https://doi.org/10.1016/j.rse.2020.112130

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], Huihui Song[d], Yuean Qiu[a,b], 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

d. Jiangsu Key Laboratory of Big Data Analysis Technology, Nanjing University of Information Science and Technology, Nanjing 210044, China

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

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 comparison studies among different fusion methods have been conducted, selecting 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 with the consideration of three influential factors between sensors, including geometric misregistration, radiometric inconsistency, and spatial resolution ratio. The results indicated 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. FSDAF could generate satisfactory results with resistance to radiometric inconsistency as well. These findings could help users to determine which method is appropriate for different remote

sensing datasets and provide guidelines for developers in the future development of novel methods.

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

1. Introduction

Time series of vegetation indices (e.g., Normalized Difference Vegetation Index, NDVI) produced by satellite sensors play a unique role in various environmental applications as important data sources, 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, due to hardware technology or budget limitations. For example, the data from sensors with dense temporal coverage usually hold 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 drawback due to a long revisit cycle (e.g., 16 days), limiting their potential in time-series analyses. Consequently, various spatiotemporal fusion methods that combine the merits of two such kinds of data have been developed and were used to produce NDVI data with high spatial and temporal resolutions (Chen et al., 2018; Liao et al., 2017; Liu et al., 2019; Maselli et al., 2019; Rao et al., 2015). Furthermore, they have been successfully applied in various fields, such as crop growth progress 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 pixel 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 target pixel through combining neighborhood pixels with empirically designed weight functions of spectral similarity, spatial distance or other related measurements (Gao et al., 2006; Wang et al., 2018; Zhu et al., 2010). Learning-based methods are relatively new, which use machine learning methods to model the relationship between coarse and fine images (Huang and Song, 2012; Liu et al., 2016; Song and Huang, 2013; Song et al., 2018). Bayesian-based methods described spatio-temporal fusions as a Maximum A Posterior (MAP) problem based on 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 spatiotemporal fusion (Li et al., 2020; Liu et al., 2019; 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 had been conducted to explore the advantages and weaknesses of different methods based on time-series data (Chen et al. 2015; Emelyanova et al., 2013; Liu et al., 2019). As a general conclusion, the performances of different fusion methods mainly depend on the sensitivity to spatial heterogeneity and temporal variations of the used data.

However, these comparison studies have neglected the influence of inevitable noise in real applications, including geometric misregistration and radiometric inconsistency. In spite of large efforts devoted on the inter-calibration and geometric registration among different sensors, adequate elimination on such inherent noises is still challenging (Chander et al. 2013a, 2013b; Claverie et al., 2018; Yan et al., 2016).Thus, numerous studies have focused on quantifying the impact of geometric misregistration error and radiometric uncertainty on land cover change detection (Dai and Khorram, 1998; Chen et al., 2014; Roy et al., 2000) and vegetation dynamic monitoring (Fan and Liu, 2018; Skakun et al., 2018; Sulla-Menashe et al., 2016). Considering the potential impacts of these noises on 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 the fusion methods (Tang et al., 2020). However, only two algorithms were explored in this study, which is not enough for most users.

Recently, open source data from sensors with fine spatial resolution have made a progress in enhancing temporal frequency, such as Sentinel-2, but they are still not sufficient in many applications due to cloud contamination (Wang et al., 2018). Thus, Sentinel-3 at 300 m resolution is also launched to supply daily observations, which is crucial for monitoring land surface dynamics. Moreover, long-term data analysis is also very important for many applications and spatiotemporal fusion plays important role for the study period without rich data. Therefore, spatiotemporal fusion methods are not only used for MODIS and Landsat images, but also images from other satellite sensors with different spatial resolutions (e.g., AMSR, ASTER, Sentinel-2/3, GF-1, Worldview, and Planet) in recent studies (Kong et al., 2016; Kwan et al., 2018; Mizuochi et al., 2017; Li et al., 2017; Wang and Atkinson, 2018). And it has been aware that the input images with different resolution ratios could lead to significant variations of different method performances (Yokoya et al., 2017). Unfortunately, to our knowledge, there are no comparative studies for evaluating the performances of spatiotemporal fusion methods based on data with different spatial resolution ratios of sensors.

