Remote Sensing of Environment 252 (2021) 112130

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Sensitivity of six typical spatiotemporal fusion methods to different influential factors: A comparative study for a normalized difference vegetation index time series reconstruction

Junxiong Zhou[a,b], Jin Chen[a,b], Xuehong Chen[a,b,*], Xiaolin Zhu[c], Yuean Qiu[a,b], Huihui Song[d], Yunhan Rao[e], Chishan Zhang[a,b], Xin Cao[a,b], Xihong Cui[a,b]

[a] 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
[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] Jiangsu Key Laboratory of Big Data Analysis Technology, Nanjing University of Information Science and Technology, Nanjing 210044, China
[e] North Carolina State University, North Carolina Institute for Climate Studies, Asheville, NC 28805, USA

ARTICLE INFO

Keywords:
Spatiotemporal fusion
Normalized difference vegetation index (NDVI)
Geometric misregistration
Radiometric inconsistency
Spatial resolution ratio

ABSTRACT

Dozens of spatiotemporal fusion methods have been developed to reconstruct vegetation index time-series data with both high spatial resolution and frequent coverage for monitoring land surface dynamics. Although several studies comparing the different fusion methods have been conducted, selecting the suitable fusion methods is still challenging, as inevitable influential factors tend to be neglected. To address this problem, this study compared six typical spatiotemporal fusion methods, including the Unmixing-Based Data Fusion (UBDF), Linear Mixing Growth Model (LMGM), Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM), Fit-FC (regression model Fitting, spatial Filtering and residual Compensation), One Pair Dictionary-Learning method (OPDL), and Flexible Spatiotemporal DAta Fusion (FSDAF), based on simulation experiments and theoretical analysis considering three influential factors between sensors: geometric misregistration, radiometric inconsistency, and spatial resolution ratio. The results indicate that Fit-FC achieved the best performance with the strongest tolerance to geometric misregistration when radiometric inconsistency was negligible; thus, it is the first recommended algorithm for blending normalized difference vegetation index (NDVI) imagery. Instead, the FSDAF could generate the best results if radiometric inconsistency was non-negligible. These findings could help users determine the method that is appropriate for different remote sensing datasets, and provide guidelines for developers in the future development of novel methods.

1. Introduction

Time series of vegetation indices (e.g., Normalized Difference Vegetation Index, NDVI) produced by satellite sensors play a unique role as important data sources in various environmental applications, such as cropland mapping (Chang et al. 2007; Wardlow et al. 2007), vegetation phenology monitoring (Bradley et al. 2007; Cao et al. 2015; Zhang et al. 2003), and disturbance detection (Verbesselt et al. 2012). However, most of the sensors onboard the launched satellites cannot acquire data with both high spatial and temporal resolutions simultaneously, owing to hardware technology or budget limitations. For example, the data from sensors with dense temporal coverage usually have coarse spatial resolution (e.g., MODIS, hereafter referred to as coarse images), imposing restrictions on capturing enough spatial details in heterogeneous areas. On the other hand, the data from sensors with fine spatial resolution (e.g., Landsat TM or ETM+, hereafter referred to as fine images) have their drawbacks due to a long revisit cycle (e.g., 16 days), which limits their potential in time-series analyses. Consequently, various spatiotemporal fusion methods that combine the merits of two such types of data have been developed and are used to produce NDVI data with high spatial and temporal resolutions (Chen et al. 2018; Liao et al. 2017; Liu et al., 2019b; Maselli et al. 2019; Rao

---

* Corresponding author at: 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.
E-mail address: chenxuehong@bnu.edu.cn (X. Chen).

https://doi.org/10.1016/j.rse.2020.112130
Received 5 March 2020; Received in revised form 25 September 2020; Accepted 10 October 2020

et al. 2015). Furthermore, they have been successfully applied in various fields, such as crop growth monitoring (Gao et al. 2017), land cover classification (Chen et al. 2017; Jia et al. 2014), biomass estimation (Zhang et al. 2016), and disturbance detection (Hilker et al. 2009).

Zhu et al. (2018) grouped the published spatiotemporal fusion methods into the following five categories according to technique principles: unmixing-based, weight function-based, learning-based, Bayesian-based, and hybrid methods. Unmixing-based methods downscale coarse pixels to fine resolution based on the linear spectral mixing theory (Rao et al. 2015; Zhukov et al. 1999; Zurita-Milla et al. 2008). Weight function-based methods estimate the value of a target pixel by combining neighborhood pixel information with empirically designed weight functions of spectral similarity, spatial distance, or other related measurements (Gao et al. 2006; Wang and Atkinson 2018; Zhu et al. 2010). Learning-based methods, which use machine learning methods to model the relationship between coarse and fine images are relatively new (Huang and Song 2012; Liu et al. 2016; Song and Huang 2013; Song et al. 2018). Bayesian-based methods describe spatiotemporal fusions as a Maximum A Posterior problem based on the Bayesian framework (Huang et al. 2013; Liao et al. 2016; Shen et al. 2016). Hybrid methods attempt to integrate two or more methods mentioned above to improve the performance of the spatiotemporal fusion (Liu et al., 2019b; Quan et al. 2018; Zhu et al. 2016).

Although the technique principles are diverse, each developed method was claimed by its original study to have unique advantages in terms of prediction accuracy, computation efficiency, or input data requirements. However, as these studies used different datasets in their method comparison, it was difficult to reach a consensus on which method outperforms all the others. Thus, it is necessary to assess the applicability of these methods to different application scenarios. Accordingly, several cross-comparison studies have been conducted to explore the advantages and weaknesses of the different methods based on time-series data (Chen et al. 2015; Emelyanova et al. 2013; Liu et al., 2019a). In general, the performances of different fusion methods depend mainly on the sensitivity to spatial heterogeneity and temporal variations of the data used.

However, these comparison studies have neglected the influence of inevitable noise in real applications, including geometric misregistration and radiometric inconsistency. Despite the large efforts devoted to the intercalibration and geometric registration among the different sensors, the complete elimination of these intrinsic noises is still challenging (Chander et al., 2013a, 2013b; Claverie et al. 2018; Yan et al. 2016). A series of studies have noticed these issues and attempted to quantify the impact of geometric and radiometric inconsistencies on land cover change detection (Dai and Khorram 1998; Chen et al. 2014; Roy, 2000) and dynamic vegetation monitoring (Fan and Liu 2018; Skakun et al. 2018; Sulla-Menashe et al. 2016). Considering the potential impacts of these noises on the spatiotemporal fusion methods (Belgiu and Stein 2019; Zhu et al. 2018) and the lack of corresponding comparative research, it is still difficult for users to choose appropriate methods for their applications. A recent study has shown an encouraging desire to address these issues by quantifying the influence of geometric errors on fusion methods (Tang et al. 2020). However, only two algorithms were explored in this study, which is not sufficient for most users.

Recently, owing to advances in sensor technology and an increased number of launched satellites, spatiotemporal fusion methods are no longer limited to MODIS and Landsat images only, and could be expanded in recent studies (Kong et al. 2016; Kwan et al. 2018; Mizuochi et al. 2017; Li et al. 2017; Wang and Atkinson 2018) to images from other satellite sensors with different spatial resolutions (e.g., AMSR, ASTER, Sentinel-2/3, GF-1, Worldview, and Planet). However, this extension brings about a new problem, that is, whether the input images with different spatial resolution ratios will affect the performance of the different methods (Yokoya et al. 2017). Unfortunately, to our knowledge, no comparative studies have evaluated the performance of the spatiotemporal fusion methods based on data with different spatial resolution ratios of sensors.

To fill the gap from the previous comparison studies, we conducted comparison experiments and theoretical analyses on the spatiotemporal fusion of NDVI time-series data by considering various influential factors, including geometric misregistration, radiometric inconsistency, and spatial resolution ratio. Six typical spatiotemporal fusion methods, that required only one fine image and two coarse images as input, including the Unmixing-Based Data Fusion (UBDF), Linear Mixing Growth Model (LMGM), Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM), regression model Fitting, spatial Filtering and residual Compensation (Fit-FC), one pair dictionary-learning method (OPDL), and Flexible Spatiotemporal DAta Fusion (FSDAF), were selected for comparison. Six methods were selected in this study, considering their unique contributions in their own categories and the availability of source codes. Moreover, the performances were evaluated on time-series data instead of individual images, to better satisfy the application requirements. In general, the goal of this study is to explore the sensitivity of the six fusion methods to three influential factors and, thus, provide useful guidelines for method selection and future method design for users and developers.

2. Methods and datasets

2.1. Experiment design

To explore the sensitivity of spatiotemporal fusion methods for NDVI time-series reconstruction to various influencing factors, experiments were specifically designed in terms of geometric misregistration, radiometric inconsistency, and different spatial resolution ratios. Similar to previous studies (Gevaert and García-Haro 2015; Liu et al., 2019b; Zhu et al. 2016), the time series of cloud-free Landsat imagery and simulated coarse resolution imagery aggregated from Landsat data were used for spatiotemporal fusion experiments and validation. The standard experiment is based on ideal simulated data without any errors; it is used as the benchmark for the later simulation experiments. To explore the effect of various influencing factors, three additional fusion experiments were designed based on the simulated data with geometric misregistration, radiometric inconsistency, and different spatial resolution ratios. Additionally, fusion experiments based on actual Landsat and MODIS data were also conducted.

Considering the wide applicability of the NDVI time series, this study focuses on NDVI data fusion. Moreover, since surface reflectance data have also received increasing attention (Hermosilla et al. 2015; Xiao et al. 2016), similar fusion experiments were also conducted on reflectance data (green, red, near infrared bands) for comparison.

2.1.1. Standard fusion experiment based on ideal simulated data

This experiment followed the experimental settings of previous studies (Gevaert and García-Haro 2015; Zhu et al. 2016). Coarse images were simulated by the aggregation of Landsat images to avoid misregistration and radiometric inconsistency between the fine and coarse images. In the standard experiment, we aggregated 8 × 8 pixels for NDVI and 16 × 16 pixels for reflectance, corresponding to the MODIS NDVI at 240 m resolution, and MODIS reflectance data at 480 m resolution, respectively. The first fine image in the time series and the corresponding simulated coarse image were used as the base-paired image input for the fusion experiment. The other simulated coarse images were then downscaled to a fine spatial resolution using different fusion methods (Fig. 1. Schematic diagram of the standard experiment.).

2.1.2. Fusion experiment based on simulated data with geometric misregistration

The settings of this experiment are similar to the standard one,

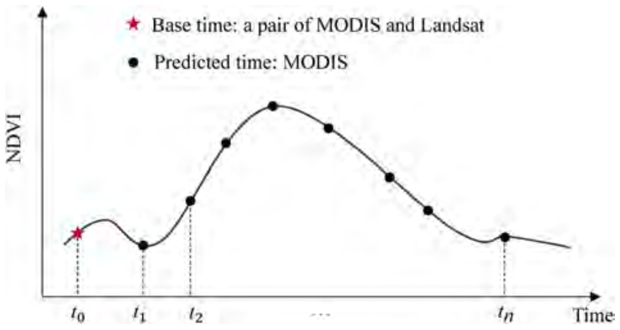

Fig. 1. Schematic diagram of the standard experiment.

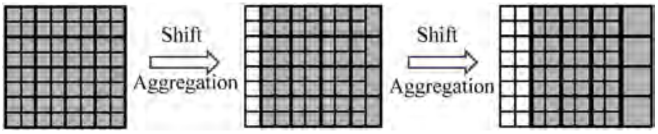

Fig. 2. Schematic diagram of the experiment with geometric misregistration.

except that the misregistration error was simulated when aggregating the Landsat images to the coarse images. Pixel shifting is one of the most serious consequences of geometric distortions. Therefore, similar to the previous study (Tang et al. 2020), the fine images were shifted by 2, 4, 6, or 8 pixels before aggregation; thus, different degrees of misregistration error were generated for the simulated coarse images (Fig. 2). This experiment compares the robustness of different methods to geometric errors. For reflectance experiments, the fine images were shifted 4, 8, 12, or 16 pixels before aggregation, considering that the resolution of coarse reflectance images was double that of the NDVI image.

2.1.3. Fusion experiment based on simulated data with radiometric inconsistency

The special experimental setting of this experiment, which is the only difference compared to the standard one, is that a linear stretch was conducted on the aggregated coarse image to simulate the radiometric inconsistency between fine and coarse sensors (Fig. 3):

$$C' = \alpha C + \beta \tag{1}$$

where C and C' are the ideal and stretched NDVI of simulated coarse pixels, respectively, and α and β are the linear stretch parameters. The parameters were referenced from an intercalibration study of vegetation indices derived from different sensors (Steven et al. 2003), in which the linear relationships of TM and MODIS, TM and AVHRR, ETM + and MODIS, POLDER and ASTR2, and QuickBird and ASTR2 were investigated (Table 1). Table 2 presents the linear relationships used in the reflectance experiments. With such a simulation, the sensitivity of different methods to radiometric inconsistency could be explored.

2.1.4. Fusion experiment based on simulated data with different spatial resolution ratios

To explore the applicability of the six fusion algorithms to various satellite products with different spatial resolutions, this experiment compares the sensitivity of these methods to different spatial resolution ratios of coarse and fine images. Coarse images are simulated at four levels of spatial resolution ratios (4, 8, 16, and 32, respectively, as shown in Fig. 4). Other experimental settings are similar to those in the standard one (i.e., without any geometric error or radiometric inconsistency).

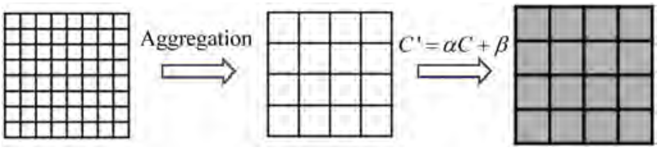

Fig. 3. Schematic diagram of the experiment with radiometric inconsistency.

Table 1
Coefficients of linear stretches for simulated radiometric inconsistency between two sensors in the NDVI experiments (Steven et al. 2003).

Satellite sensors	Slope (α)	Intercept (β)
TM-MODIS	1.002	−0.012
TM-AVHRR	1.106	−0.007
ETM + -MODIS	1.023	−0.013
POLDER-ASTR2	1.008	−0.110
QuickBird-ASTR2	0.928	−0.105

Table 2
Coefficients of linear stretches for simulated radiometric inconsistency between two sensors in the reflectance experiments.

Simulations	Slope (α)	Intercept (β)
Simu1	0.9	0.0
Simu2	1.1	0.0
Simu3	1.0	−0.05
Simu4	1.0	0.05

2.1.5. Fusion experiment based on actual MODIS images

Actual MODIS images were used for this fusion experiment. Additionally, simulated MODIS images with certain geometric errors and radiometric inconsistencies were also used for comparison. Although the geolocation accuracy of MODIS was 50 m at nadir (Wolfe et al. 2002), the large scan angle and reprojection and resampling procedures could further enlarge the geolocation error. Thus, it is reasonable to assume an averaged misregistration error size of approximately a half-pixel (120 m), corresponding to a four-pixel shifting of the fine images. The parameters of the linear stretches for the TM-MODIS (Table 1) were used to simulate the radiometric inconsistency. This experiment was conducted to illustrate how much of the fusion error of the results using the actual MODIS data could be accounted for by the results based on the simulated MODIS data with simulated geometric and radiometric errors.

2.1.6. Accuracy indices for evaluation

Two accuracy indices, the root mean square error (RMSE) and correlation coefficient (*r*), were used to evaluate the performance of the different fusion methods. The RMSE was calculated using all the pairs of predicted and true images throughout the time series. The correlation coefficient (*r*) was calculated between the predicted and true NDVI time series for each fine pixel. Then, an averaged *r* of the whole image was used to represent the overall accuracy of the predicted time-series data. Different aspects of the fusion results were assessed. The image-based RMSE evaluates the average pixel-wise prediction errors, which has drawn the attention of quantitative remote sensing studies. The coefficient *r* is the similarity between the predicted temporal profile and the true temporal profile, which will benefit dynamic monitoring research. With the above two indices, the overall performances on the time series, instead of the individual images, were evaluated for the different fusion methods under different experimental scenarios.

Additionally, to further explore the relationship between the fusion accuracy and the temporal variation of the input data, an absolute relative difference index (ADRI) was calculated to represent the temporal change between the base and predicted time.

$$\mathrm{ARDI} = |F_2 - F_1|/F_1 \tag{2}$$

where F_1 and F_2 denote the NDVI or reflectance of the fine images at the base and predicted time, respectively.

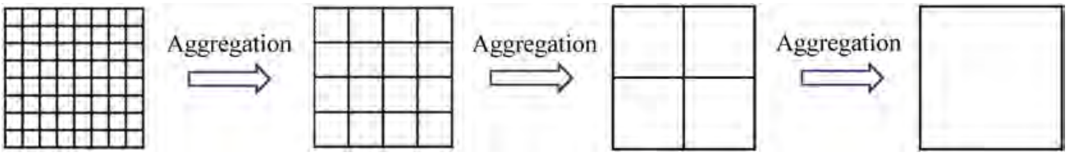

Fig. 4. Schematic diagram of the experiment with different spatial resolution ratios.

