R) between the accuracy of FSDAF and ARDI and the correlation coefficient between the accuracy of the FSDAF and HD are shown in Table 4. ARDI has a high correlation with

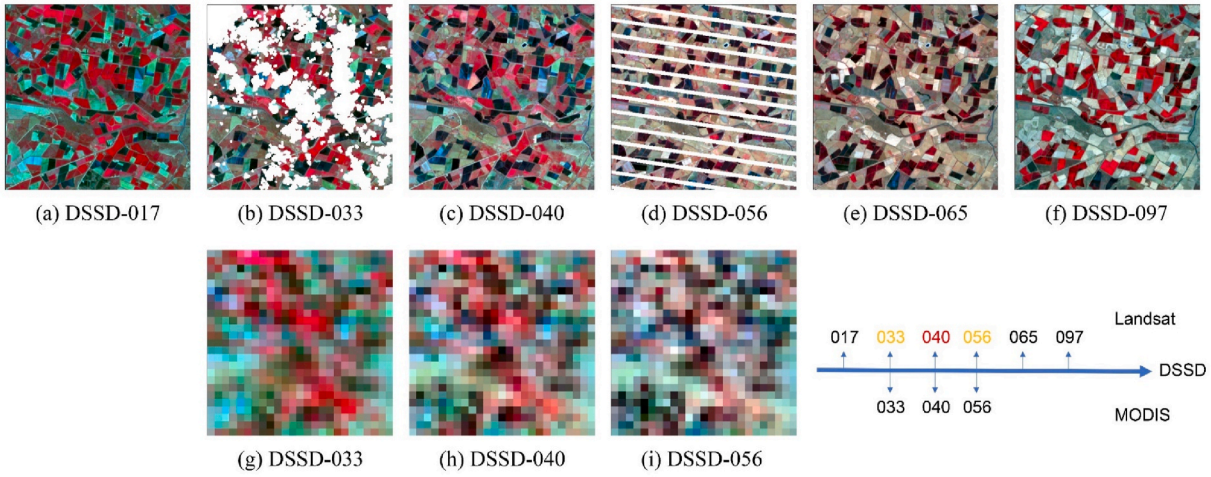

Fig. 17. Data preparation for STAIR. (a), (e), (f) clear Landsat image on DSSD-017, -065, −097, (b) simulated cloudy Landsat image on DSSD-033, (c) original reference Landsat image on DSSD-040, (d) simulated un-scanned Landsat image on DSSD-056, (g)–(i) clear MODIS image on DSSD-033, -040, −056. A Schematic diagram at the bottom right shows the input data in timeseries, number above the timeline means Landsat available, numbers below the timeline means MODIS available, number in black means clear image, number in yellow means cloudy or un-scanned image, number in red means the target image to be predicted. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