To fill the gap in the previous comparison studies, we conducted comparison experiments and theoretical analyses on the spatiotemporal fusion of NDVI time-series data with considering various influential factors, including geometric misregistration, radiometric inconsistency, and spatial resolution ratio. Six typical spatiotemporal fusion methods requiring only one fine image and two coarse images as input, including the UBDF (Unmixing-Based Data Fusion), LMGM (Linear Mixing Growth Model), STARFM (Spatial and Temporal Adaptive Reflectance Fusion Model), Fit-FC (regression model Fitting, spatial Filtering and residual Compensation), OPDL (One Pair Dictionary-Learning method), and FSDAF (Flexible Spatiotemporal DAta Fusion), were selected for comparison. The 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 requirement. 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 to users and developers.

2. Methods and datasets

2.1 Experiment design

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

NDVI fusion is the main object of this study considering the widely application of

NDVI time series. As the time series of surface reflectance also receive attentions (Hermosilla et al., 2015; Xiao et al., 2016), similar fusion experiments were also conducted on reflectance data (green, red, near infrared bands) for a 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 fine and coarse images. In the standard experiment, we aggregate 8×8 pixels for NDVI and 16×16 pixels for reflectance to simulate the MODIS NDVI at 250m resolution and MODIS reflectance data at 500m resolution. The first fine image in the time series and 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 by different fusion methods (Fig. 1. Schematic diagram of the standard experiment.).

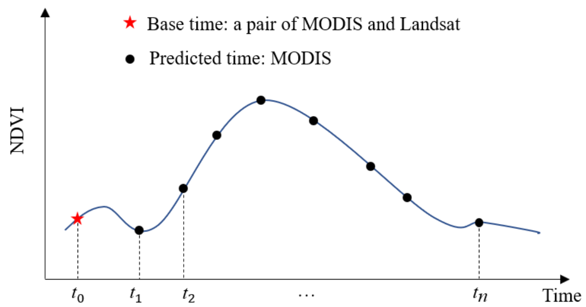

Fig. 1. Schematic diagram of the standard experiment.

2.1.2 Fusion experiment based on simulated data with geometric misregistration

The settings of this experiment are similar to the standard one, except that misregistration error was simulated when aggregating the Landsat images to the coarse images.

Specifically, pixel shifting is one of the most serious consequences caused by geometric distortions. Therefore, similar to the previous study (Tang et al., 2020), the fine images were shifted 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 the geometric error. For experiments of reflectance, the fine images were shifted 4, 8, 12, or 16 pixels before aggregation considering the resolution of coarse reflectance images was doubled as that of NDVI image.

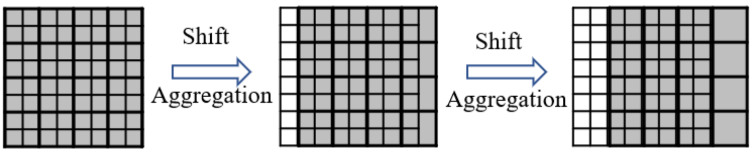

Fig. 2. Schematic diagram of the experiment with geometric misregistration.

2.1.3 Fusion experiment based on simulated data with radiometric inconsistency

The special experiment setting of this experiment, which is the only difference from 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 the stretched NDVI of simulated coarse pixels, respectively; 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, as well as QuickBird and ASTR2 were investigated (Table 1). And Table 2 presents the linear relationships used in the reflectance experiments. With such a simulation, the sensitivity of different methods to the radiometric inconsistency could be explored.