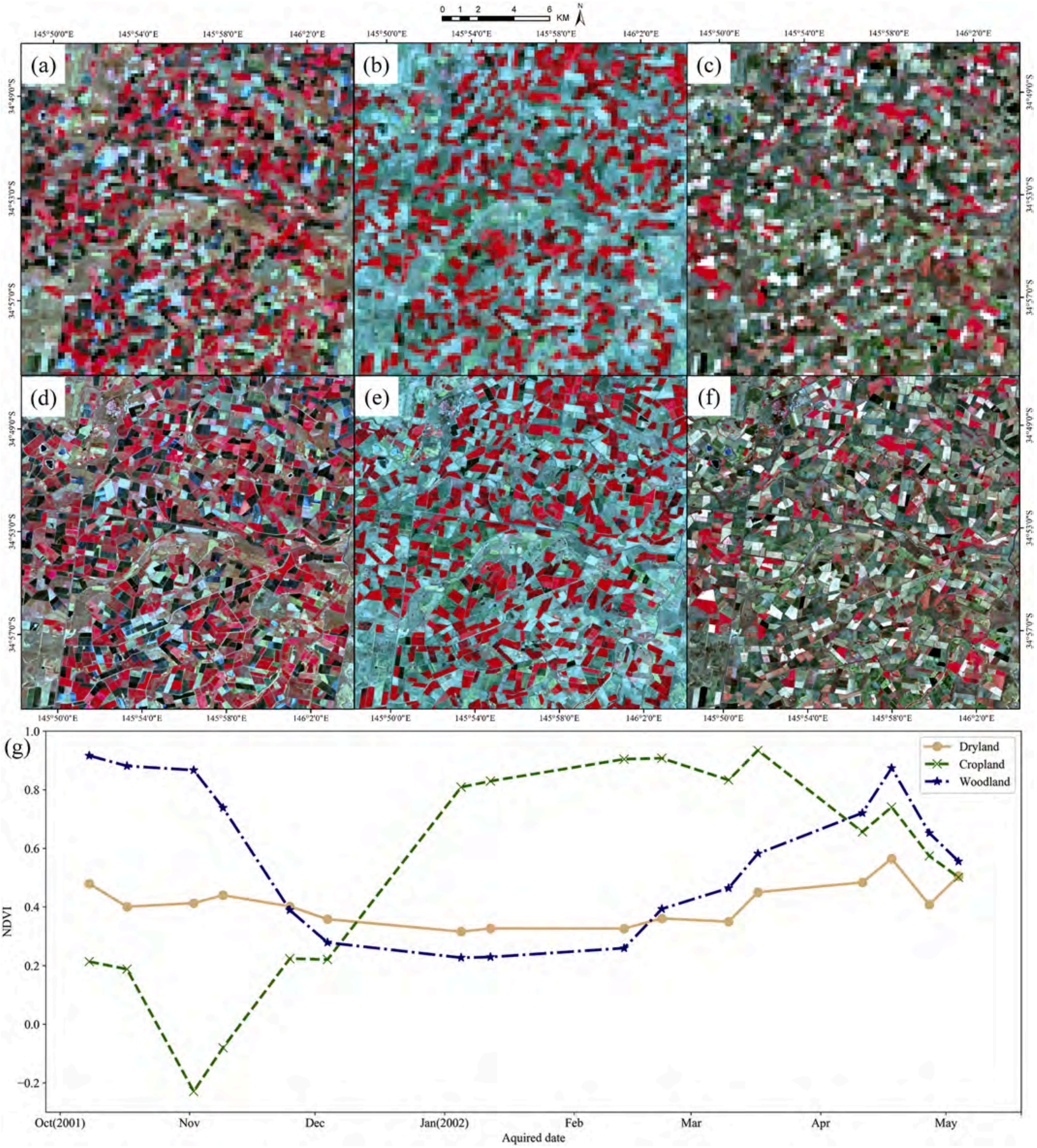

Fig. 5. Test data from the CIA site: 240 m simulated coarse images and corresponding 30 m fine images acquired on (a) and (d) November 9, 2001, (b) and (e) February 13, 2002, and (c) and (f) May 4, 2002; and (g) NDVI time series of three typical land covers. All images use NIR–red–green as RGB. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.2. Experimental datasets

For a unified comparison, the typical datasets from previous spatiotemporal fusion studies (Emelyanova et al. 2013; Li et al. 2020): the Coleambally Irrigation Area (CIA) in southern New South Wales (145.10°E, 34.05°S), Gwydir Catchment (GWY) in northern New South Wales (149.63°E, 29.77°S), and Tianjin area in northern China (117.20°E, 39.30°N), were used in this study. The CIA site was dominated by woodlands, cropland, and dryland cover types. A total of 16 cloud-free pairs of Landsat-7 ETM+ (800 × 800 pixels at 30 m spatial resolution) data were collected in this area from October 2001 to May 2002. As shown in Fig. 5(a) and (b), there are fragmented cropland and woodland parcels in this area, resulting in a heterogeneous landscape. Additionally, woodlands, croplands, and drylands show distinctive NDVI profiles during this period (Fig. 5(c)). The main purpose of using this dataset with high heterogeneity and complex NDVI seasonality is to compare the performance of the fusion methods in monitoring seasonal changes in fragmented cropland landscapes. The GWY site was dominated by winter crops and natural vegetation. A total of 14 cloud-free pairs of Landsat-5 TM (800 × 800 pixels at 30 m spatial resolution) data were collected in this area from April 2004 to April 2005. This site was relatively homogeneous, displaying relatively large parcels of crop fields and natural vegetation (Fig. 6(a) and (b)). However, a flood occurred in December 2004, leading to a sudden drop in the NDVI of the inundated areas (Fig. 6(c)). Thus, this dataset is employed to test the performance of fusion methods in capturing abrupt land cover changes. As for the third site, Tianjin, the main land cover types are impervious surfaces, winter crops, summer crops, and waterbodies. The more complex land cover types and small built-up patches resulted in a more heterogeneous urban landscape at this site. There were 11 cloud-free pairs of Landsat-8 OLI (800 × 800 pixels at 30 m spatial resolution) data collected at the Tianjin site. As shown in Fig. 7, each land cover had a unique NDVI temporal profile. The main purpose of this dataset is to test the accuracy of the fusion methods in detecting seasonal variations in urban landscapes. For all three sites, the true MODIS surface reflectance (MODIS Terra MOD09GQ collection 6, resampled to 240 m spatial resolution) acquired in the corresponding periods were also downloaded for comparison.

2.3. Spatiotemporal fusion methods

We selected 1–2 typical methods for each category mentioned by Zhu et al. (2018) for the comparison experiments, except for the Bayesian-based methods because of the lack of open-source code. To quantify the error propagation of the fusion results caused by misregistration and radiometric consistency between two sensors, key concepts and equations of each method were introduced here for the convenience of the later theoretical analysis in the discussion section. For simplification, the algorithms were reintroduced here based on a consistent denotation (Table 3).

2.3.1. UBDF

As an unmixing-based method, UBDF employs a constrained least square with a moving to unmix coarse images for appropriate results (Zurita-Milla et al. 2008). Based on the linear spectral mixing model, the NDVI at a coarse pixel is regarded as a linear combination of the NDVIs of its endmembers. Assuming that the fine pixels are pure enough to be endmembers, the NDVI at the coarse pixels (*x*,*y*) and the predicted time is

$$C_2(x,y) = \sum_{i=1}^{c} f_i(x,y) F_2^i(x,y) + \varepsilon(x,y) \tag{3}$$

where $F_2{}^i(x,y)$ is the NDVI of the *i*th land cover type in the coarse pixel (*x*,*y*), $f_i(x,y)$ is the fraction of the *i*th endmember in the coarse pixel, *c* is the number of endmembers, and ε is the residual error. The $f_i(x,y)$ is calculated based on the classification result of the fine images at t_1, as land cover is assumed to be unchanged in UBDF. With another assumption that endmembers are consistent in a moving window of coarse pixels ($\mathbb{M}$), $F_2{}^i(x,y)$ can be solved by the following equations with constrained corresponding to the mixing models in a moving window ($m \times m$ coarse pixels):

$$\begin{bmatrix} C_2(1,1) \\ \vdots \\ C_2(x,y) \\ \vdots \\ C_2(m,m) \end{bmatrix} = \begin{bmatrix} f_1(1,1) & f_2(1,1) & \cdots & f_c(1,1) \\ \vdots & \vdots & & \vdots \\ f_1(x,y) & f_2(x,y) & \cdots & f_c(x,y) \\ \vdots & \vdots & & \vdots \\ f_1(m,m) & f_2(m,m) & \cdots & f_c(m,m) \end{bmatrix} \begin{bmatrix} F_2^1 \\ F_2^2 \\ \vdots \\ F_2^c \end{bmatrix} \tag{4}$$

Also, Eq. (4) can be written in matrix form for convenience:

$$\mathbf{C}_2(\mathbb{M}) = \mathbf{f}(\mathbb{M})\mathbf{F}_2(\mathbb{M}) \tag{5}$$

Thus, $F_2{}^i$ can be estimated by the least-squares method:

$$\widehat{\mathbf{F}}_2(\mathbb{M}) = [\mathbf{f}^T(\mathbb{M})\mathbf{f}(\mathbb{M})]^{-1}\mathbf{f}^T(\mathbb{M})\mathbf{C}_2(\mathbb{M}) \tag{6}$$

Finally, the fine image at t_2 can be generated by assigning the estimated $F_2{}^i$ to the corresponding fine pixels based on the classification result of t_1.

2.3.2. LMGM

To further enhance the spatial details in the unmixing-based fusion results, LMGM makes use of F_1 (Rao et al. 2015). It assumes that the growth rate of the same land cover is constant over a short period of time. Therefore, LMGM estimates the growth rates of endmembers ($\Delta F = F_2 - F_1$) by unmixing the growth rate of coarse pixels ($\Delta C = C_2 - C_1$), as shown in Eq. (7):

$$\begin{bmatrix} \Delta C(1,1) \\ \vdots \\ \Delta C(x,y) \\ \vdots \\ \Delta C(m,m) \end{bmatrix} = \begin{bmatrix} f_1(1,1) & f_2(1,1) & \cdots & f_c(1,1) \\ \vdots & \vdots & & \vdots \\ f_1(x,y) & f_2(x,y) & \cdots & f_c(x,y) \\ \vdots & \vdots & & \vdots \\ f_1(m,m) & f_2(m,m) & \cdots & f_c(m,m) \end{bmatrix} \begin{bmatrix} \Delta F^1 \\ \Delta F^2 \\ \vdots \\ \Delta F^c \end{bmatrix} \tag{7}$$

Then, the LMGM calculates $\widehat{F}_2$ by adding the estimated growth rate of class *i* ($\Delta\widehat{F}^i$) to $F_{1:}$

$$\widehat{F}_2(x,y) = F_1(x,y) + \Delta\widehat{F}^i \tag{8}$$

2.3.3. STARFM

STARFM is the most typical and popular fusion method based on a weight function (Gao et al. 2006). It assumes that the systematic bias between two sensors does not change over time. STARFM first resamples the coarse images to the same spatial resolution as the fine image. Thus, $F_2(x, y)$ can be estimated as

$$\widehat{F}_2(x,y) = F_1(x,y) + \Delta C(x,y) \tag{9}$$

Considering the issues of mixed pixel and land cover change, the information of similar neighboring pixels is introduced for the final estimation of F_2:

$$\widehat{F}_2(x,y) = \sum_{i=1}^{n_s} W_i(F_1(x_i,y_i) + \Delta C(x_i,y_i)) \tag{10}$$

where n_s is the number of similar pixels in the moving window, and W_i is the weight of the *i*th similar pixel. The definition of spectral neighbor similar pixels is that they belong to the same class. The calculation of the weight W_i combines the spatial distance (D_i) and spectral difference between the coarse and fine images (S_i) (Gao et al. 2006; Gao et al. 2015):

$$D_i = \sqrt{(x_{w/2} - x_i)^2 + (y_{w/2} - y_i)^2} \tag{11}$$

$$S_i = |F_1(x_i,y_i) - C_1(x_i,y_i)| \tag{12}$$

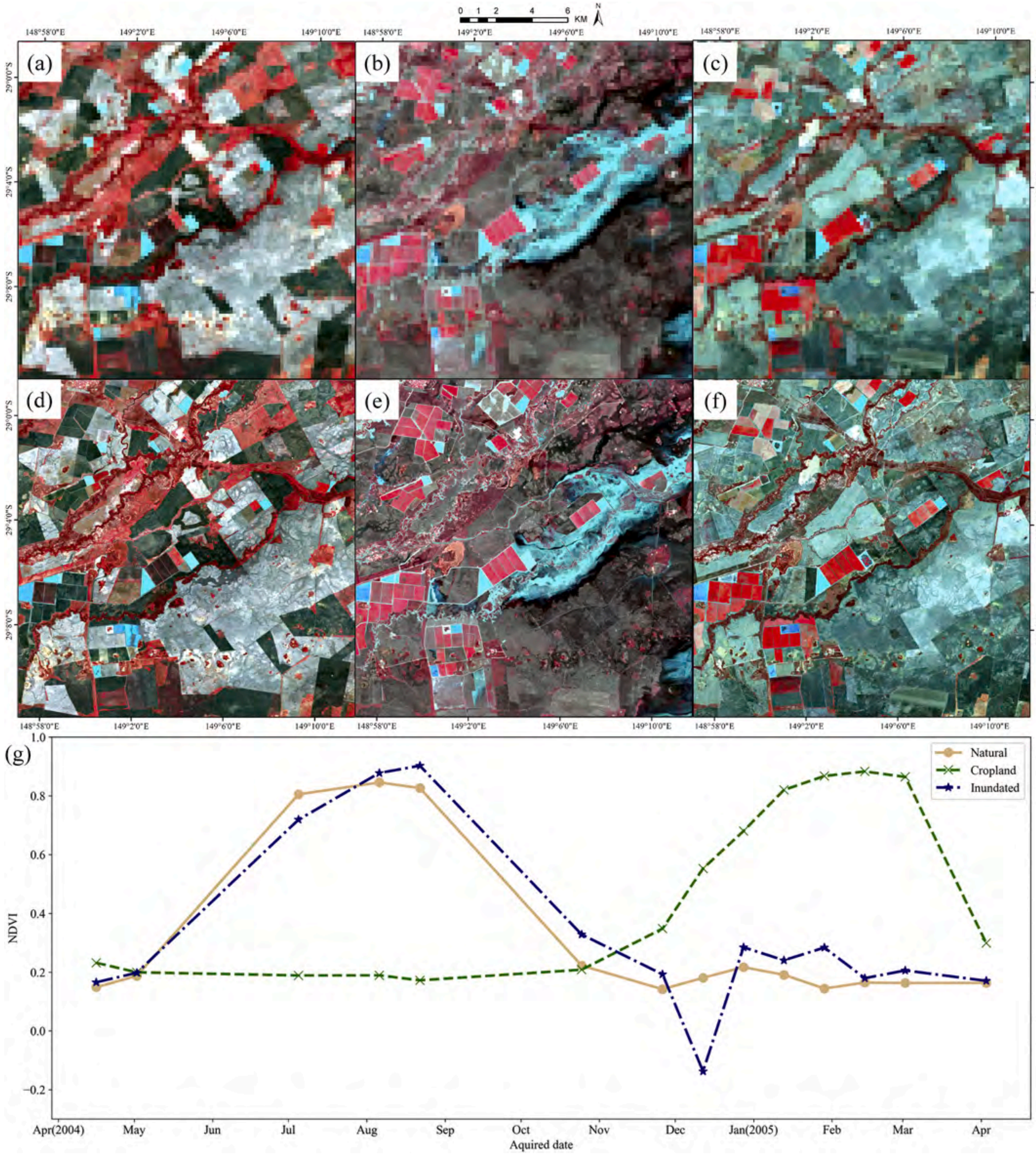

Fig. 6. Test data from the GWY site: 240 m simulated coarse images and corresponding 30 m fine images acquired on (a) and (d) May 2, 2004, (b) and (e) December 12, 2004, and (c) and (f) April 4, 2005; and (g) NDVI time series of three typical land covers. All images use NIR-red-green as RGB. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

where $(x_{w/2}, y_{w/2})$ and (x_i, y_i) are the central pixels of the moving window and candidate similar neighboring pixels, respectively. The spatial closer similar pixel with smaller spectral difference holds higher weight.