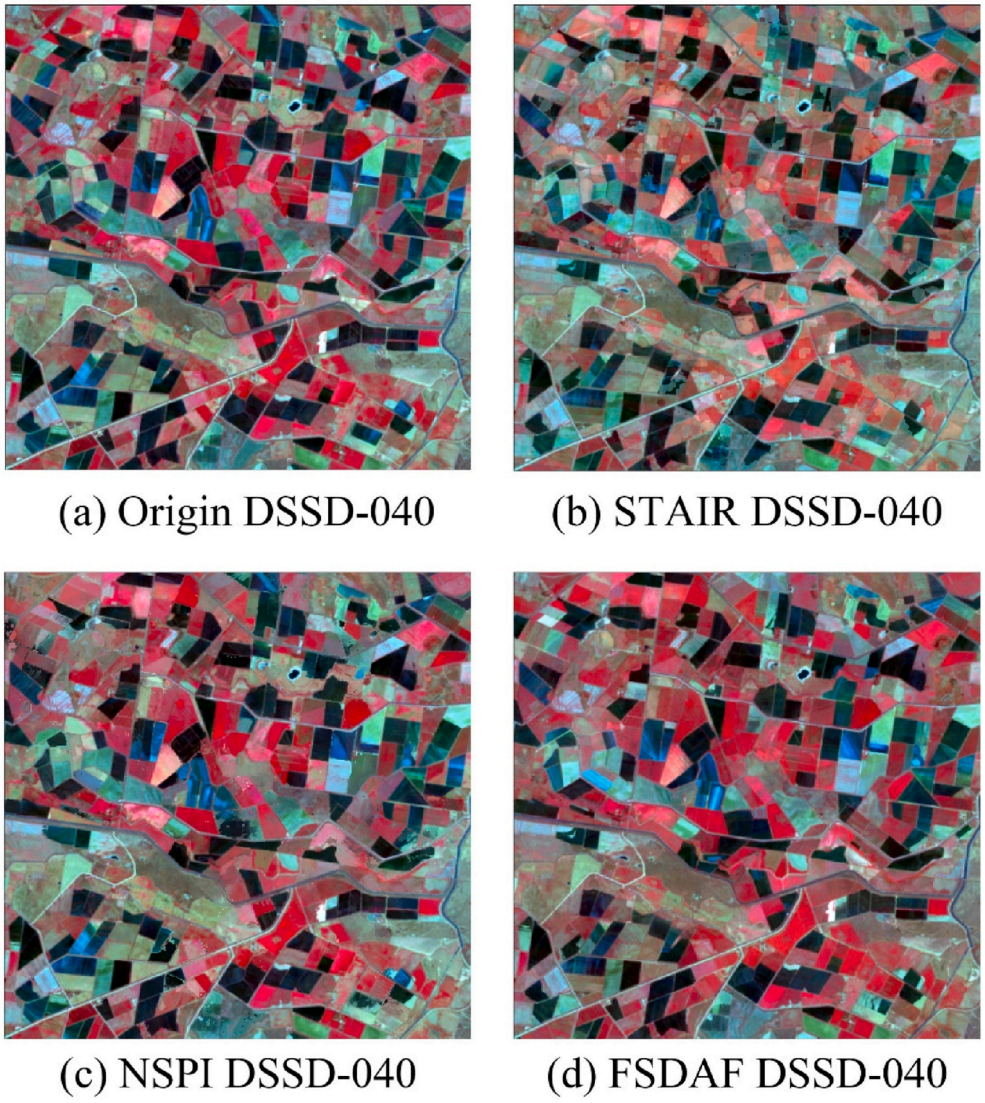

Fig. 18. Spatial pattern of three reconstructed Landsat-like images on DSSD-040 comparing original reference image.

Table 5
Quantitative accuracy (RMSE and R) of STAIR, NSPI and FSDAF at different bands and computing time using a laptop computer with an Intel Core i7-9750H CPU and 16 GB memory.

		NIR	Green	R	Blue	Computing time
STAIR	RMSE	0.026	0.011	0.019	0.009	7.1 s (MATLAB)
	R	0.951	0.843	0.846	0.852[a]	
NSPI	RMSE	0.030	0.010[a]	0.018[a]	0.008[a]	25.8 s (Cloud:
	R	0.930	0.779	0.773	0.792	37.19%)
FSDAF	RMSE	0.025[a]	0.013	0.018[a]	0.011	51.9 s (MODIS 480
	R	0.962[a]	0.845[a]	0.863[a]	0.850	m)

[a] The best accuracy among three methods.

Table 6
Recommended strategies in different scenarios.

Scenarios	NSPI	FSDAF for images with smaller scale difference	FSDAF for images with larger scale difference
Small cloud patches	✓	×	×
Medium cloud patches	×	✓	×
Large cloud patches	×	✓	✓
Auxiliary data with short time interval	✓	✓	×
Auxiliary data with long time interval	✓	×	×
Heterogeneous areas	✓	✓	×
Homogeneous areas	✓	✓	✓

the accuracy of FSDAF in Area 1 ($R_{FSDAF\text{-}240m} = 0.889$ and $R_{FSDAF\text{-}480m} = 0.891$), and HD has a high correlation with the accuracy of FSDAF in Area 2 ($R_{FSDAF\text{-}240m} = 0.774$ and $R_{FSDAF\text{-}480m} = 0.773$). This result suggests that the unmixing step could bring more errors in areas with large spatial heterogeneity since the temporal changes are more diverse in heterogeneous areas. On the contrary, the TPS spatial interpolation brings more errors than the unmixing step in homogenous areas due to the redistribution of the residuals in FSDAF.