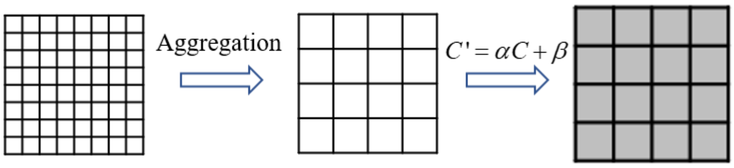

Fig. 3. Schematic diagram of the experiment with radiometric inconsistency

Table 1

Coefficients of linear stretches for simulated radiometric inconsistency in the NDVI experiments between sensors (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 in the reflectance experiments between sensors.

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

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

To explore the applicability of 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 4 levels of spatial resolution ratios (4, 8, 16, and 32; Fig. 4). Other experiment settings are similar to those in the standard one (i.e., without any geometric error or radiometric inconsistency).

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

2.1.5 Fusion experiment based on actual MODIS images

The actual MODIS images were used for this fusion experiment. In addition, simulated MODIS images with certain geometric error and radiometric inconsistency were also used for comparison. Although the geolocation accuracy of MODIS achieved 50 m at nadir (Wolfe et al., 2002), the large scan angle and procedures of reprojection and resampling could further enlarge the geolocation error. Thus, it should be reasonable to assume an averaged misregistration error of approximately a half-pixel size (120 m), corresponding to four pixel shifting of fine images. The parameters of linear stretches for TM-MODIS (Table 1) were used for simulating 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, root mean square error (RMSE) and, correlation coefficient

(r) were used to evaluate the performance of different fusion methods. The RMSE was calculated using all pairs of the predicted and true images throughout the time series. The correlation coefficient (r) was calculated between the predicted and the 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 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 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 different fusion methods under different experimental scenarios.

In addtion, 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 temporal change between 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 fine images at based and predicted time.

2.2 Experimental Datasets

For a unified comparison, the typical datasets in previous spatiotemporal fusion studies (Emelyanova et al., 2013; Jun et al., 2020), Coleambally irrigated 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 in northern China (117.20°E, 39.30°N) were used in this study. The CIA site was dominated by woodlands, cropland, and dryland land 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 woodlands parcels in this area, resulting in a heterogeneous landscape. In addition, woodlands, croplands, and drylands show distinctive NDVI profiles during this period (Fig. 5(c)). The main purpose using this dataset with the high heterogeneity and complex NDVI seasonality is to compare the performance of fusion methods for the monitoring phenology changes in fragmented cropland landscape. 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 for capturing abrupt land cover change. As for the third site, Tianjin, the main land cover were impervious surface, cropland and waterbody. Many small impervious surfaces (e.g., buildings, and roads) were distributed in this site, resulting in a heterogenous landscape. There were 11 cloud-free pairs of Landsat-8 OLI (800×800 pixels at 30 m spatial resolution) collected for Tianjin Site. As shown in Fig. Fig. 7, each land cover had the unique NDVI temporal profile. The main purpose using this dataset is to test the accuracy of fusion methods for the detecting phenology changes in urban land-scapes. For all three sites, 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.