2.3.4. Fit-FC

To capture the temporal changes of the fine pixels accurately, Fit-FC introduces a linear regression model established from coarse images (Wang and Atkinson 2018). A local linear regression model is first established between C_2 and C_1 within a moving window $\mathbb{M}$:

$$\mathbf{C}_2(\mathbb{M}) = a \times \mathbf{C}_1(\mathbb{M}) + b + \mathbf{R}(\mathbb{M}) \tag{13}$$

where $\mathbf{R}(\mathbb{M})$ are the coarse residuals in the moving window, and a and b are the regression coefficients. Then, the regression coefficients are

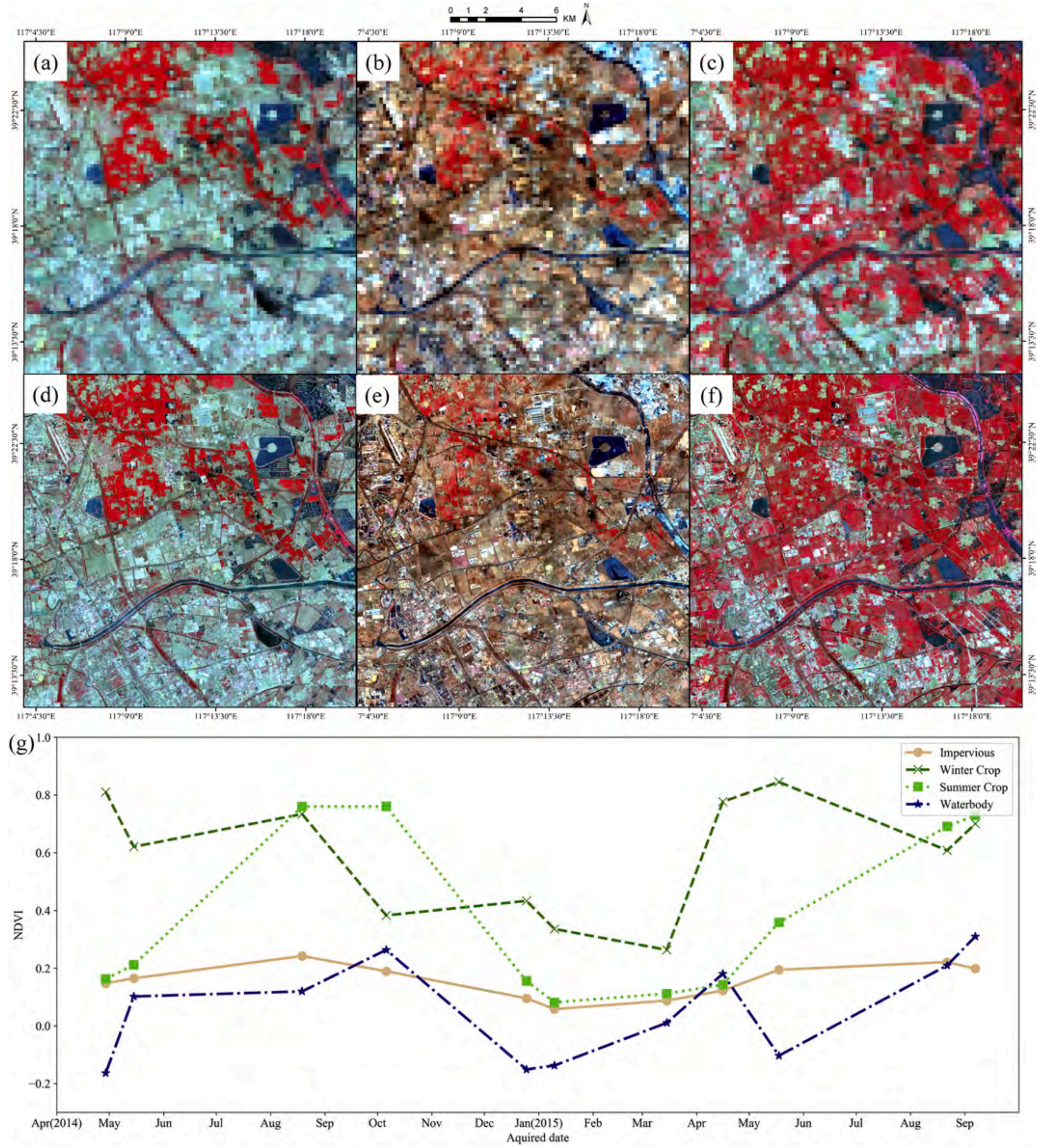

Fig. 7. Test data from the Tianjin site: 240 m simulated coarse images and corresponding 30 m fine images acquired on (a) and (d) April 29, 2014, (b) and (e) December 25, 2014, and (c) and (f) August 2, 2015; and (g) NDVI time series of three typical land covers. All images use NIR-red-green as RGB. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

applied to the fine pixels within a moving window corresponding to the coarse moving window (𝕄) for the RM (i.e., regression model) prediction. Then, a spatial filtering by weighting similar neighboring pixels with spatial distance is adopted to address the problem of blocky artifacts while considering the residuals of the regression model:

$$\widehat{F_2}(x,y) = \sum_{i=1}^{n_s} W_i(a \times F_1(x_i,y_i) + b + r(x_i,y_i)) \tag{14}$$

where $r(x_i,y_i)$ is the residual at the fine pixel (x_i,y_i), which is resampled from $\mathbf{R}(\mathbb{M})$ by bicubic interpolation.

Table 3
Common variables used in different spatiotemporal fusion methods.

Symbol	Meaning
(x, y)	Geolocation of specific pixel
t1	Base time
t2	Predicted time
C1	The input coarse image at t1
C2	The input coarse image at t2
F1	The input fine image at t1
F2	The output image at t2
$\mathbb{M}$	The moving window of pixel (x, y)

2.3.5. *OPDL*

Dictionary-learning based methods reconstruct images with an overcomplete dictionary and the corresponding coefficients of sparse representation (Huang and Song 2012). Song and Huang (2013) proposed the dictionary-based learning method OPDL, which requires only one image pair. The key idea of OPDL is that coarse and fine images acquired at the same location share the same sparse representation coefficients, and the overcomplete dictionary trained from images acquired at base time should be time-invariant. Therefore, C_1 and F_1 provide the dictionary, and C_2 provides the corresponding coefficients to generate the transition image T_2. With the same process, T_1 can also be produced. Finally, a high-pass modulation is introduced to transfer the temporal change from transition images to F_1 for prediction:

$$\widehat{F}_2(x,y)=\frac{T_2(x,y)}{T_1(x,y)}F_1(x,y) \tag{15}$$

Owing to the large spatial resolution difference between the fine and coarse images, OPDL is implemented in a two-layer framework (Song and Huang 2013). The first layer produces an image with an intermediate resolution between the coarse and fine images. Subsequently, the second layer generates final results using the image synthesized by the first layer.

2.3.6. *FSDAF*

FSDAF (Zhu et al. 2016) is a hybrid method that combines unmixing, spatial interpolation, and similar neighboring pixel smoothing for robust fusion results. First, similar to LMGM, FSDAF estimates the temporal change of a fine pixel (ΔF^{tp}) by an unmixing-based method to produce the temporal prediction (${F_2}^{tp}$), except that the unmixing procedure is conducted in the whole image instead of a moving window. Then, with the TPS interpolation (Dubrule 1984), the spatial prediction (${F_2}^{sp}$) of F_2 can be generated. The residuals between the sum of ΔF and ΔC are considered in the FSDAF:

$$R(x,y)=\Delta C(x,y)-\frac{1}{n}\left[\sum_{i=1}^{n}F_2^{tp}(x_i,y_i)-\sum_{i=1}^{n}F_1(x_i,y_i)\right] \tag{16}$$

where $R(x, y)$ is the residual in the coarse pixel at location (x, y), n is the number of fine pixels inside a coarse pixel, and the fine pixel at location (x_i, y_i) is inside the coarse pixel at location (x, y). In a homogenous area, the spatial prediction performs well, which is applied to calculate a new residual:

$$R_{ho}(x,y)=F_2^{sp}(x,y)-F_2^{tp}(x,y) \tag{17}$$

Thus, a weighted function (w_h) integrates the two residuals (i.e., R_{ho} and R) using a homogeneity index for residual compensation. The final prediction of FSDAF can be expressed as

$$\widehat{F}_2(x,y)=F_1(x,y)+\sum_{i=1}^{n_s}W_i\left(\Delta F^{tp}(x_i,y_i)+n\times R(x_i,y_i)\times w_h(x_i,y_i)\right) \tag{18}$$

where W_i is the weight of a similar pixel as that of the Fit-FC.

Table 4
Key parameters of five fusion methods (c is the class number, m is the moving window size, n_s is the number of similar neighboring pixels, m_s is the moving window size for searching similar neighboring pixels, and R is the spatial resolution ratio of the coarse and fine images).

	c			m	n_s	m_s
	CIA	GWY	Tianjin			
UBDF	6	5	7	5 × 5	N/A	N/A
LMGM	6	5	7	5 × 5	N/A	N/A
STARFM	6	5	7	N/A	N/A	1.5 × R + 1
Fit-FC	N/A	N/A	N/A	3 × 3	1.5 × R	1.5 × R + 1
FSDAF	6	5	7	N/A	1.5 × R	1.5 × R + 1

2.4. *Parameter settings of the six spatiotemporal fusion methods*

Based on previous studies (Gao et al. 2006; Song and Huang 2013; Rao et al. 2015; Wang and Atkinson, 2018; Zhu et al. 2016; Zurita-Milla et al. 2008), the parameters of the six spatiotemporal fusion methods were carefully tuned for different experimental sites and different resolution ratios. Table 4 shows the key parameters of UBDF, LMGM, STARFM, Fit-FC, and FSDAF. Note that we set the same values for the parameters with similar functions in different methods, to achieve a fair comparison (i.e., similar neighboring pixel smoothing in STARFM, Fit-FC, and FSDAF). The key parameters of the OPDL are shown separately in Table 5 as they are very different from those of the other five fusion methods. The patch size of the dictionary representation in the two layers was consistently set as 3 and 4 for all of the experimental sites and resolution ratios. The fusion methods were implemented with their authors' source codes except that UBDF was rewritten by ourselves. Among them, The source codes of STARFM, FSDAF, and Fit-FC are publicly available (https://www.ars.usda.gov/research/software/download/?softwareid=432; https://github.com/qunmingwang/Fit-FC; https://xiaolinzhu.weebly.com/open-source-code.html).

3. Results

3.1. *Standard comparison*

The performances of the six methods at the three sites were evaluated using the ideal simulation data. Table 6 shows the averaged RMSE and r for each method. In general, Fit-FC performed best, followed by FSDAF. Among the other four methods, STARFM performed better than UBDF, LMGM, and OPDL. As shown in Fig. 8, all the methods performed worse when the ADRI increased, which is similar to previous study (Xie et al. 2018), while Fit-FC and FSDAF always generated better results than the other four methods for all the images in the time series.

3.2. *Geometric misregistration*

The performances of the six methods with the simulated misregistration errors between the coarse and fine images are presented in this section. The extent of misregistration was quantitatively measured as the shifting pixel distance. Based on a visual comparison of the NDVI fusion results (Fig. 9), there are few distortions of the results fused by Fit-FC under eight pixel shifting. However, the results fused by the other five fusion methods are obviously different from the reference results without geometric errors. When evaluated by quantitative indices (Fig. 10), it is apparent that the accuracy of each method generally decreases as the shifting distance increases. Fit-FC is the most robust method for misregistration, followed by UBDF, as their evaluation index values vary the slowest. The other four methods, LMGM, STARFM, OPDL, and FSDAF, are more sensitive to the geometric error, as shown in Fig. 10 (a) and (b), where they all have sheer accuracy

Table 5
Key parameters of the OPDL method (resolution ratio is equal to the product of the scale factors of the two layers).

Dictionary size (Layer 1, Layer 2)		Resolution ratio (scale factors of the two layers)			
		4 (2 × 2)	8 (2 × 4)	16 (4 × 4)	32 (4 × 8)
Experimental site	CIA	(1500,1500)	(700,1500)	(700,1500)	(50,1500)
	GWY	(900,1500)	(200,1500)	(200,1500)	(100,1500)
	Tianjin	(600,1500)	(200,1500)	(200,1500)	(50,1200)

Table 6
Standard comparison evaluated by averaged RMSE and *r* at the three sites.

		UBDF	LMGM	STARFM	Fit-FC	OPDL	FSDAF
CIA	RMSE	0.1533	0.1816	0.1048	0.0816	0.131	0.1006
	r	0.7606	0.7717	0.8756	0.8979	0.8555	0.8758
GWY	RMSE	0.1124	0.1133	0.0729	0.0643	0.0718	0.0669
	r	0.8681	0.8726	0.9207	0.9226	0.9072	0.9175
Tianjin	RMSE	0.1346	0.1296	0.0919	0.0788	0.0843	0.0797
	r	0.8769	0.8749	0.9379	0.9409	0.9329	0.9385

drops along with an increase in the shifting distance. The results of the reflectance images (Fig. A1, Fig. A2, and Fig. A3) are similar to those of NDVI.

3.3. Radiometric inconsistency

Fig. 11 shows the robustness of different methods to the radiometric inconsistency between two sensors (i.e., the linear stretches of QuickBird-ASTR2). There are significant distortions in the results fused by UBDF, STARFM, and Fit-FC. The results evaluated by the quantitative indices are in good agreement with those of the visual comparison. When the fusion methods were applied to the datasets with small radiometric inconsistencies (Table 1), such as TM-MODIS, TM-AVHRR, and ETM + -MODIS, all produced accurate results. However, when there were larger radiometric inconsistencies, such as POLDER-ASTR2, QuickBird-ASTR2, UBDF, STARFM, and Fit-FC showed larger errors than the other methods. In contrast, LMGM, OPDL, and FSDAF are more robust to radiometric inconsistencies between the two sensors. As for the reflectance images results (Fig. A4, Fig. A5, and Fig. A6), the sensitivity of the fusion methods to radiometric inconsistency is consistent with that of the NDVI. Especially, Fit-FC and STARFM were very sensitive to radiometric bias for green and red bands due to the relatively small reflectance values for these two bands.

3.4. Spatial resolution ratio

Fig. 13 presents the accuracies of the six fusion methods in the scenarios of different spatial resolution ratios between the coarse and fine images. In general, all the methods perform worse when the spatial resolution ratio increases. Among these methods, OPDL is the most sensitive to the spatial resolution ratio. The accuracy of the OPDL fusion results decreases the fastest as the spatial resolution ratio increases (Fig. 14). STARFM is also highly sensitive to the spatial resolution ratio, especially in heterogeneous sites such as CIA. In contrast, UBDF, LMGM, FSDAF, and Fit-FC are somewhat less sensitive to the spatial resolution ratio. The results of reflectance images are similar to those of the NDVI (Fig. A7, Fig. A8, and Fig. A9).

3.5. Actual MODIS data

The performances of the six fusion methods based on the actual MODIS images were compared with those based on the simulated MODIS images (Fig. 15). Generally, the former results were worse than those based on the simulated images with only partial errors, and

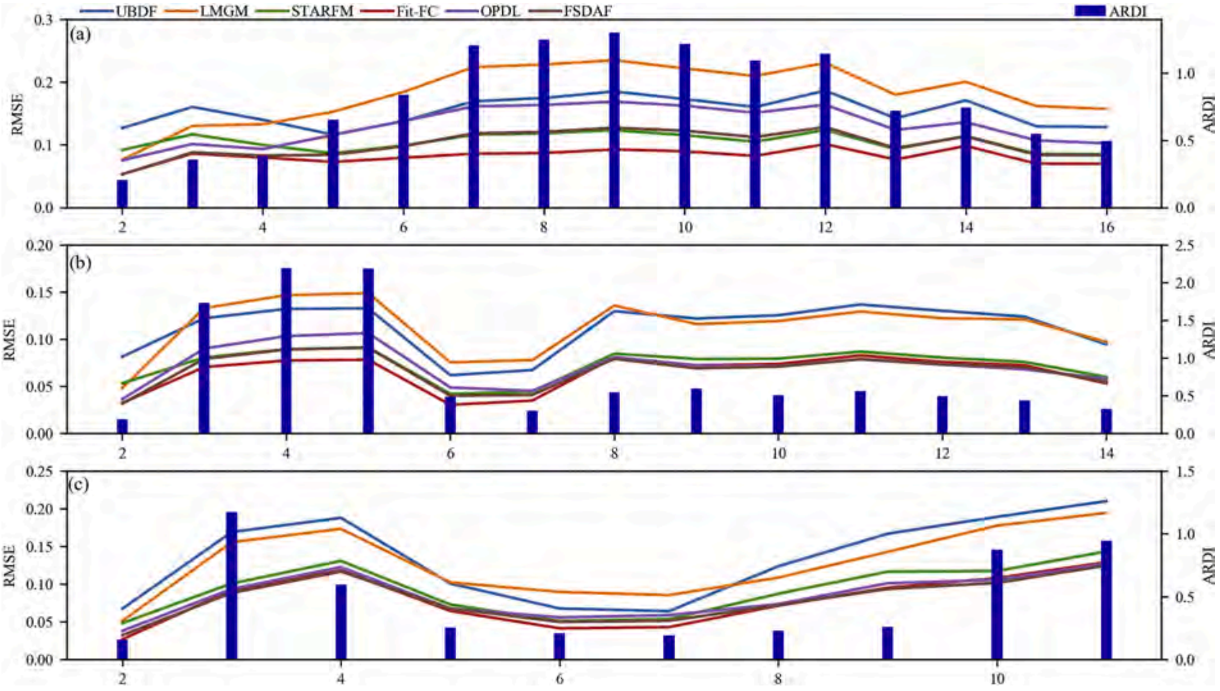

Fig. 8. The relationship between the prediction accuracy and the temporal variation at the three sites: (a) CIA; (b) GWY; (c) Tianjin. Image number is the number of the predicted image in the image time series.