It is worth noting that the performance of NSPI and FSDAF may be affected by the uncertainty in similar pixel selection. In addition, the different temporal variations among similar pixels also lead to large errors in the prediction by both NSPI and FSDAF. Similar pixels for NSPI and FSDAF can be selected based on either unsupervised classification or supervised classification depending on the specific application and data availability (i.e., the ground reference data). Although supervised classification can potentially generate more accurate class maps thus better similar pixels, it does increase the complexity of the methods and

therefore decrease their applicability. Accordingly, there are two advantages in selecting similar pixels based on unsupervised classification as used in this paper: (1) making the whole filling or fusing process automatic, (2) increased applicability of the methods.

5.3. Combination of filling and fusing strategies

Combination of filling and fusing could potentially produce more accurate high-quality fine-resolution time-series images if the combination can take advantage of both the filling and fusing methods and avoid the weakness from both. Otherwise, the improvement with combination may be limited due to the error propagation. Some filling and fusing integrated methods have been developed in previous studies, and they often use a two-step mode, i.e., filling then fusing (Luo et al., 2018) or fusing then filling (Li et al., 2019).

To test whether the combination of filling and fusing can produce better results than using filling or fusing alone, we compared the performance of STAIR (SaTellite dAta IntegRation proposed in Luo et al., 2018) with NSPI and FSDAF. STAIR is a generic and fully-automated two-step method to generate a high-resolution time series, which first fill gap pixels in the time series of Landsat-MODIS pairs, and then use a fusion framework for fusing gap-filled Landsat and MODIS images. We generated a series of input data in Area 1 of CIA (Fig. 17) to meet the requirement of STAIR method for predicting the DSSD-040 image. In addition, we use the same dataset to test NSPI and FSDAF. For NSPI, we select clear Landsat image on DSSD-017, -065, 097 and a simulated cloudy image (cloud coverage: 37.19%) on DSSD-040. For FSDAF, we select the temporally-closest Landsat-MODIS pair on DSSD-033 and clear MODIS image on DSSD-040.

For these three predicted Landsat-like images on DSSD-040 using STAIR, NSPI and FSDAF, we compare them in visual and quantitative ways. Fig. 18 shows the spatial pattern of three reconstructed Landsat-like images on DSSD-040 comparing original reference image, while Table 5 shows the quantitative accuracy (RMSE and R) at different bands and computation speed of three methods. It is obvious that these three reconstructed Landsat-like images and the original reference image have very similar spatial pattern, and accuracy of three methods are quite similar. NSPI and FSDAF use longer computing time than STAIR, but they can be further speeded up using parallel computing (Gao et al., 2021) for processing massive images due to their simple structure. The comparison results shows that the combination method, STAIR, cannot outperform NSPI and FSDAF in some spectral bands, suggesting that combination methods may not fully take advantage of the strength of filling and fusing strategy. The two-step combination of filling and fusing may introduce additional errors.

Although single-strategy-method like NSPI and FSDAF still has its superiority in real-world applications, more and more studies have developed new strategies to combine filling and fusing for image reconstruction (Li et al., 2021). However, as our experiment on STAIR showed, the combination of two strategies does not necessarily perform better than a single strategy. Before developing a better combination strategy, it is critical to investigate the strength and weakness of both the filling and fusing methods. Our study fills this gap by comparing the influencing factors of two representative filling and fusing methods (NSPI and FSDAF), shedding light for future development of combination methods.