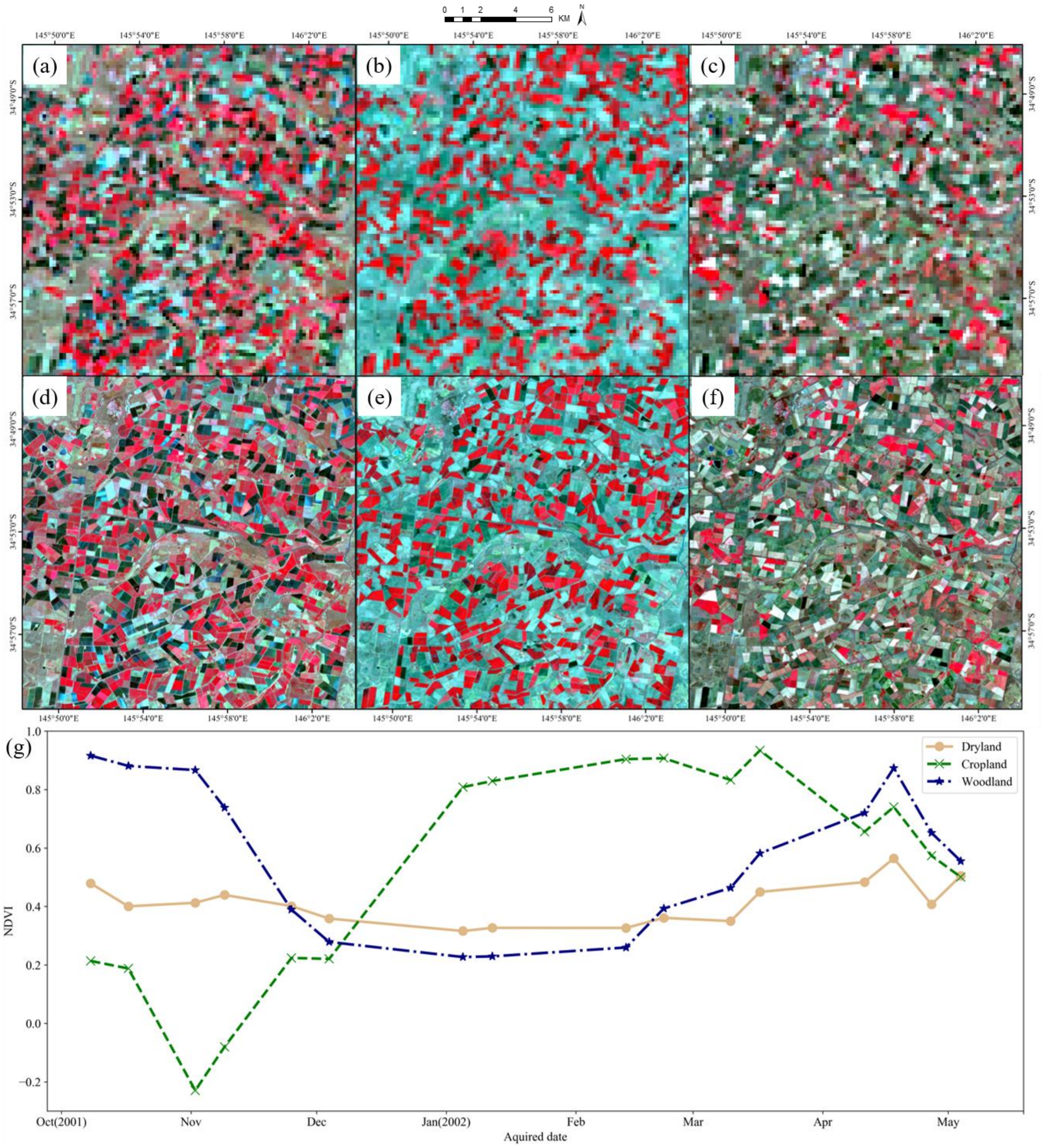

Fig. 5. Test data in CIA site. 240 m simulated coarse images and corresponding 30 m fine images acquired on (a) and (d) November 9, 2001, (b) and (e) February 13, 2002,

and (c) and (f) May 4, 2002; (g) NDVI time-series of three typical land covers. All images use NIR-red-green as RGB.