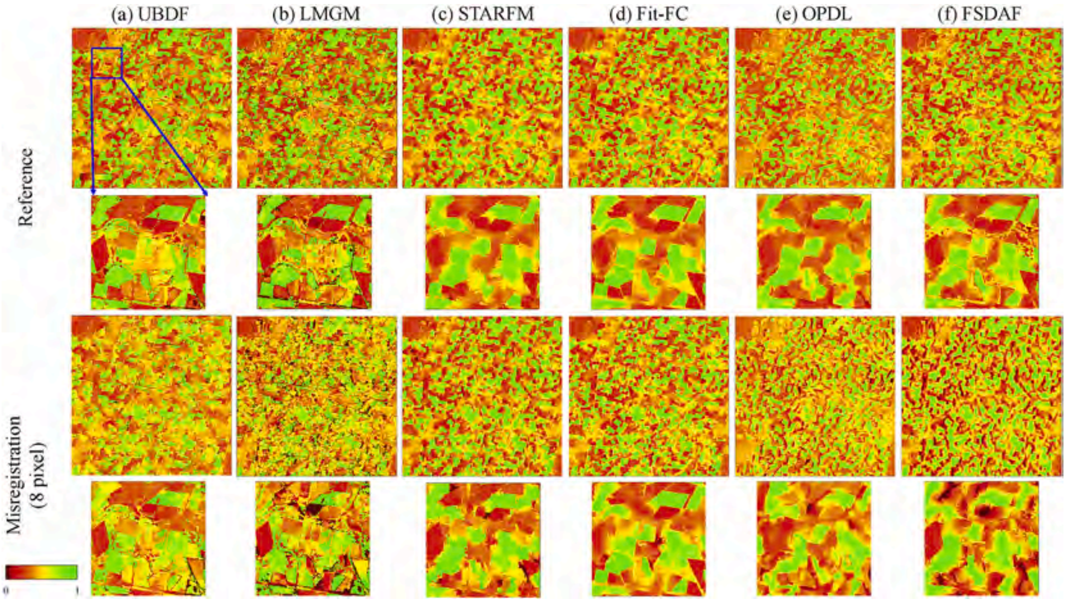

Fig. 9. Based on the fusion results of the CIA site on February 13, 2002, the visual comparison of the NDVI results without geometric errors (e.g., reference) and with misregistration (eight pixel shifting) by six methods: (a) UBDF; (b) LMGM; (c) STARFM; (d) Fit-FC; (e) OPDL; and (f) FSDAF.

further worse than those based on ideal images without any errors. This indicates that the simulated geometric and radiometric errors, to some extent, account for the performance differences between the experiments based on the ideal simulated MODIS data and those based on the actual MODIS data. For the experiment based on actual MODIS data, Fit-FC performed best at the CIA and GWY sites, whereas OPDL performed best at the Tianjin site. FSDAF ranked second for the GWY and Tianjin sites, whereas UBDF ranked second for the CIA site. These results imply the complexity of fusion experiments based on actual data. However, in general, Fit-FC and FSDAF are still worth recommending, considering their relative stability.

4. Discussions

To further explore the sensitivity of the six fusion methods to various factors, theoretical derivations were conducted to analyze the geometric and radiometric error propagation from the input data to the results. As for the spatial resolution ratio, its influence was similar to the influence of the spatial heterogeneity of the input data.

For comparison convenience, the different fusion methods (except OPDL) were grouped into three types: origin weighting (Eq. (19)), increment weighting (Eq. (20)), and regression weighting methods (Eq. (21)):

$$\widehat{F_2}(x,y) = \sum_i w_i C_{2i} \tag{19}$$

$$\widehat{F_2}(x,y) = F_1(x,y) + \sum_i w_i (C_{2i} - C_{1i}) \tag{20}$$

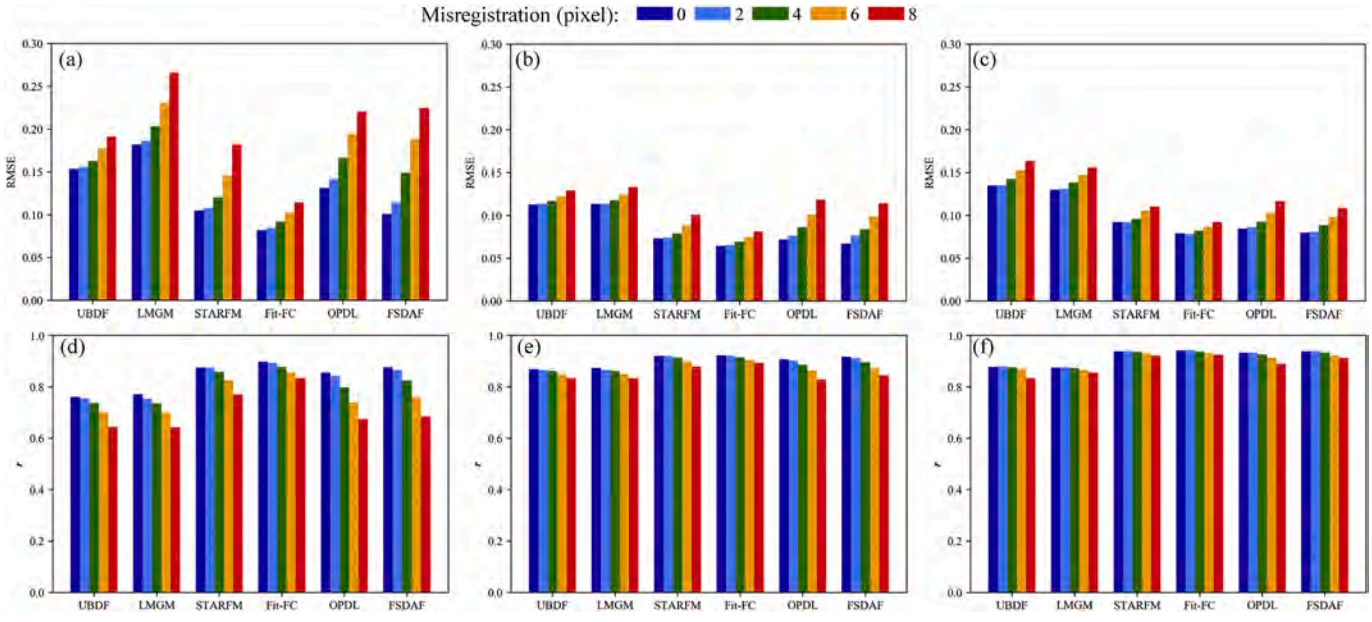

Fig. 10. Quantitative comparison of the NDVI fusion results under different levels of geometric errors from 0 to 8 (misregistration pixels): (a) RMSE in CIA; (b) RMSE in GWY; (c) RMSE in Tianjin; (d) *r* in CIA; (e) *r* in GWY; and (f) *r* in Tianjin.

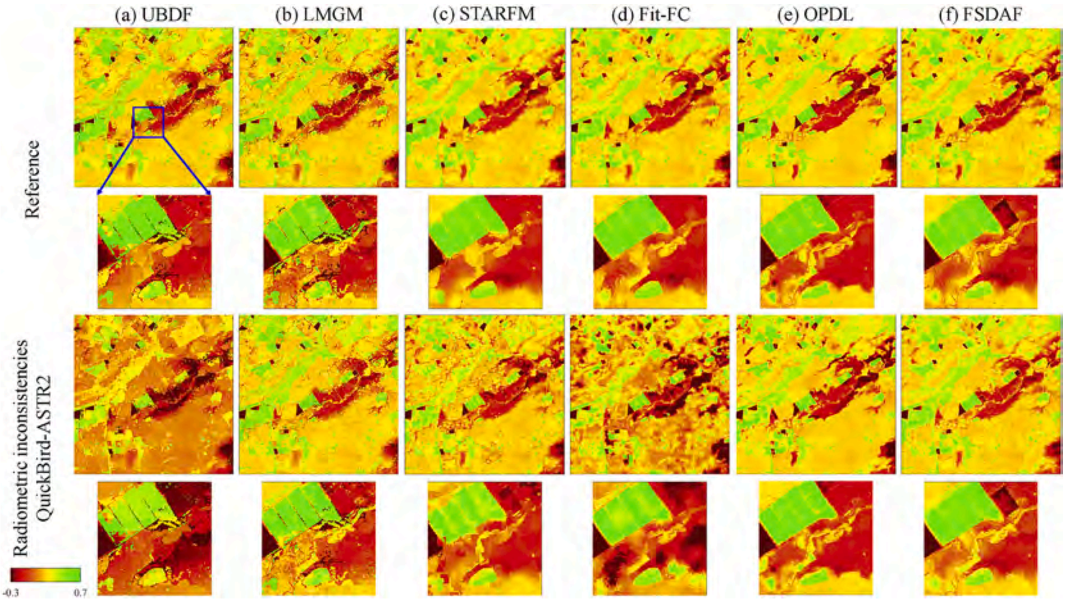

Fig. 11. Based on the fusion results of the GWY site on December 12, 2004, the visual comparison of the NDVI results without any errors (e.g., reference) and with radiometric inconsistency by six methods: (a) UBDF; (b) LMGM; (c) STARFM; (d) Fit-FC; (e) OPDL; and (f) FSDAF.

$$\widehat{F}_2(x,y)=\sum_i w_i(a\times F_{1i}+b+r_i) \tag{21}$$

where F_{1i}, C_{1i}, C_{2i}, and r_i are denoted as $F_1(X_i,Y_i)$, respectively, for simplification. UBDF is a typical origin weighting method. As shown in Eq. (19), the fused result is calculated by weighting different coarse pixels acquired at t_2; the w_i is calculated by $[\mathbf{f}^T(\mathbb{M})\mathbf{f}(\mathbb{M})]^{-1}\mathbf{f}^T(\mathbb{M})$ for UBDF. LMGM, STARFM, and FSDAF belong to increment weighting methods. As shown in Eq. (20), the fused result is calculated by weighting the temporal increments from t_1 to t_2 of different coarse pixels; the w_i is calculated in different ways for different algorithms. Fit-FC is a novel developed regression weighting method. As shown in Eq. (21), the fused result is calculated by weighing the linear transformation of fine pixels acquired at t_1; w_i is calculated based on a similar pixel smoothing strategy. For a convenient theoretical analysis, Eq. (22) was further simplified by replacing r_i with $C_{2i}-aC_{1i}-b$:

$$\begin{aligned}\widehat{F}_2(x,y)&=\sum_i w_i(a\times F_{1i}+b+C_{2i}-aC_{1i}-b)\\&=\sum_i w_i(a(F_{1i}-C_{1i})+C_{2i})\end{aligned}. \tag{22}$$

4.1. Propagation of geometric errors to fusion results

As the fine images were considered as reference, the NDVI error induced by geometric errors could be expressed only in coarse images. Although the geometric error is a kind of systematic error, the induced NDVI error is random. Thus, the NDVI error on coarse pixels induced at t_1 and t_2 are assumed to be random variables δC_1 and δC_2. Subsequently, although w_i is calculated in different ways by the five fusion methods, it is mainly determined by the information of the fine pixels that are not affected by geometric errors. Thus, w_i could be considered as a constant in the error propagation procedure. The fusion errors ($\delta\widehat{F}^{G}{}_{2}$) induced by geometric misregistration can be estimated based on the error propagation equation. For UBDF, the fusion uncertainty of UBDF induced by geometric errors can be derived as

$$std(\delta F_2^G(x,y))=\sqrt{\sum_i w_i^2 var(\delta C_{2i})} \tag{23}$$

where *std* and *var* are the standard deviation and variance, respectively. Similarly, the standard deviation of fusion errors of LMGM, STARFM, and FSDAF can be derived as

$$\begin{aligned}std(\delta F_2^G(x,y))&=\sqrt{\sum_i var(w_i\delta(C_{2i}-C_{1i}))}\\&=\sqrt{\sum_i w_i^2 var(\delta C_{2i})+var(\delta(C_{1i}))-2\,cov(\delta C_{2i},\delta C_{1i})}\end{aligned} \tag{24}$$

where *cov* is the covariance. If δC_1 and δC_2 are independent (i.e., the temporal change between t_1 and t_2 is significant), the term *cov* $(\delta C_{1i},\delta C_{2i})$ approaches zero. Thus, $var(\delta(C_{2i}-C_{1i}))$ is larger than $var(\delta C_{2i})$ because of error accumulation, which is also confirmed in the simulated data in most cases (Fig. 16). Therefore, LMGM, STARFM, and FSDAF are more sensitive to geometric errors than UBDF in general. The standard deviation of the fusion error of Fit-FC can also be derived as

$$\begin{aligned}std(\delta F_2^G(x,y))&=\sqrt{\sum_i var(w_i\delta(C_{2i}-aC_{1i}))}\\&=\sqrt{\sum_i w_i^2 var(\delta C_{2i})+var(\delta(aC_{1i}))-2\,cov(\delta C_{2i},\delta(aC_{1i}))}\end{aligned}. \tag{25}$$

As a is the regression coefficient between C_1 and C_2, $var(\delta(aC_{1i}))$is strongly correlated with δC_2. Thus, $var(\delta(C_{2i}-aC_{1i}))$ is smaller than $var(\delta C_{2i})$ because of error compensation, which is also shown in Fig. 16. Therefore, Fit-FC is the most robust method for geometric error.

Analysis of the error propagation of OPDL is difficult because of the nonlinear optimization in the dictionary-learning procedure. The sensitivity to the geometric error could depend on different learned features; thus, it varies case by case.

4.2. Propagation of radiometric error to fused result

As the radiometric inconsistency is usually a systematic error, the linear stretch was used to express the radiometric inconsistency. Thus,

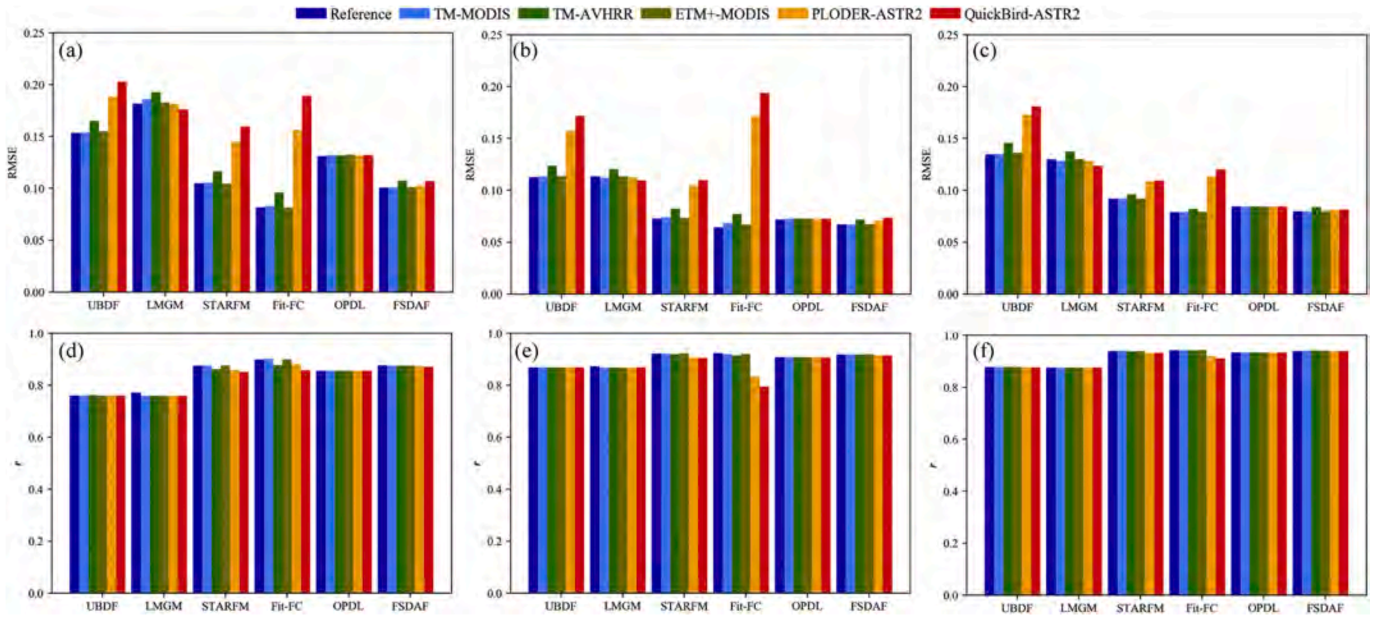

Fig. 12. Quantitative comparison of the NDVI fusion results under different levels of radiometric inconsistencies: (a) RMSE in CIA; (b) RMSE in GWY; (c) RMSE in Tianjin; (d) r in CIA; (e) r in GWY; and (f) r in Tianjin. Reference means that there is no radiometric inconsistency.

the fusion error of the UBDF induced by the radiometric inconsistency can be derived as

$$\begin{aligned}\Delta\widehat{F}^{R}_{2}(x,y) &= \sum_i w_i(\alpha C_{2i}+\beta)-\sum_i w_i C_{2i}\\ &= \sum_i w_i((\alpha-1)C_{2i}+\beta)\end{aligned} \tag{26}$$

where α and β are the coefficients for simulating radiometric inconsistency (i.e., the slope and intercept in Table 1). The fusion error of the UBDF induced by the radiometric inconsistency depends linearly on the two stretching parameters.