5.4. Suggestion on the selection of appropriate strategy

The performance of NSPI and FSDAF varied in different landscapes and different input data. In the experiment of time interval in the highly heterogeneous area, NSPI performs better than FSDAF, and the NIR band is the most sensitive band among the three selected bands. By contrast, the performance of the red and green bands remains stable except for the FSDAF-480 m fusion result. In the experiment of cloud size in the highly heterogeneous area, NSPI is significantly affected by the size of cloud patches. There is an intersection point between the performance of NSPI and the FSDAF-240 m as the cloud size increases for the NIR band (Fig. 14 (a)), suggesting that NSPI performs better at a small cloud size, while with the increase in cloud size, FSDAF-240 m performs better than NSPI and the average accuracy of NSPI decreases quickly. Different from Area 1, all bands simulated in the experiments of time interval and cloud size in Area 2 have similar performance between NSPI and FSDAF due to its homogeneity landscape. Table 6 lists the recommended strategies in different scenarios based on this study, serving as a guideline for further studies.

6. Conclusion

Filling and fusing are two main strategies to reconstruct high-quality fine-resolution satellite time series data. This study selected NSPI and FSDAF as the representative method in each strategy and designed experiments according to their potential influential factors (i.e., the time interval of auxiliary data, size of cloud patches, heterogeneity of landscapes) to compare their performance. In addition, we gave recommendations on the selection of appropriate methods in future research. NSPI and FSDAF have their superiority for reconstructing fine-resolution time series in different scenarios. The results suggest that the accuracy of both methods decreases with the time interval between the predicted image and the auxiliary image, and NSPI is a better choice when the auxiliary data has a larger time interval. For the image covered by clouds with different sizes, the accuracy of NSPI significantly decreases with cloud size, suggesting that FSDAF is a better choice when the images are covered by large clouds. Scale ratio is a significant factor affecting the performance of FSDAF, especially for heterogeneous areas, so a smaller scale ratio is recommended for using FSDAF to reconstruct fine-resolution time series. The findings of this study can help users select the appropriate method to reconstruct satellite time series for their specific applications. The procedure employed in this study can be used by future studies to investigate more filling and fusing methods.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (Project No.42022060) and the Hong Kong Polytechnic University (Project No. ZVN6).

References

Chen, J., Zhu, X., Vogelmann, J.E., Gao, F., Jin, S., 2011. A simple and effective method for filling gaps in Landsat ETM+ SLC-off images. Remote Sens. Environ. 115, 1053–1064. https://doi.org/10.1016/j.rse.2010.12.010.

Cheng, Q., Shen, H., Zhang, L., Li, P., 2014. Inpainting for remotely sensed images with a multichannel nonlocal total variation model. IEEE Trans. Geosci. Rem. Sens. 52, 175–187. https://doi.org/10.1109/TGRS.2012.2237521.

Emelyanova, I.V., McVicar, T.R., Van Niel, T.G., Li, L.T., van Dijk, A.I.J.M., 2013. 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. 133, 193–209. https://doi.org/10.1016/j.rse.2013.02.007.

Gao, F., Masek, J., Schwaller, M., Hall, F., 2006. On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance. IEEE Trans. Geosci. Rem. Sens. 44, 2207–2218. https://doi.org/10.1109/TGRS.2006.872081.

Gao, H., Zhu, X., Guan, Q., Yang, X., Yao, Y., Zeng, W., Peng, X., 2021. cuFSDAF: an enhanced flexible spatiotemporal data fusion algorithm parallelized using graphics processing units. IEEE Trans. Geosci. Rem. Sens. 60, 1–16. https://doi.org/10.1109/TGRS.2021.3080384.

Guo, D., Shi, W., Hao, M., Zhu, X., 2020. FSDAF 2.0: improving the performance of retrieving land cover changes and preserving spatial details. Remote Sens. Environ. 248, 111973. https://doi.org/10.1016/j.rse.2020.111973.

Hansen, M.C., Roy, D.P., Lindquist, E., Adusei, B., Justice, C.O., Altstatt, A., 2008. A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in the Congo Basin. Remote Sens. Environ. 112, 2495–2513. https://doi.org/10.1016/j.rse.2007.11.012.

Huang, B., Song, H., 2012. Spatiotemporal reflectance fusion via sparse representation. IEEE Trans. Geosci. Rem. Sens. 50, 3707–3716. https://doi.org/10.1109/TGRS.2012.2186638.