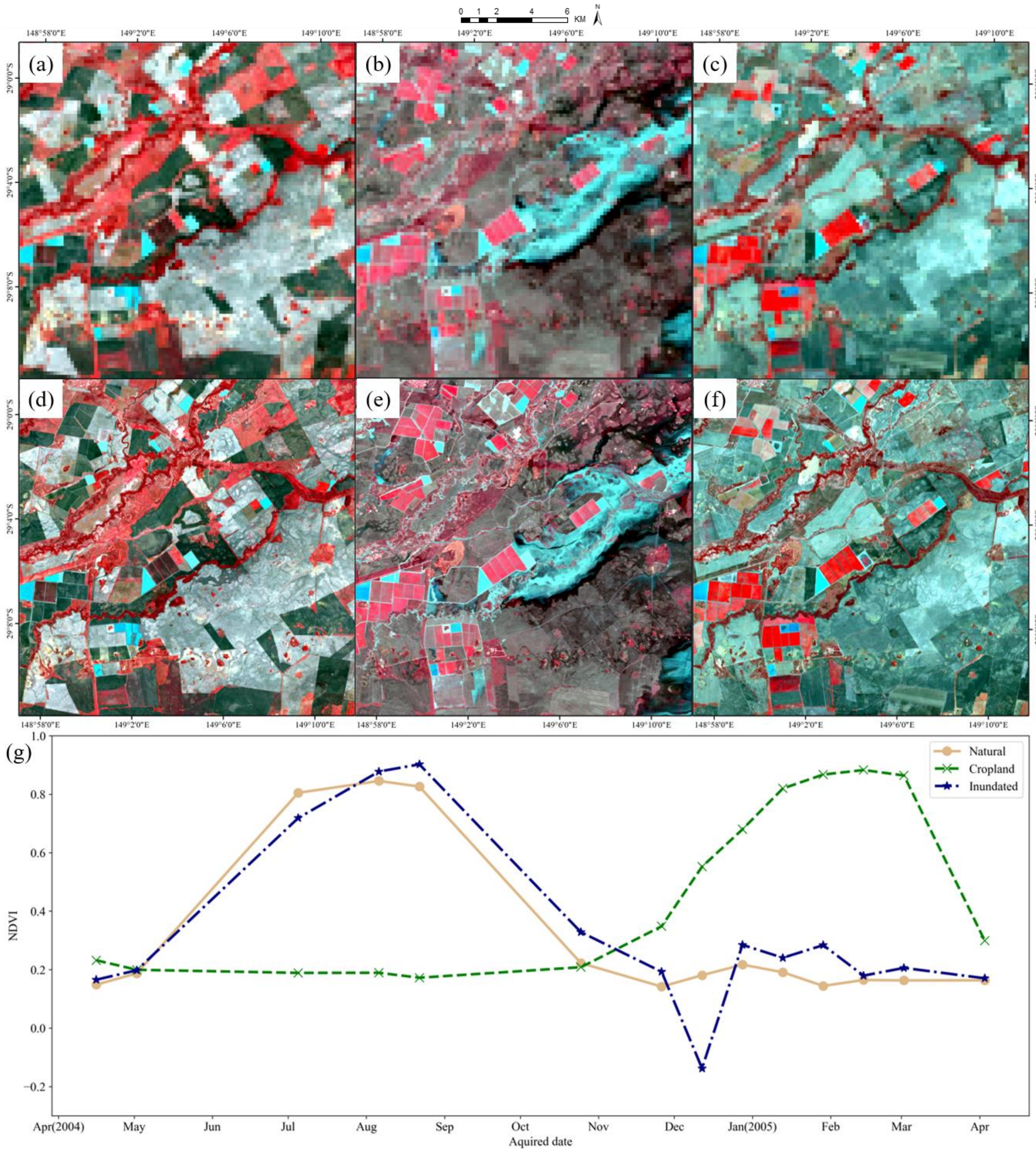

Fig. 6. Test data in GWY site. 240 m simulated coarse images and corresponding 30 m fine images acquired on (a) and (d) May 2, 2004, (b) and (e) December 12, 2004, and (c) and (f) April 4, 2005; (g) NDVI time-series of three typical land covers. All images use NIR-red-green as RGB.