For STARFM, LMGM, and FSDAF, the fusion error induced by the radiometric inconsistency can be derived as

$$\begin{aligned}\Delta\widehat{F}^{R}_{2}(x,y) &= \sum_i w_i[(\alpha C_{2i}+\beta)-(\alpha C_{1i}+\beta)]-\sum_i w_i(C_{2i}-C_{1i})\\ &= \sum_i w_i(\alpha-1)(C_{2i}-C_{1i})\end{aligned} \tag{27}$$

Therefore, the intercept term (β) is removed in terms of ΔC. Theoretically, these three methods are less sensitive to radiometric inconsistency compared to UBDF. However, STARFM shows a high sensitivity to radiometric inconsistencies in the experiments (Fig. 11 and Fig. 12), which is inconsistent with the above theoretical analysis. This is because the weight (w_i) calculation of the similar pixel smoothing in STARFM includes a term of absolute NDVI difference between coarse and fine pixels (Eq. (12)), which is sensitive to radiometric inconsistency. If the weight calculation in the original STARFM is modified as that in Fit-FC, the modified STARFM will also be as robust to radiometric inconsistency as in the theoretical analysis (Fig. 17).

Similarly, for Fit-FC, the fusion error induced by the radiometric inconsistency can be expressed as

$$\begin{aligned}&\Delta\widehat{F}^{R}_{2}(x,y)\\ &= \sum_i w_i(a(F_{1i}-\alpha C_{1i}-\beta)+\alpha C_{2i}+\beta)-\sum_i w_i(a(F_{1i}-C_{1i})+C_{2i})\\ &= \sum_i W_i((\alpha-1)(C_{2i}-aC_{1i})+(1-a)\beta)\end{aligned} \tag{28}$$

Compared with the second group (LMGM, STARFM, and FSDAF), Fit-FC is more sensitive to radiometric inconsistency because α and β both influence the fusion result. Subsequently, OPDL is robust to radiometric inconsistency because it employs a linear regression model for the intercalibration of the coarse and fine images.

4.3. Influence of spatial resolution ratio on spatiotemporal fusion

The spatial resolution ratio of the sensors determines the information gap between the coarse and fine images acquired at the same time. In other words, as the spatial resolution ratio increases, coarse pixels contain more fine pixels and, thus, become more mixed; this is a similar effect as the increase in spatial heterogeneity. Thus, methods that perform relatively better in heterogeneous images should also be less sensitive to the spatial resolution ratio. As the unmixing module employed in fusion methods can better capture the spatial heterogeneity, UBDF, LMGM, and FSDAF, which employ the unmixing module, are less sensitive to the spatial resolution ratio than STARFM and OPDL. Fit-FC is also relatively less sensitive to the spatial resolution ratio, although the unmixing module is not employed in this method. This is because only two land cover types (i.e., vegetation and non-vegetation) need to be considered in the NDVI fusion; therefore, the linear regression model in Fit-FC with two degrees of freedom (i.e., two coefficients a and b) plays a similar unmixing role, which is adequate for capturing the temporal changes of the two land cover types. Furthermore, it implies that Fit-FC is particularly more suitable for the spatiotemporal fusion of the NDVI data than of the reflectance data.

4.4. Method selection and guidance for future research

The above comparison and analyses can guide the selection of suitable spatiotemporal fusion methods in applications. In addition to the influential factors of geometric misregistration, radiometric inconsistency, and spatial resolution ratio, the performances of the fusion methods mainly depend on the spatiotemporal variations of the input datasets. The selection of a suitable method should, therefore, consider the influence extent of all the factors, and balance the pros and cons of each method according to the characteristics of their data and

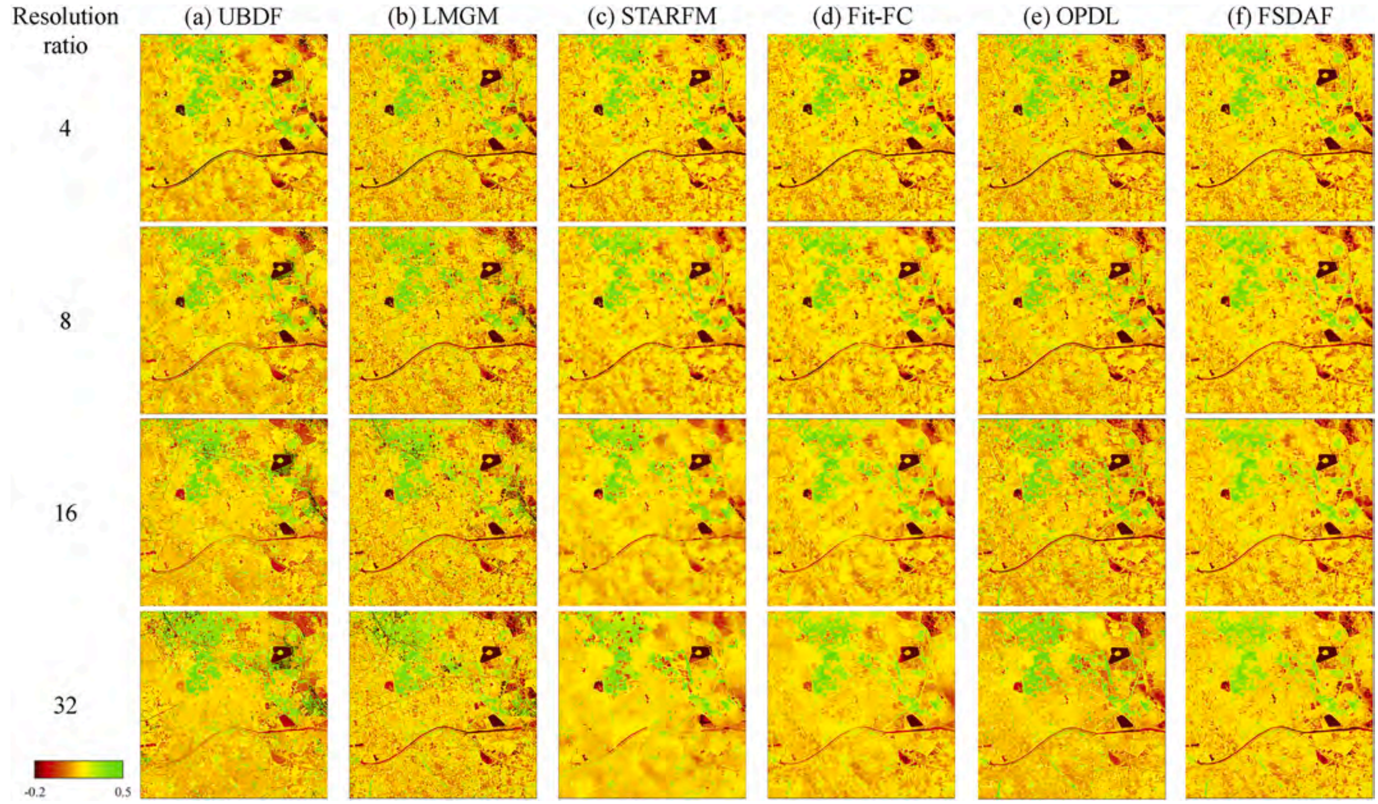

Fig. 13. Based on the NDVI fusion results of the Tianjin site on December 25, 2014, the visual comparison under different levels of spatial resolution ratios, from 4 to 32, by the six methods: (a) UBDF; (b) LMGM; (c) STARFM; (d) Fit-FC; (e) OPDL; and (f) FSDAF.

applications. Similar to a recent comparative study (Liu et al., 2019a), Fit-FC and FSDAF were shown to have better performances than the other three methods (i.e., UBDF, STARFM, and OPDL) for the actual MODIS data, indicating that Fit-FC and FSDAF are robust to different spatiotemporal variations. For a comprehensive comparison, the advantages and disadvantages of the six fusion methods are summarized in Table 7. The most recommended algorithm is Fit-FC, which can produce accurate results with high efficiency for NDVI fusion. However,

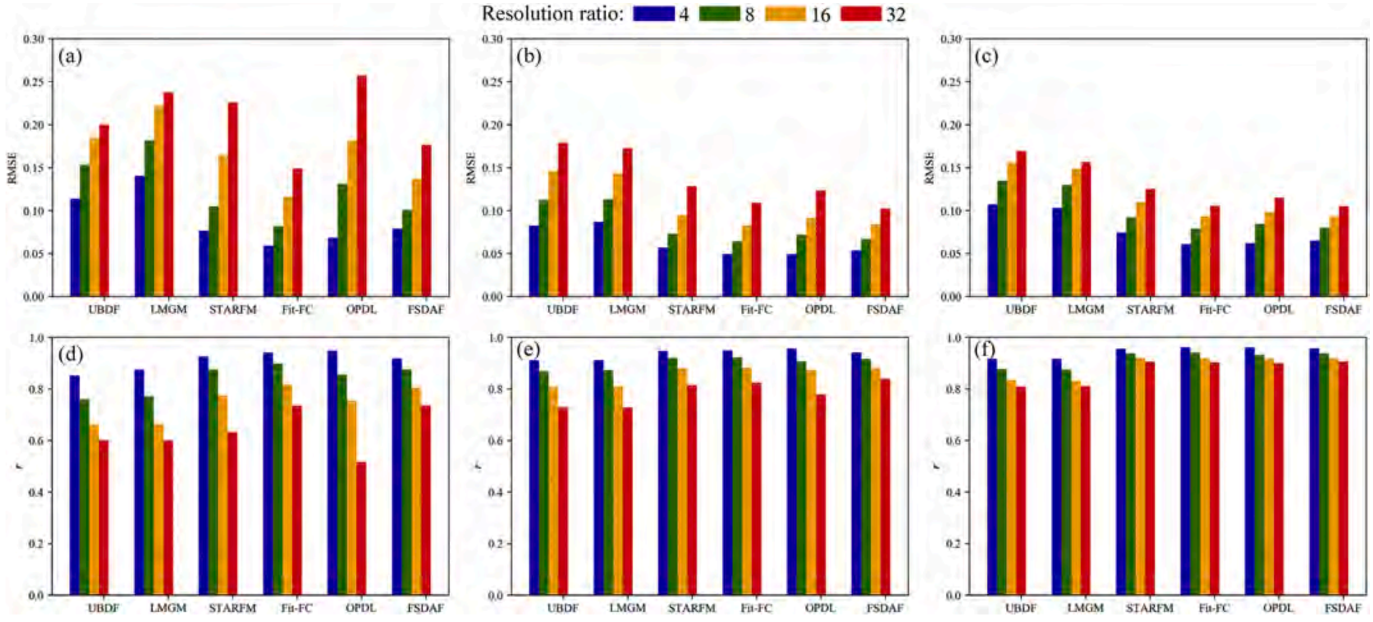

Fig. 14. Quantitative comparison of the NDVI fusion results under different levels of spatial resolution ratios, from 4 to 32: (a) RMSE in CIA; (b) RMSE in GWY; (c) RMSE in Tianjin; (d) r in CIA; (e) r in GWY; and (f) r in Tianjin.

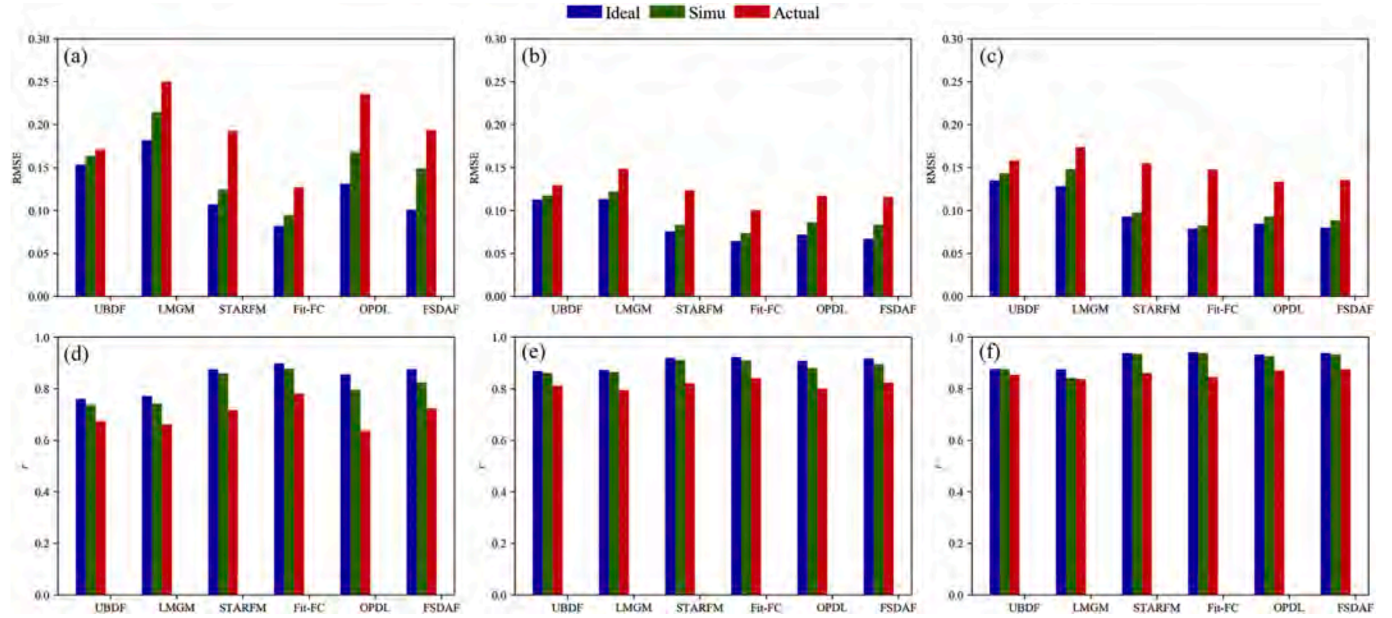

Fig. 15. Quantitative comparison of the NDVI fusion results based on actual MODIS images: (a) RMSE in CIA; (b) RMSE in GWY; (c) RMSE in Tianjin; (d) *r* in CIA; (e) *r* in GWY; and (f) *r* in Tianjin. *Ideal*, *Simu*, and *Actual* imply that the input coarse images are simulated ideally without any errors, simulated with geometric and radiometric errors, and are the actual MODIS images, respectively.

it should be noted that Fit-FC needs to be implemented with radiometric normalization (Gao et al. 2010; Gevaert and García-Haro 2015) considering its sensitivity to systematic radiometric error. FSDAF is another favorable method with high accuracy if the geometric misregistration can be well corrected.