Johnson, M.D., Hsieh, W.W., Cannon, A.J., Davidson, A., Bédard, F., 2016. Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods. Agric. For. Meteorol. 218–219, 74–84. https://doi.org/10.1016/j.agrformet.2015.11.003.

Ju, J., Roy, D.P., 2008. The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally. Remote Sens. Environ. 112, 1196–1211. https://doi.org/10.1016/j.rse.2007.08.011.

Li, S., Xu, L., Jing, Y., Yin, H., Li, X., Guan, X., 2021. High-quality vegetation index product generation: a review of NDVI time series reconstruction techniques. Int. J. Appl. Earth Obs. Geoinf. 105, 102640. https://doi.org/10.1016/j.jag.2021.102640.

Li, X., Foody, G.M., Boyd, D.S., Ge, Y., Zhang, Y., Du, Y., Ling, F., 2020. SFSDAF: an enhanced FSDAF that incorporates sub-pixel class fraction change information for spatio-temporal image fusion. Remote Sens. Environ. 237, 111537. https://doi.org/10.1016/j.rse.2019.111537.

Li, X., Ling, F., Foody, G.M., Ge, Y., Zhang, Y., Du, Y., 2017. Generating a series of fine spatial and temporal resolution land cover maps by fusing coarse spatial resolution remotely sensed images and fine spatial resolution land cover maps. Remote Sens. Environ. 196, 293–311. https://doi.org/10.1016/j.rse.2017.05.011.

Li, X., Wang, L., Cheng, Q., Wu, P., Gan, W., Fang, L., 2019. Cloud removal in remote sensing images using nonnegative matrix factorization and error correction. ISPRS J. Photogrammetry Remote Sens. 148, 103–113. https://doi.org/10.1016/j.isprsjprs.2018.12.013.

Liao, L., Song, J., Wang, J., Xiao, Z., Wang, J., 2016. Bayesian method for building frequent Landsat-like NDVI datasets by integrating MODIS and Landsat NDVI. Rem. Sens. 8, 452. https://doi.org/10.3390/rs8060452.

Liu, M., Yang, W., Zhu, X., Chen, J., Chen, X., Yang, L., Helmer, E.H., 2019. An Improved Flexible Spatiotemporal DAta Fusion (IFSDAF) method for producing high spatiotemporal resolution normalized difference vegetation index time series. Remote Sens. Environ. 227, 74–89. https://doi.org/10.1016/j.rse.2019.03.012.

Liu, X., Deng, C., Wang, S., Huang, G.-B., Zhao, B., Lauren, P., 2016. Fast and accurate spatiotemporal fusion based upon extreme learning machine. IEEE Geosci. Rem. Sens. Lett. 13, 2039–2043. https://doi.org/10.1109/LGRS.2016.2622726.

Luo, Y., Guan, K., Peng, J., 2018. STAIR: a generic and fully-automated method to fuse multiple sources of optical satellite data to generate a high-resolution, daily and cloud-/gap-free surface reflectance product. Remote Sens. Environ. 214, 87–99. https://doi.org/10.1016/j.rse.2018.04.042.

Masek, J.G., Huang, C., Wolfe, R., Cohen, W., Hall, F., Kutler, J., Nelson, P., 2008. North American forest disturbance mapped from a decadal Landsat record. Remote Sens. Environ. 112, 2914–2926. https://doi.org/10.1016/j.rse.2008.02.010.

Maxwell, S.K., Schmidt, G.L., Storey, J.C., 2007. A multi-scale segmentation approach to filling gaps in Landsat ETM+ SLC-off images. Int. J. Rem. Sens. 28, 5339–5356. https://doi.org/10.1080/01431160601034902.

Rao, Y., Zhu, X., Chen, J., Wang, J., 2015. An improved method for producing high spatial-resolution NDVI time series datasets with multi-temporal MODIS NDVI data and Landsat TM/ETM+ images. Rem. Sens. 7, 7865–7891. https://doi.org/10.3390/rs70607865.