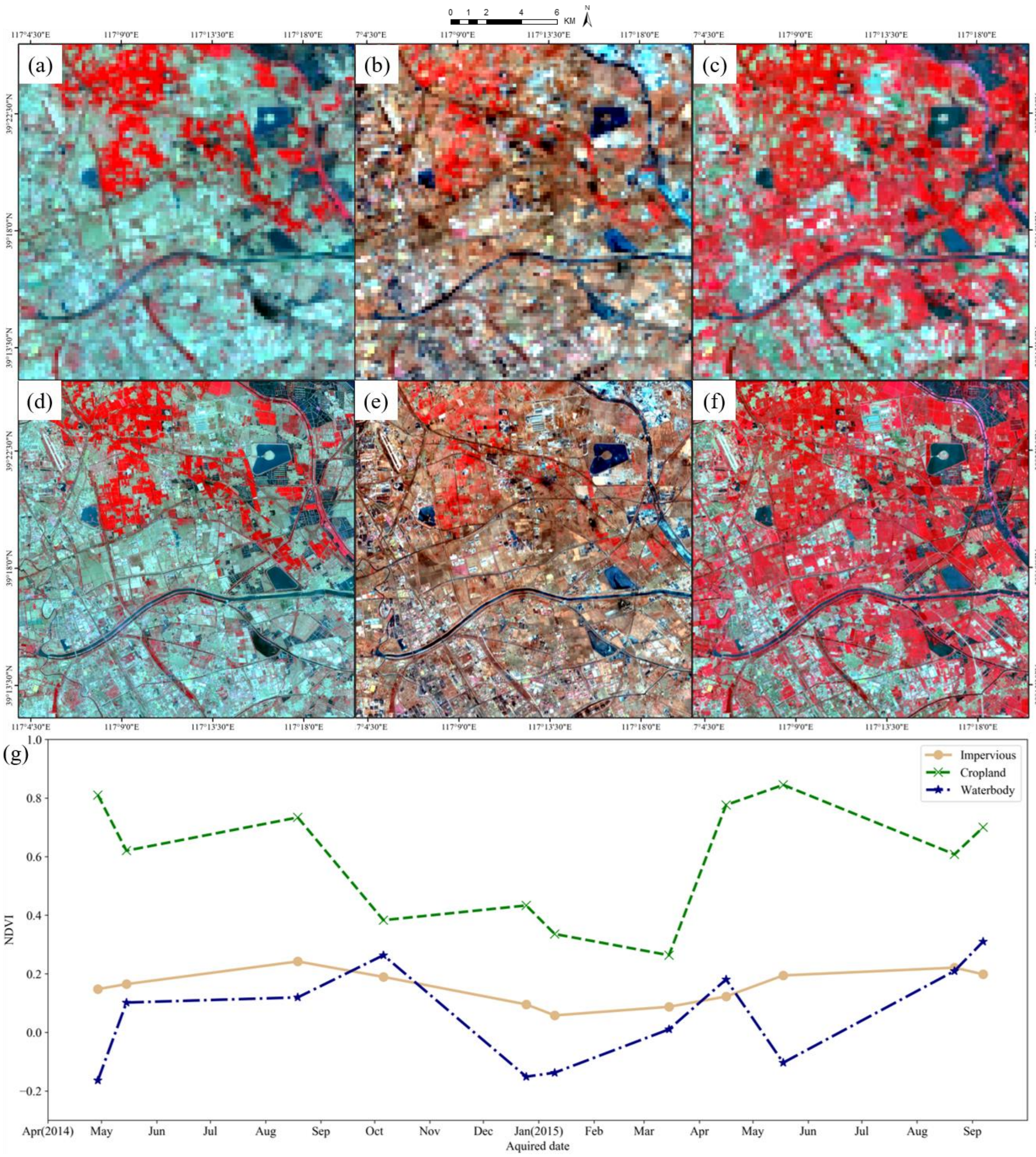

Fig. 7. Test data in Tianjin site. 240 m simulated coarse images and corresponding 30 m fine images acquired on (a) and (d) April 29, 2014, (b) and (e) December 25, 2014, and (c) and (f) August 2, 2015; (g) NDVI time-series of three typical land covers. All images use NIR-red-green as RGB.

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 due to the lack of open-source code. For quantifying the error propagation of fusion results caused by misregistration and radiometric consistency between sensors, key concepts and equations of each method were introduced here for the convenience of the later theoretical analysis in the discussion part.

For simplification, the algorithms were reintroduced here based on a consistent denotation (Table 3).