This study can also provide guidance for the future development of spatiotemporal fusion methods. Previous developments in spatiotemporal fusion methods were generally designed without considering inevitable geometric and radiometric errors. For example, increment weighting (Eq. (20)) is commonly used in a large group of fusion methods (e.g., STARFM, LMGM, and FSDAF) because it can maintain good spatial details and reduce the radiometric inconsistency of sensors to some extent. However, the above analysis indicates that it is highly sensitive to geometric error. In contrast, the regression model

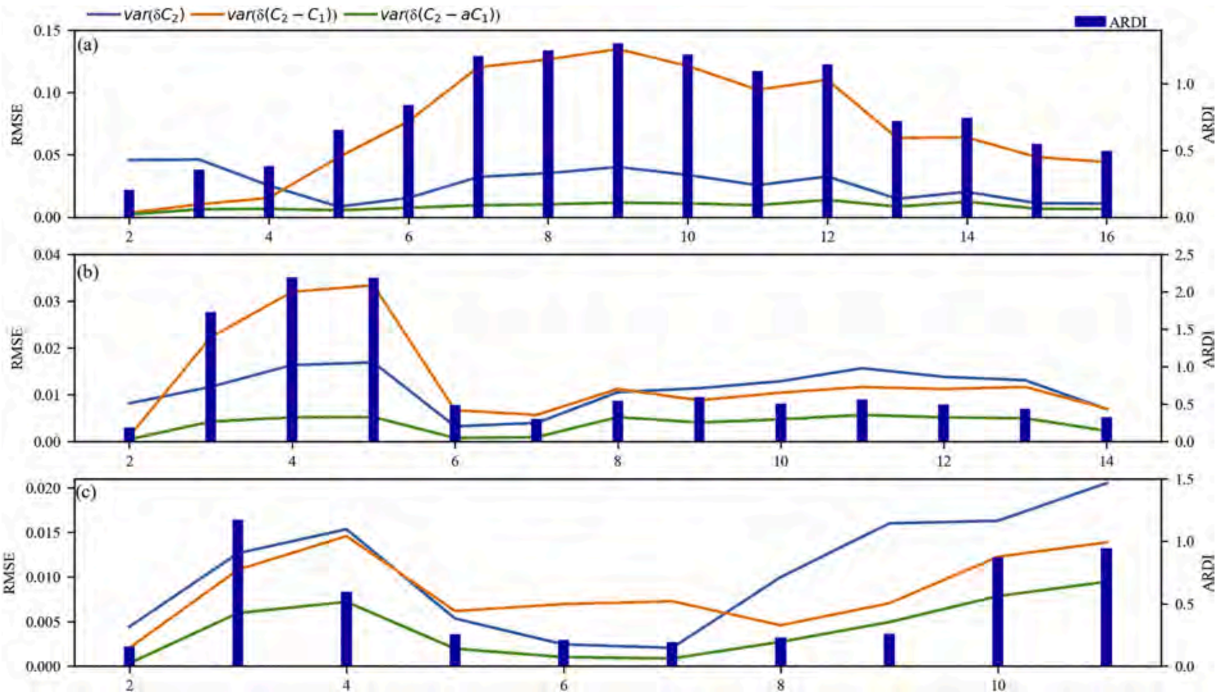

Fig. 16. Comparison of the variances of the three weighting terms in the three sites: (a) CIA, (b) GWY, and (c) Tianjin; image number is the number of the predicted image in the image time series. Fine images were shifted 8 pixels before aggregation.

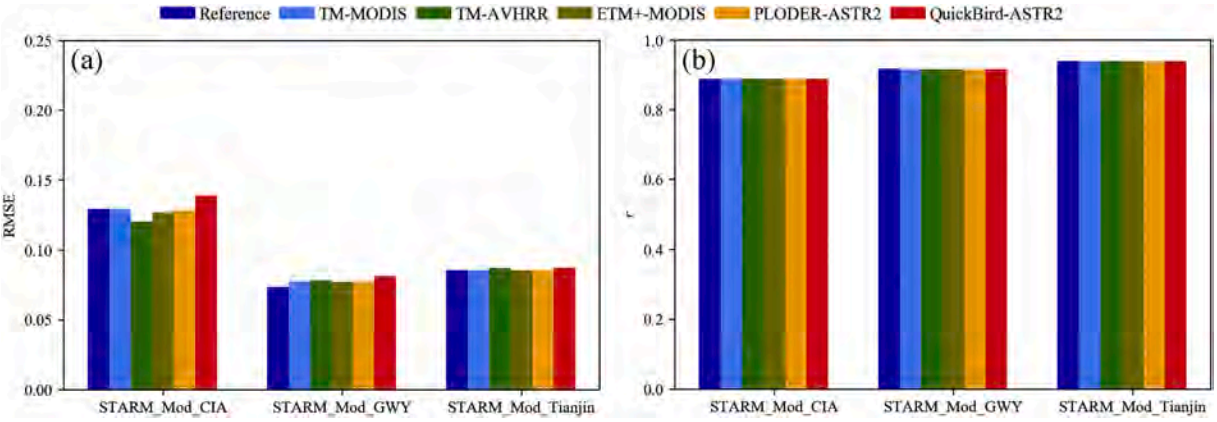

Fig. 17. Quantitative comparison of the NDVI fusion results of the STARFM_Mod_[Site] (i.e., results of the modified STARFM in different sites) under different levels of radiometric inconsistencies: (a) RMSE, and (b) *r*. Reference means that there is no radiometric inconsistency.

Table 7
The pros and cons of the six typical fusion methods under comparison of different influential factors (worst: 1, good: 2–4, best: 5). Due to the dominance of the spatiotemporal variations in the fusion method performances, a triple weight has been used in the calculations of the total scores. (Variations = Spatiotemporal variations; Ratio = Spatial Resolution ratio).

Method	Variations	Geometric	Radiometric	Ratio	Total
UBDF	2	3	2	3	14
LMGM	1	2	5	3	13
STARFM	3	2	3	1	15
Fit-FC	5	5	1	5	26
OPDL	3	3	5	1	18
FSDAF	4	2	5	3	22

employed in Fit-FC is resistant to geometric errors, whereas it is sensitive to radiometric inconsistency. Therefore, combining the strength of Fit-FC and increment weighting might be a promising strategy in the future development of novel methods; other techniques that can mitigate these errors should also be taken into consideration.

It should be noted that this study did not completely consider all the influential factors. The geometric and radiometric errors were simply simulated in this study by pixel shifting and linear transformation. However, there are more complex errors between sensors, including complex geometric errors from imagery scaling, rotation, and skewing (Dai and Khorram 1998; Toutin 2004) and radiometric inconsistencies caused by nonlinear distortion, such as analogous bands between sensors with different spectral response functions, radiometric resolution difference, and the angle effect that solar-sensor geometry bidirectional reflectance distribution function changes over time (Chander et al., 2013b; Gao et al. 2006; Roy et al. 2008). These errors could cause large uncertainties in the fusion results. This is shown in the actual MODIS experiments and should be considered carefully in the future. The selection of typical fusion methods might be another issue. It is impossible in this study to compare all spatiotemporal fusion methods due to limitations of source code availability and heavy work. Notwithstanding the as much representative as possible methods that we selected, some better methods were probably missed. Organizing a programming contest with a standard dataset and assessment protocol could be a solution for engaging more algorithm developers and generating a fair comparison of the different spatiotemporal fusion methods in the near future.

Recently, with advances in sensor technology and an increased number of launched satellites, more and more remotely sensed data, with improved spatial resolution and temporal frequency for land surface observations, are becoming available for the public. For example, Sentinel-2A and − 2B provide high spatial and temporal resolution data for vegetation index (VI) time series together with Landsat free of charge (3–4 days revisit). Commercial satellite constellations, such as Planet, can provide daily observations at 5 m spatial resolution, although it is still expensive for most users. This progress reduces the urgency of data fusion approaches for the current VI reconstruction. However, the significance of data fusion approaches still remains for long-term data analysis, due to the lack of data in the earlier period. Moreover, sensors with fine spatial resolution, such as Sentinel-2, have made progress in enhancing temporal frequency, although they are still not adequate in many study areas with heavy cloud contamination (Wang and Atkinson 2018). Thus, spatiotemporal fusion will still be a widely used technique in future researches (Kimm et al. 2020; Moreno-Martínez et al. 2020).

5. Conclusions

In addition to the spatiotemporal variations of the input datasets, this study presents the necessity of considering the sensitivity of fusion methods to three influential factors (i.e., geometric misregistration, radiometric inconsistency, and spatial resolution ratio) when they are employed in real life applications. These influencing factors could affect different fusion methods to different degrees. The simulation experiment and theoretical analysis showed that Fit-FC achieved the best performance for both sites, with the best resistance to geometric errors among the six typical spatiotemporal fusion methods when the radiometric inconsistency between sensors was negligible, suggesting that it is the first recommended algorithm for NDVI time-series reconstruction. However, Fit-FC is sensitive to systematic radiometric error and thus performs poorly if there is a significant radiometric inconsistency between the two sensors. The FSDAF could also generate satisfactory results through its ability to reduce radiometric inconsistency; however, it is sensitive to geometric errors. Therefore, precise geometric registration is required when using FSDAF. These findings could not only help users select suitable methods according to the characteristics of their data and applications, but could also provide guidance for developers in designing novel algorithms, more robust to different influential factors in the future.

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.

Acknowledgement

This study was supported by the National Natural Science Foundation of China (No. 41830108, No. 41871224, and No. 61872189). The authors would like to thank Dr. Feng Gao and Dr. Qunming Wang for providing the codes of STARFM and Fit-FC and confirmation of the experimental results in this study.

Appendix A. Appendix

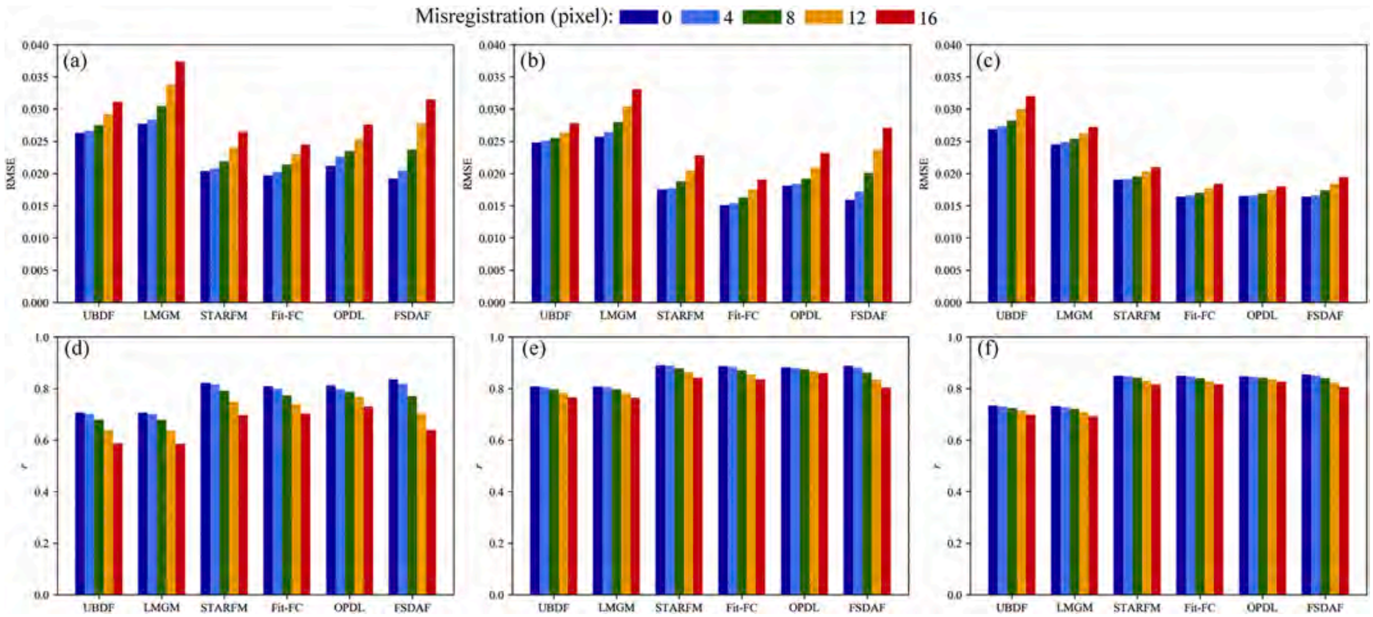

Fig. A1. Quantitative comparison of green band fusion results under different levels of geometric errors, from 0 to 16, (misregistration pixel): (a) RMSE in CIA; (b) RMSE in GWY; (c) r in CIA; and (d) r in GWY. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

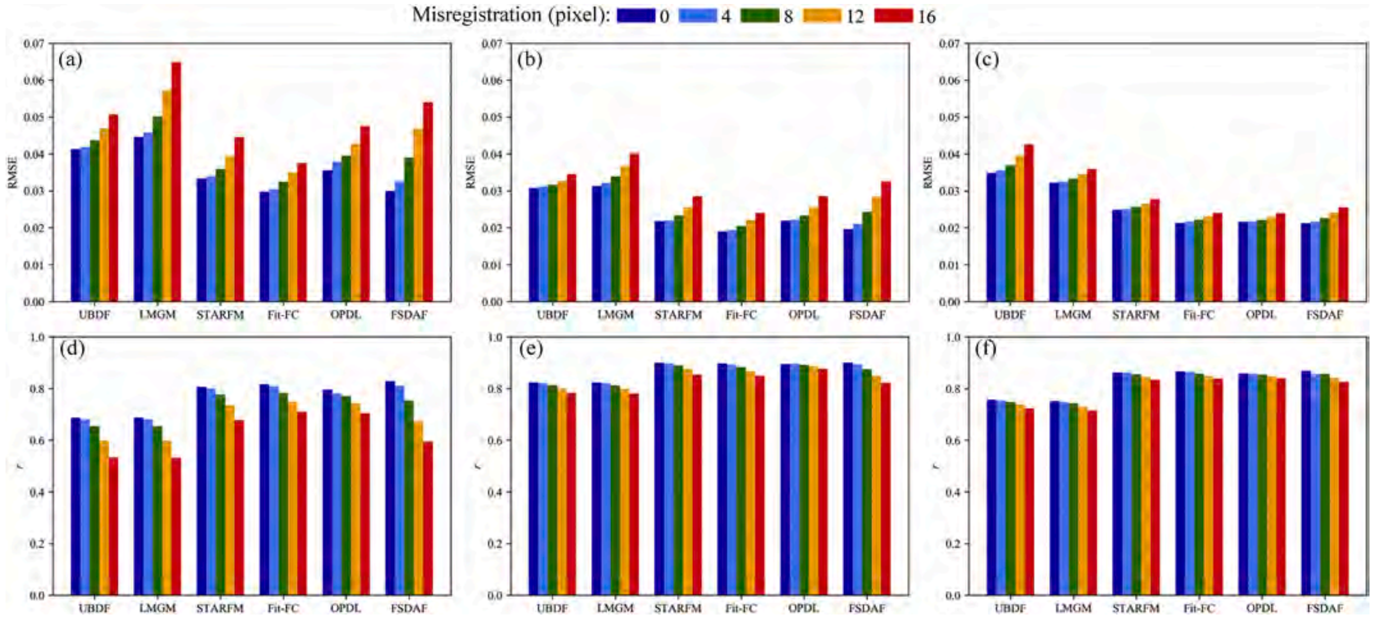

Fig. A2. Quantitative comparison of red band fusion results under different levels of geometric errors, from 0 to 16, (misregistration pixel): (a) RMSE in CIA; (b) RMSE in GWY; (c) r in CIA; and (d) r in GWY. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

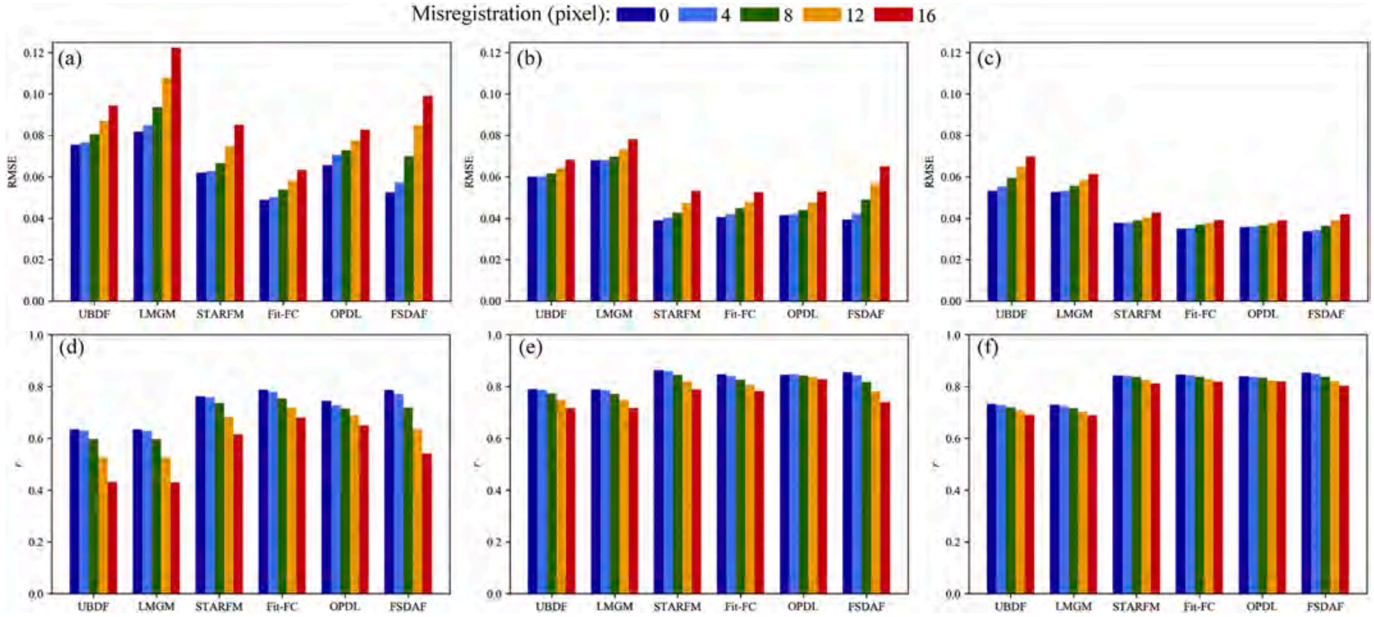

Fig. A3. Quantitative comparison of NIR band fusion results under different levels of geometric errors, from 0 to 16, (misregistration pixel): (a) RMSE in CIA; (b) RMSE in GWY; (c) *r* in CIA; and (d) *r* in GWY.