Roy, D.P., Ju, J., Lewis, P., Schaaf, C., Gao, F., Hansen, M., Lindquist, E., 2008. Multi-temporal MODIS–Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data. Remote Sens. Environ. 112, 3112–3130. https://doi.org/10.1016/j.rse.2008.03.009.

Shen, H., Meng, X., Zhang, L., 2016. An integrated framework for the spatio–temporal–spectral fusion of remote sensing images. IEEE Trans. Geosci. Rem. Sens. 54, 7135–7148. https://doi.org/10.1109/TGRS.2016.2596290.

Shen, M., Tang, Y., Chen, J., Zhu, X., Zheng, Y., 2011. Influences of temperature and precipitation before the growing season on spring phenology in grasslands of the central and eastern Qinghai-Tibetan Plateau. Agric. For. Meteorol. 151, 1711–1722. https://doi.org/10.1016/j.agrformet.2011.07.003.

Shi, C., Wang, X., Zhang, M., Liang, X., Niu, L., Han, H., Zhu, X., 2019. A comprehensive and automated fusion method: the enhanced flexible spatiotemporal DAta fusion model for monitoring dynamic changes of land surface. Appl. Sci. 9, 3693. https://doi.org/10.3390/app9183693.

Townshend, J., Justice, C., Li, W., Gurney, C., McManus, J., 1991. Global land cover classification by remote sensing: present capabilities and future possibilities. Remote Sens. Environ. 35, 243–255. https://doi.org/10.1016/0034-4257(91)90016-Y.

USGS, 2004. Phase 2 gap-fill algorithm: SLC-off gap-filled products gap-fill algorithm methodology. Available online at. https://www.usgs.gov/media/files/landsat-7-slc-gap-filled-products-phase-two-methodology. (Accessed 27 December 2021).

Wang, B., Jia, K., Wei, X., Xia, M., Yao, Y., Zhang, X., Liu, D., Tao, G., 2020. Generating spatiotemporally consistent fractional vegetation cover at different scales using spatiotemporal fusion and multiresolution tree methods. ISPRS J. Photogrammetry Remote Sens. 167, 214–229. https://doi.org/10.1016/j.isprsjprs.2020.07.006.

Wang, Q., Wang, L., Li, Z., Tong, X., Atkinson, P.M., 2021a. Spatial-spectral radial basis function-based interpolation for Landsat ETM+ SLC-off image gap filling. IEEE Trans. Geosci. Rem. Sens. 59, 7901–7917. https://doi.org/10.1109/TGRS.2020.3038878.

Wang, Q., Wang, L., Wei, C., Jin, Y., Li, Z., Tong, X., Atkinson, P.M., 2021b. Filling gaps in Landsat ETM+ SLC-off images with Sentinel-2 MSI images. Int. J. Appl. Earth Obs. Geoinf. 101, 102365. https://doi.org/10.1016/j.jag.2021.102365.

Weiss, D.J., Atkinson, P.M., Bhatt, S., Mappin, B., Hay, S.I., Gething, P.W., 2014. An effective approach for gap-filling continental scale remotely sensed time-series. ISPRS J. Photogrammetry Remote Sens. 98, 106–118. https://doi.org/10.1016/j.isprsjprs.2014.10.001.

Wu, M., Niu, Z., Wang, C., Wu, C., Wang, L., 2012. Use of MODIS and Landsat time series data to generate high-resolution temporal synthetic Landsat data using a spatial and temporal reflectance fusion Model. J. Appl. Remote Sens. 6 https://doi.org/10.1117/1.JRS.6.063507, 063507.

Zhang, L., Weng, Q., Shao, Z., 2017. An evaluation of monthly impervious surface dynamics by fusing Landsat and MODIS time series in the Pearl River Delta, China, from 2000 to 2015. Remote Sens. Environ. 201, 99–114. https://doi.org/10.1016/j.rse.2017.08.036.

Zhang, M., Yuan, N., Lin, H., Liu, Y., Zhang, H., 2022. Quantitative estimation of the factors impacting spatiotemporal variation in NPP in the Dongting Lake wetlands using Landsat time series data for the last two decades. Ecol. Indicat. 135, 108544. https://doi.org/10.1016/j.ecolind.2022.108544.