Table 3 Common variables used in different spatiotemporal fusion methods.

Symbol	Meaning
(x, y)	geolocation of specific pixel
t_1	base time;
t_2	predicted time;
C_1	the input coarse image at t_1;
C_2	the input coarse image at t_2;
F_1	the input fine image at t_1;
$\hat{F}_2$	the output image at t_2
$\mathbf{M}$	the moving window of pixel (x, y).

2.3.1 UBDF

As an unmixing-based method, UBDF employ 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, 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 a predicted time is:

$$C_2\left(x,y\right)=\sum_{i=1}^{c} f_i(x,y)F_2^i(x,y)+\varepsilon(x,y), \tag{2}$$

where $F_2^i(x,y)$ is the NDVI of the ith land cover type in the coarse pixel (x,y); $f_i(x,y)$ is the fraction of the ith endmember in the coarse pixel; c is the number of endmembers; and ε is the residual error. $f_i(x,y)$ is calculated based on the classification result of 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 (**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):

$$\begin{bmatrix} C_2(1,1) \\ \vdots \\ C_2(x,y) \\ \vdots \\ C_2(m,m) \end{bmatrix} = \begin{bmatrix} f_1(1,1) & f_2(1,1) & \cdots & f_c(1,1) \\ \vdots & \vdots & & \vdots \\ f_1(x,y) & f_2(x,y) & \cdots & f_c(x,y) \\ \vdots & \vdots & & \vdots \\ f_1(m,m) & f_2(m,m) & \cdots & f_c(m,m) \end{bmatrix} \begin{bmatrix} F_2^1 \\ F_2^2 \\ \\ \vdots \\ F_2^c \end{bmatrix}. \tag{3}$$

Also, Eq. (3) could be written in a matrix form for convenience

$$\mathbf{C}_2(\mathbf{M}) = \mathbf{f}(\mathbf{M})\mathbf{F}_2(\mathbf{M}). \tag{4}$$

Thus, F_2^i could be estimated by the least-square method:

$$\hat{\mathbf{F}}_2(\mathbf{M}) = \left[\mathbf{f}^{\mathrm{T}}(\mathbf{M})\mathbf{f}(\mathbf{M})\right]^{-1}\mathbf{f}^{\mathrm{T}}(\mathbf{M})\mathbf{C}_2(\mathbf{M}). \tag{5}$$

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 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 in a short period. Therefore, LMGM estimates the growth rates of endmembers (ΔF=F_2-F_1) by unmixing the growth rate of coarse pixels (ΔC=C_2-C_1), as shown in Eq. (6):

$$\begin{bmatrix} \Delta C(1,1) \\ \vdots \\ \Delta C(x,y) \\ \vdots \\ \Delta C(m,m) \end{bmatrix} = \begin{bmatrix} f_1(1,1) & f_2(1,1) & \cdots & f_c(1,1) \\ \vdots & \vdots & & \vdots \\ f_1(x,y) & f_2(x,y) & \cdots & f_c(x,y) \\ \vdots & \vdots & & \vdots \\ f_1(m,m) & f_2(m,m) & \cdots & f_c(m,m) \end{bmatrix} \begin{bmatrix} \Delta F^1 \\ \Delta F^2 \\ \\ \vdots \\ \Delta F^c \end{bmatrix}. \tag{6}$$

Then, LMGM calculates $\hat{F}_2$ by adding the estimated growth rate of class i ($\Delta\hat{F}^i$) to F_1

$$\hat{F}_2(x,y) = F_1(x,y) + \Delta\hat{F}^i . \tag{7}$$

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 firstly resamples the coarse images to the same spatial resolution as the fine image. Thus, $F_2(x,y)$ can be estimated as:

$$\hat{F}_2(x,y) = F_1(x,y) + \Delta C(x,y). \tag{8}$$

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:

$$\hat{F}_2(x,y) = \sum_{i=1}^{n_s} W_i \left(F_1(x_i, y_i) + \Delta C(x_i, y_i) \right), \tag{9}$$