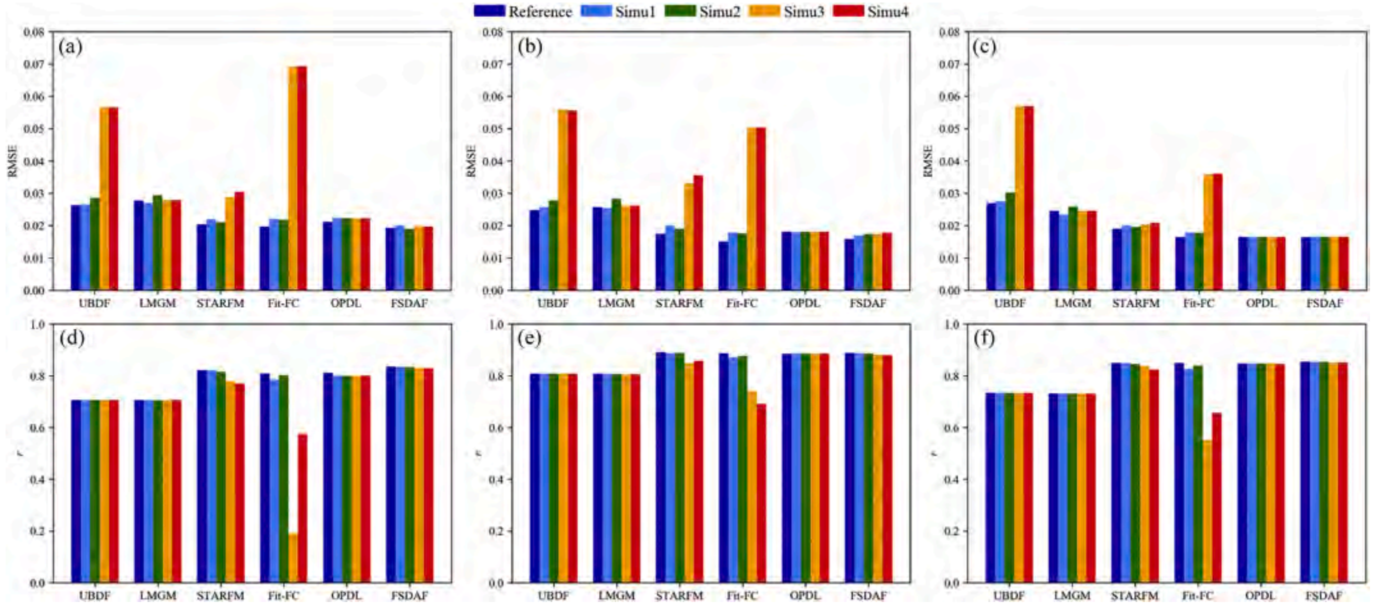

Fig. A4. Quantitative comparison of green band fusion results under different levels of radiometric inconsistencies: (a) RMSE in CIA; (b) RMSE in GWY; (c) *r* in CIA; and (d) *r* in GWY. Reference means that there is no radiometric inconsistency. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

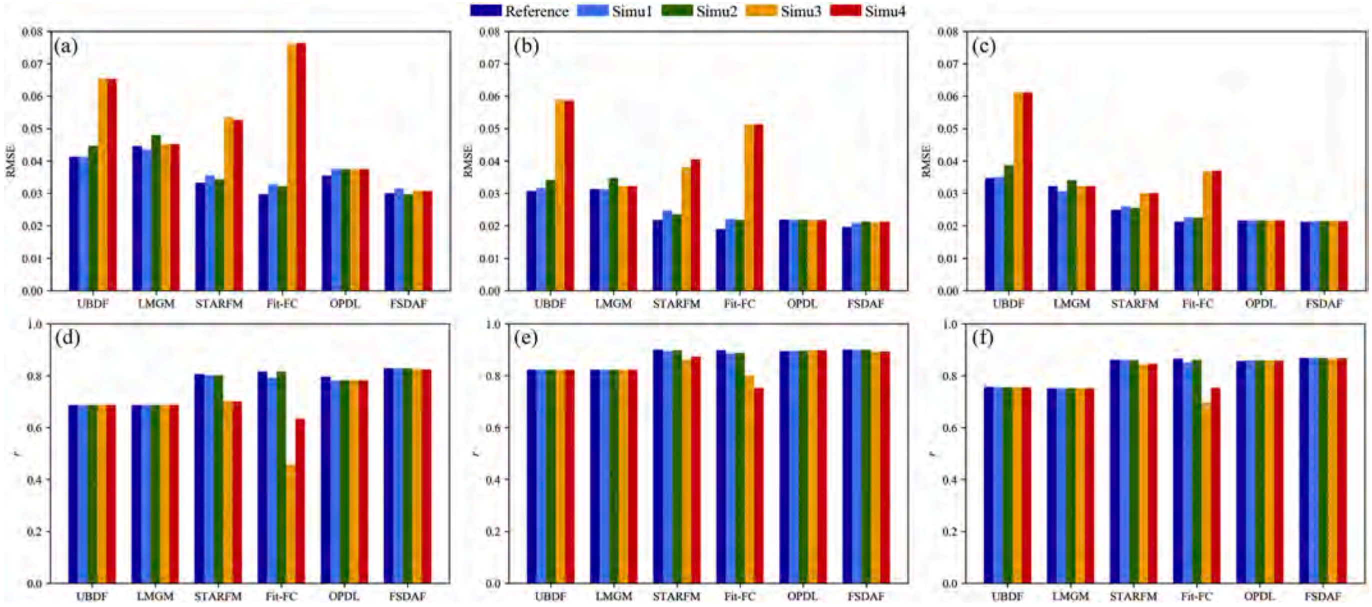

Fig. A5. Quantitative comparison of red band fusion results under different levels of radiometric inconsistencies: (a) RMSE in CIA; (b) RMSE in GWY; (c) *r* in CIA; and (d) *r* in GWY. Reference means that there is no radiometric inconsistency. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

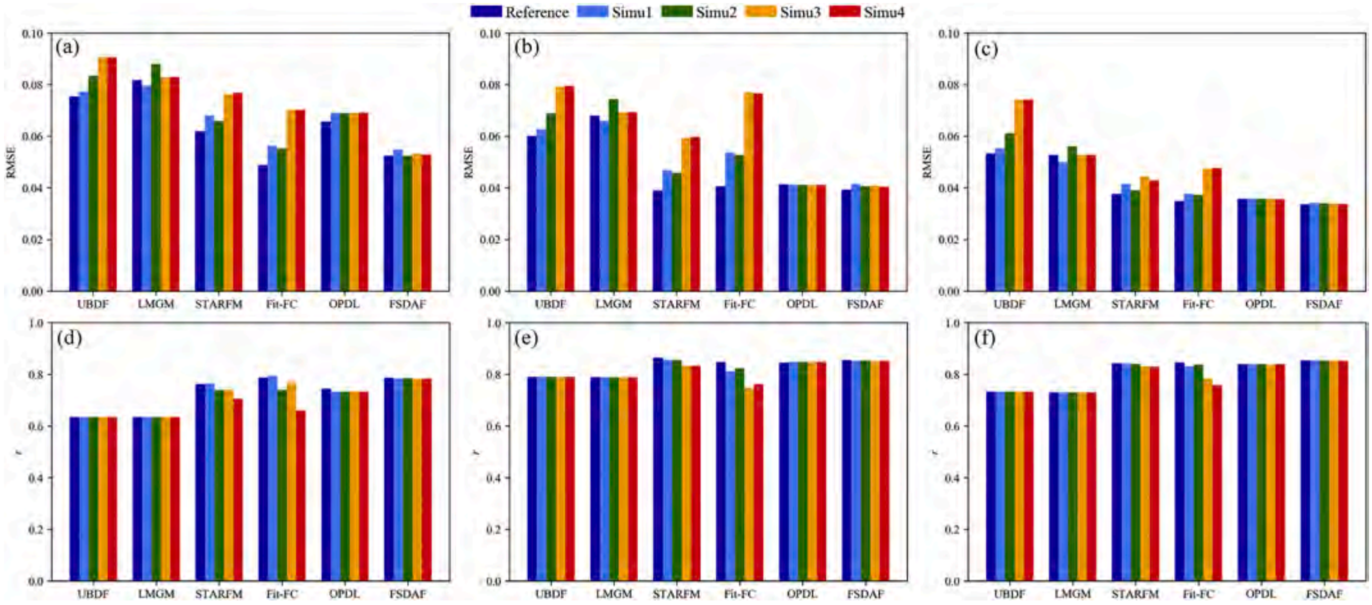

Fig. A6. Quantitative comparison of NIR band fusion results under different levels of radiometric inconsistencies: (a) RMSE in CIA; (b) RMSE in GWY; (c) *r* in CIA; and (d) *r* in GWY. Reference means that there is no radiometric inconsistency.

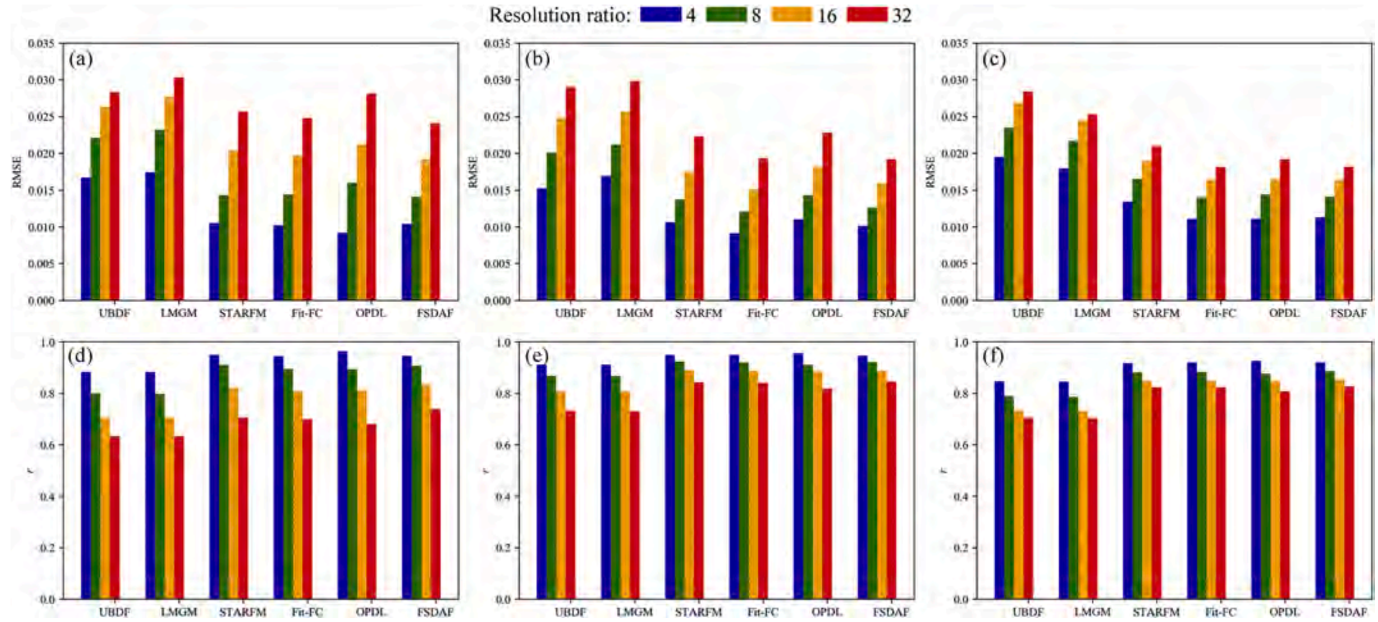

Fig. A7. Quantitative comparison of green band fusion results under different levels of spatial resolution ratios, from 4 to 32: (a) RMSE in CIA; (b) RMSE in GWY; (c) *r* in CIA; and (d) *r* in GWY. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

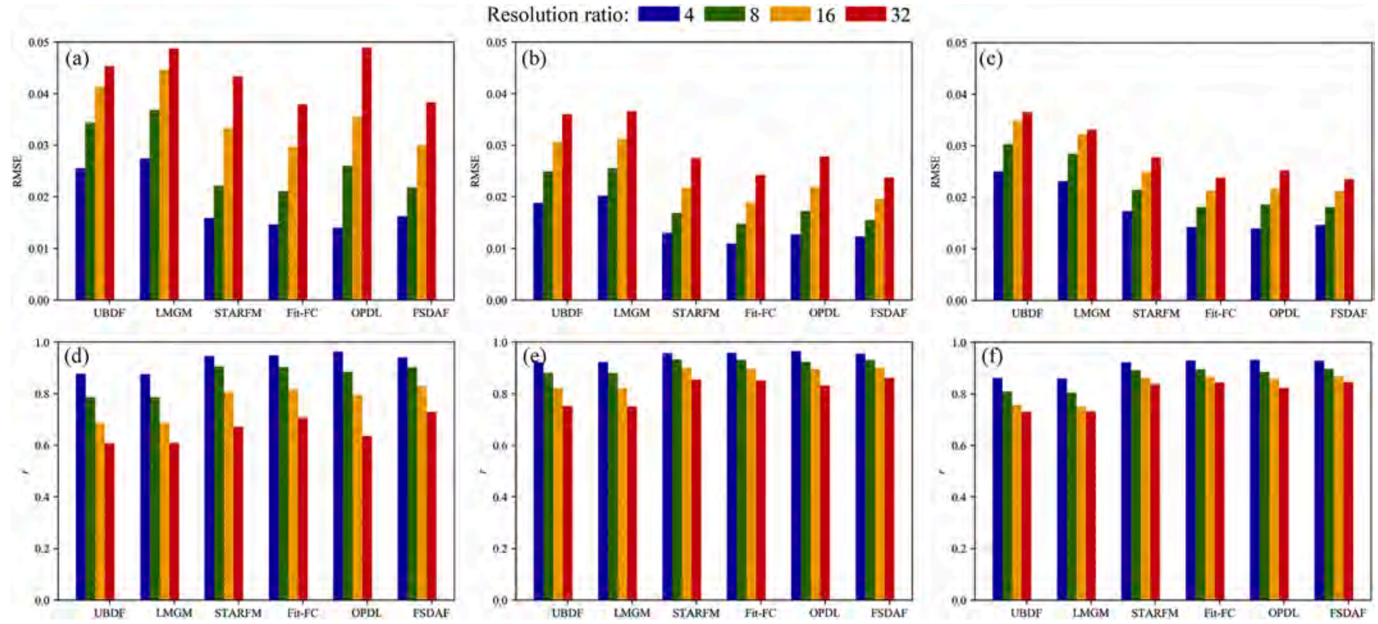

Fig. A8. Quantitative comparison of red band fusion results under different levels of spatial resolution ratios, from 4 to 32, (a) RMSE in CIA; (b) RMSE in GWY; (c) *r* in CIA; and (d) *r* in GWY. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

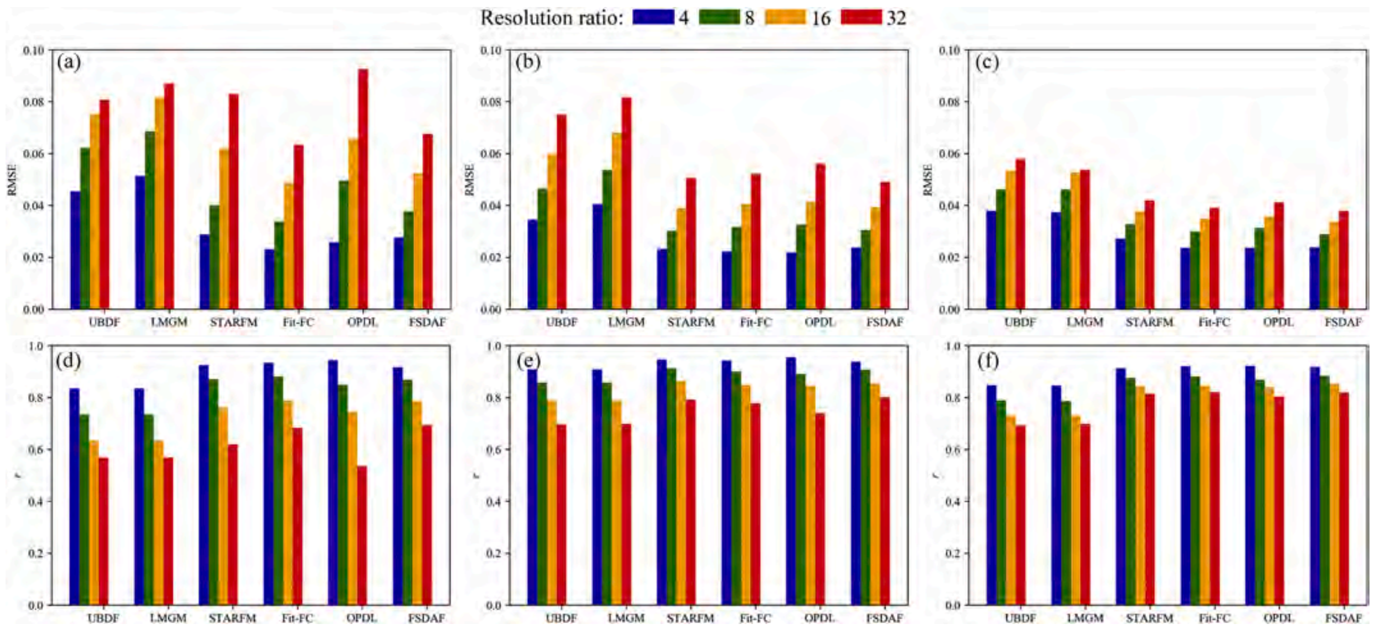

Fig. A9. Quantitative comparison of NIR band fusion results under different levels of spatial resolution ratios, from 4 to 32: (a) RMSE in CIA; (b) RMSE in GWY; (c) *r* in CIA; and (d) *r* in GWY.