Zhao, R., Li, Y., Chen, J., Ma, M., Fan, L., Lu, W., 2021. Mapping a paddy rice area in a cloudy and rainy region using spatiotemporal data fusion and a phenology-based algorithm. Rem. Sens. 13, 4400. https://doi.org/10.3390/rs13214400.

Zhou, J., Chen, J., Chen, X., Zhu, X., Qiu, Y., Song, H., Rao, Y., Zhang, C., Cao, X., Cui, X., 2021. Sensitivity of six typical spatiotemporal fusion methods to different influential factors: a comparative study for a normalized difference vegetation index time series reconstruction. Remote Sens. Environ. 252, 112130. https://doi.org/10.1016/j.rse.2020.112130.

Zhu, X., Cai, F., Tian, J., Williams, T.K.-A., 2018a. Spatiotemporal fusion of multisource remote sensing data: literature survey, taxonomy, principles, applications, and future directions. Rem. Sens. 10, 527. https://doi.org/10.3390/rs10040527.

Zhu, X., Chen, J., Gao, F., Chen, X., Masek, J.G., 2010. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. Remote Sens. Environ. 114, 2610–2623. https://doi.org/10.1016/j.rse.2010.05.032.

Zhu, X., Gao, F., Liu, D., Chen, J., 2012a. A modified neighborhood similar pixel interpolator approach for removing thick clouds in Landsat images. IEEE Geosci. Rem. Sens. Lett. 9, 521–525. https://doi.org/10.1109/LGRS.2011.2173290.

Zhu, X., Helmer, E.H., 2018. An automatic method for screening clouds and cloud shadows in optical satellite image time series in cloudy regions. Remote Sens. Environ. 214, 135–153. https://doi.org/10.1016/j.rse.2018.05.024.

Zhu, X., Helmer, E.H., Chen, J., Liu, D., 2018b. An Automatic System for Reconstructing High-Quality Seasonal Landsat Time-Series. In: Weng, Q. (Ed.), Remote Sensing: Time Series Image Processing. Taylor and Francis Series in Imaging Science. CRC Press, Boca Raton, pp. 25–42.

Zhu, X., Helmer, E.H., Gao, F., Liu, D., Chen, J., Lefsky, M.A., 2016. A flexible spatiotemporal method for fusing satellite images with different resolutions. Remote Sens. Environ. 172, 165–177. https://doi.org/10.1016/j.rse.2015.11.016.

Zhu, X., Liu, D., 2014. Accurate mapping of forest types using dense seasonal Landsat time-series. ISPRS J. Photogrammetry Remote Sens. 96, 1–11. https://doi.org/10.1016/j.isprsjprs.2014.06.012.

Zhu, X., Liu, D., Chen, J., 2012b. A new geostatistical approach for filling gaps in Landsat ETM+ SLC-off images. Remote Sens. Environ. 124, 49–60. https://doi.org/10.1016/j.rse.2012.04.019.

Zhu, Z., Woodcock, C.E., 2014. Continuous change detection and classification of land cover using all available Landsat data. Remote Sens. Environ. 144, 152–171. https://doi.org/10.1016/j.rse.2014.01.011.

Zhu, Z., Woodcock, C.E., Holden, C., Yang, Z., 2015. Generating synthetic Landsat images based on all available Landsat data: predicting Landsat surface reflectance at any given time. Remote Sens. Environ. 162, 67–83. https://doi.org/10.1016/j.rse.2015.02.009.

Zhukov, B., Oertel, D., Lanzl, F., Reinhackel, G., 1999. Unmixing-based multisensor multiresolution image fusion. IEEE Trans. Geosci. Rem. Sens. 37, 1212–1226. https://doi.org/10.1109/36.763276.

Zurita-Milla, R., Clevers, J.G.P.W., Schaepman, M.E., 2008. Unmixing-based Landsat TM and MERIS FR data fusion. IEEE Geosci. Rem. Sens. Lett. 5, 453–457. https://doi.org/10.1109/LGRS.2008.919685.