References

Belgiu, M., Stein, A., 2019. Spatiotemporal image fusion in remote sensing. Remote Sens. (Basel) 11, 818. https://doi.org/10.3390/rs11070818.

Bradley, B.A., Jacob, R.W., Hermance, J.F., Mustard, J.F., 2007. A curve fitting procedure to derive inter-annual phenologies from time series of noisy satellite NDVI data. Remote Sens. Environ. 106, 137–145. https://doi.org/10.1016/j.rse.2006.08.002.

Cao, R., Chen, J., Shen, M., Tang, Y., 2015. An improved logistic method for detecting spring vegetation phenology in grasslands from MODIS EVI time-series data. Agric. For. Meteorol. 200, 9–20. https://doi.org/10.1016/j.agrformet.2014.09.009.

Chander, G., Helder, D.L., Aaron, D., Mishra, N., Shrestha, A.K., 2013a. Assessment of spectral, misregistration, and spatial uncertainties inherent in the cross-calibration study. IEEE Trans. Geosci. Remote Sens. 51, 1282–1296. https://doi.org/10.1109/TGRS.2012.2228008.

Chander, G., Hewison, T.J., Fox, N., Wu, X., Xiong, X., Blackwell, W.J., 2013b. Overview of intercalibration of satellite instruments. IEEE Trans. Geosci. Remote Sens. 51, 1056–1080. https://doi.org/10.1109/TGRS.2012.2228654.

Chang, J., Hansen, M.C., Pittman, K., Carroll, M., DiMiceli, C., 2007. Corn and soybean mapping in the United States using MODIS time-series data sets. Agron. J. 99, 1654–1664. https://doi.org/10.2134/agronj2007.0170.

Chen, G., Zhao, K., Powers, R., 2014. Assessment of the image misregistration effects on object-based change detection. ISPRS J. Photogramm. Remote Sens. 87, 19–27. https://doi.org/10.1016/j.isprsjprs.2013.10.007.

Chen, B., Huang, B., Xu, B., 2015. Comparison of spatiotemporal fusion models: a review. Remote Sens. (Basel) 7, 1798–1835. https://doi.org/10.3390/rs70201798.

Chen, B., Huang, B., Xu, B., 2017. Multi-source remotely sensed data fusion for improving land cover classification. ISPRS J. Photogramm. Remote Sens. 124, 27–39. https://doi.org/10.1016/j.isprsjprs.2016.12.008.

Chen, X., Liu, M., Zhu, X., Chen, J., Zhong, Y., Cao, X., 2018. Blend-then-Index or Index-then-Blend: A Theoretical Analysis for Generating High-resolution NDVI Time Series by STARFM. Photocrammetric Eng. Remote Sens. 84, 65–73. https://doi.org/10.14358/PERS.84.2.65.

Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.C., Skakun, S.V., Justice, C., 2018. The harmonized Landsat and Sentinel-2 surface reflectance data set. Remote Sens. Environ. 219, 145–161. https://doi.org/10.1016/j.rse.2018.09.002.

Dai, X., Khorram, S., 1998. The effects of image misregistration on the accuracy of remotely sensed change detection. IEEE Trans. Geosci. Remote Sens. 36, 1566–1577. https://doi.org/10.1109/36.718860.

Dubrule, O., 1984. Comparing splines and kriging. Comput. Geosci. 10, 327–338. https://doi.org/10.1016/0098-3004(84)90030-X.

Emelyanova, I.V., McVicar, T.R., Van Niel, T.G., Li, L.T., van Dijk, A.I., 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.

Fan, X., Liu, Y., 2018. Multisensor normalized difference vegetation index Intercalibration: a comprehensive overview of the causes of and solutions for multisensor differences. IEEE Geosci. Remote Sens. Mag. 6, 23–45. https://doi.org/10.1109/MGRS.2018.2859814.

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. Remote Sens. 44, 2207–2218. https://doi.org/10.1109/TGRS.2006.872081.

Gao, F., Masek, J.G., Wolfe, R.E., Huang, C., 2010. Building a consistent medium resolution satellite data set using moderate resolution imaging spectroradiometer products as reference. J. Appl. Remote. Sens. 4, 043526. https://doi.org/10.1117/1.3430002.

Gao, F., Hilker, T., Zhu, X., Anderson, M., Masek, J., Wang, P., Yang, Y., 2015. Fusing Landsat and MODIS data for vegetation monitoring. IEEE Geosci. Remote Sens. Mag. 3, 47–60. https://doi.org/10.1109/MGRS.2015.2434351.

Gao, F., Anderson, M.C., Zhang, X., Yang, Z., Alfieri, J.G., Kustas, W.P., Mueller, R., Johnson, D.M., Prueger, J.H., 2017. Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. Remote Sens. Environ. 188, 9–25. https://doi.org/10.1016/j.rse.2016.11.004.

Gevaert, C.M., García-Haro, F.J., 2015. A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion. Remote Sens. Environ. 156, 34–44. https://doi.org/10.1016/j.rse.2014.09.012.

Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., 2015. Regional detection, characterization, and attribution of annual forest change from 1984 to 2012 using Landsat-derived time-series metrics. Remote Sens. Environ. 170, 121–132. https://doi.org/10.1016/j.rse.2015.09.004.

Hilker, T., Wulder, M.A., Coops, N.C., Linke, J., McDermid, G., Masek, J.G., Gao, F., White, J.C., 2009. A new data fusion model for high spatial-and temporal-resolution mapping of forest disturbance based on Landsat and MODIS. Remote Sens. Environ. 113, 1613–1627. https://doi.org/10.1016/j.rse.2009.03.007.

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

Huang, B., Zhang, H., Song, H., Wang, J., Song, C., 2013. Unified fusion of remote-sensing imagery: generating simultaneously high-resolution synthetic spatial–temporal–spectral earth observations. Remote Sens Lett. 4, 561–569. https://doi.org/10.1080/2150704X.2013.769283.

Jia, K., Liang, S., Wei, X., Yao, Y., Su, Y., Jiang, B., Wang, X., 2014. Land cover classification of Landsat data with phenological features extracted from time series MODIS NDVI data. Remote Sens. (Basel) 6, 11518–11532. https://doi.org/10.3390/rs61111518.

Kimm, H., Guan, K., Jiang, C., Peng, B., Gentry, L.F., Wilkin, S.C., Wang, S., Cai, Y., Bernacchi, C.J., Jian, Peng, Luo, Y., 2020. Deriving high-spatiotemporal-resolution leaf area index for agroecosystems in the US Corn Belt using planet labs CubeSat and STAIR fusion data. Remote Sens. Environ. 239, 111615. https://doi.org/10.1016/j.rse.2019.111615.

Kong, F., Li, X., Wang, H., Xie, D., Li, X., Bai, Y., 2016. Land cover classification based on fused data from GF-1 and MODIS NDVI time series. Remote Sens. (Basel) 8, 741. https://doi.org/10.3390/rs8090741.

Kwan, C., Zhu, X., Gao, F., Chou, B., Perez, D., Li, J., Shen, Y., Koperski, K., Marchisio, G., 2018. Assessment of spatiotemporal fusion algorithms for planet and worldview images. Sensors. 18, 1051. https://doi.org/10.3390/s18041051.

Li, Y., Huang, C., Hou, J., Gu, J., Zhu, G., Li, X., 2017. Mapping daily evapotranspiration based on spatiotemporal fusion of ASTER and MODIS images over irrigated agricultural areas in the Heihe River basin. N. W. Chin. Agr. Forest Meteorol. 244, 82–97. https://doi.org/10.1016/j.agrformet.2017.05.023.

Li, J., Li, Y., He, L., Chen, J., Plaza, A., 2020. Spatio-temporal fusion for remote sensing data: an overview and new benchmark. Sci. China Inf. Sci. 63, 1–140301. https://doi.org/10.1007/s11432-019-2785-y.

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. Remote Sens. (Basel) 8, 452. https://doi.org/10.3390/rs8060452.

Liao, C., Wang, J., Pritchard, I., Liu, J., Shang, J., 2017. A spatio-temporal data fusion model for generating NDVI time series in heterogeneous regions. Remote Sens. (Basel) 9, 1125. https://doi.org/10.3390/rs9111125.

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 Trans. Geosci. Remote Sens Lett. 13, 2039–2043. https://doi.org/10.1109/LGRS.2016.2622726.

Liu, M., Ke, Y., Yin, Q., Chen, X., Im, J., 2019a. Comparison of five Spatio-temporal satellite image fusion models over landscapes with various spatial heterogeneity and temporal variation. Remote Sens. (Basel) 11, 2612. https://doi.org/10.3390/rs11222612.

Liu, M., Yang, W., Zhu, X., Chen, J., Chen, X., Yang, L., Helmer, E.H., 2019b. 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.

Maselli, F., Chiesi, M., Pieri, M., 2019. A new method to enhance the spatial features of multitemporal NDVI image series. IEEE Trans. Geosci. Remote Sens. 57, 4967–4979. https://doi.org/10.1109/TGRS.2019.2894850.

Mizuochi, H., Hiyama, T., Ohta, T., Fujioka, Y., Kambatuku, J.R., Iijima, M., Nasahara, K.N., 2017. Development and evaluation of a lookup-table-based approach to data fusion for seasonal wetlands monitoring: an integrated use of AMSR series, MODIS, and Landsat. Remote Sens. Environ. 199, 370–388. https://doi.org/10.1016/j.rse.2017.07.026.

Moreno-Martínez, Á., Izquierdo-Verdiguier, E., Maneta, M.P., Camps-Valls, G., Robinson, N., Muñoz-Marí, J., Sedano, F., Clinton, N., Running, S.W., 2020. Multispectral high resolution sensor fusion for smoothing and gap-filling in the cloud. Remote Sens. Environ. 247, 111901. https://doi.org/10.1016/j.rse.2020.111901.

Quan, J., Zhan, W., Ma, T., Du, Y., Guo, Z., Qin, B., 2018. An integrated model for generating hourly Landsat-like land surface temperatures over heterogeneous landscapes. Remote Sens. Environ. 206, 403–423. https://doi.org/10.1016/j.rse.2017.12.003.

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. Remote Sens. (Basel) 7, 7865–7891. https://doi.org/10.3390/rs70607865.

Roy, D.P., 2000. The impact of misregistration upon composited wide field of view satellite data and implications for change detection. IEEE Trans. Geosci. Remote Sens. 38, 2017–2032. https://doi.org/10.1109/36.851783.

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. Remote Sens. 54, 7135–7148. https://doi.org/10.1109/TGRS.2016.2596290.

Skakun, S., Justice, C.O., Vermote, E., Roger, J.C., 2018. Transitioning from MODIS to VIIRS: an analysis of inter-consistency of NDVI data sets for agricultural monitoring. Int. J. Remote Sens. 39, 971–992. https://doi.org/10.1080/01431161.2017.1395970.

Song, H., Huang, B., 2013. Spatiotemporal satellite image fusion through one-pair image learning. IEEE Trans. Geosci. Remote Sens. 51, 1883–1896. https://doi.org/10.1109/TGRS.2012.2213095.

Song, H., Liu, Q., Wang, G., Hang, R., Huang, B., 2018. Spatiotemporal satellite image fusion using deep convolutional neural networks. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11, 821–829. https://doi.org/10.1109/JSTARS.2018.2797894.

Steven, M.D., Malthus, T.J., Baret, F., Xu, H., Chopping, M.J., 2003. Intercalibration of vegetation indices from different sensor systems. Remote Sens. Environ. 88, 412–422. https://doi.org/10.1016/j.rse.2003.08.010.

Sulla-Menashe, D., Friedl, M.A., Woodcock, C.E., 2016. Sources of bias and variability in long-term Landsat time series over Canadian boreal forests. Remote Sens. Environ. 177, 206–219. https://doi.org/10.1016/j.rse.2016.02.041.

Tang, Y., Wang, Q., Zhang, K., Atkinson, P., 2020. Quantifying the effect of registration error on Spatio-temporal fusion. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. https://doi.org/10.1109/JSTARS.2020.2965190.

Toutin, T., 2004. Geometric processing of remote sensing images: models, algorithms and methods. Int. J. Remote Sens. 25, 1893–1924. https://doi.org/10.1080/0143116031000101611.

Verbesselt, J., Zeileis, A., Herold, M., 2012. Near real-time disturbance detection using satellite image time series. Remote Sens. Environ. 123, 98–108. https://doi.org/10.1016/j.rse.2012.02.022.

Wang, Q., Atkinson, P.M., 2018. Spatio-temporal fusion for daily Sentinel-2 images. Remote Sens. Environ. 204, 31–42. https://doi.org/10.1016/j.rse.2017.10.046.

Wardlow, B.D., Egbert, S.L., Kastens, J.H., 2007. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the US central Great Plains. Remote Sens. Environ. 108, 290–310. https://doi.org/10.1016/j.rse.2006.11.021.

Wolfe, R.E., Nishihama, M., Fleig, A.J., Kuyper, J.A., Roy, D.P., Storey, J.C., Patt, F.S., 2002. Achieving sub-pixel geolocation accuracy in support of MODIS land science. Remote Sens. Environ. 83, 31–49. https://doi.org/10.1016/S0034-4257(02)00085-8.

Xiao, Z., Liang, S., Wang, J., Xiang, Y., Zhao, X., Song, J., 2016. Long-time-series global land surface satellite leaf area index product derived from MODIS and AVHRR surface reflectance. IEEE Trans. Geosci. Remote Sens. 54, 5301–5318. https://doi.org/10.1109/TGRS.2016.2560522.

Xie, D., Gao, F., Sun, L., Anderson, M., 2018. Improving spatial-temporal data fusion by choosing optimal input image pairs. Remote Sens. (Basel) 10, 1142. https://doi.org/10.3390/rs10071142.

Yan, L., Roy, D.P., Zhang, H., Li, J., Huang, H., 2016. An automated approach for sub-pixel registration of Landsat-8 operational land imager (OLI) and Sentinel-2 multi spectral instrument (MSI) imagery. Remote Sens. (Basel) 8, 520. https://doi.org/10.3390/rs8060520.

Yokoya, N., Grohnfeldt, C., Chanussot, J., 2017. Hyperspectral and multispectral data fusion: a comparative review of the recent literature. IEEE Geosci. Remote Sens. Mag. 5, 29–56. https://doi.org/10.1109/MGRS.2016.2637824.

Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., Hodges, J.C., Gao, F., Reed, B.C., Huete, A., 2003. Monitoring vegetation phenology using MODIS. Remote Sens. Environ. 84, 471–475. https://doi.org/10.1016/S0034-4257(02)00135-9.

Zhang, B., Zhang, L., Xie, D., Yin, X., Liu, C., Liu, G., 2016. Application of synthetic NDVI time series blended from Landsat and MODIS data for grassland biomass estimation. Remote Sens. (Basel) 8, 10. https://doi.org/10.3390/rs8010010.

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., Helmer, E.H., Gao, F., Liu, D., Chen, J., Lefsky, M.A., 2016. A flexible spatio-temporal 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., Cai, F., Tian, J., Williams, T., 2018. Spatiotemporal fusion of multisource remote sensing data: literature survey, taxonomy, principles, applications, and future directions. Remote Sens. (Basel) 10, 527. https://doi.org/10.3390/rs10040527.

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

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