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

$$D_i = \sqrt{\left(x_{w/2} - x_i\right)^2 + \left(y_{w/2} - y_i\right)^2}\ , \tag{10}$$

$$S_i = \left|F_1\left(x_i, y_i\right) - C_1\left(x_i, y_i\right)\right|, \tag{11}$$

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

2.3.4 Fit-FC

For capturing the temporal changes of fine pixels accurately, Fit-FC introduces a linear regression model established based on coarse images (Wang and Atkinson, 2018). A local linear regression model is firstly established between C_2 and C_1 within a moving window **M**:

$$\mathbf{C}_2\left(\mathbf{M}\right) = a \times \mathbf{C}_1\left(\mathbf{M}\right) + b + \mathbf{R}\left(\mathbf{M}\right), \tag{12}$$

where **R**(**M**) are the coarse residuals in the moving window and a and b are the regression coefficients. Then, the regression coefficients are applied to the fine pixels within a moving window corresponding to the coarse moving window (**M**) for the RM (i.e., Regression Model) prediction. Finally, unlike STARFM with the spectral difference between coarse and fine images, an another searching similar neighboring pixels approach only with spatial distance using threshold is adopted to address the problem of blocky artifacts while considering the residuals of the regression model:

$$\hat{F}_2\left(x, y\right) = \sum_{i=1}^{n_s} W_i \left(a \times F_1\left(x_i, y_i\right) + b + r\left(x_i, y_i\right)\right), \tag{13}$$

where $r(x_i,y_i)$ is the residual at the fine pixel (x_i,y_i), which is resampled from **R**(**M**) by bicubic interpolation.

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 image and fine image 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 transition image T_2. And with the same process, T_1 can be also produced. Finally, the high-pass modulation is introduced to transfer the temporal change from transition images to F_1 for prediction:

$$\hat{F}_2(x, y) = \frac{T_2(x, y)}{T_1(x, y)} F_1(x, y). \tag{14}$$

Due to the large spatial resolution difference between the fine image and coarse image, OPDL is implemented in a two-layer framework (Song and Huang, 2013). The first layer produces an image with the intermediate resolution between the coarse and fine image. Subsequently, the second layer generates the 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. Firstly, 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 FSDAF:

$$R(x_i, y_i) = \Delta C(x_i, y_i) - \frac{1}{n}\left[\sum_{j=1}^{n} F_2^{tp}(x_{ij}, y_{ij}) - \sum_{j=1}^{n} F_1(x_{ij}, y_{ij})\right], \tag{15}$$

where $R(x_i, y_i)$ is the residual in the coarse pixel at location (x, y) and 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:

$$R_{ho}(x, y) = F_2^{sp}(x, y) - F_2^{tp}(x, y). \tag{16}$$

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

$$\hat{F}_2(x, y) = F_1(x, y) + \sum_{i=1}^{n_s} W_i\left(\Delta F^{tp}(x_i, y_i) + n \times R(x_i, y_i) \times w_h(x_i, y_i)\right). \tag{17}$$

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

2.4 Parameter settings of six spatiotemporal fusion methods

Referring to previous studies (Gao et al., 2006; Song and Huang, 2013; Rao et al., 2015; Wang et al., 2018; Zhu et al, 2016; Zurita-Milla et al., 2008), parameters of 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. It is noted 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 OPDL are separately shown in

Table **5** as they are very different from those of other five fusion methods. The patch size of dictionary representation in two layers was consistently set as 3 and 4 for all of the experimental sites and resolution ratios.

Table 4 Key parameters of five fusion methods (*c:* class number, *m:* moving window size, n_s: number of similar neighboring pixels, m_s: moving window size for searching similar neighboring pixels, *R*: Spatial resolution ratio of 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

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

Dictionary size (Layer 1, Layer 2)		Resolution ratio (scale factors of two layers)			
		4 (2×2)	8 (2×4)	16 (4×4)	32 (4×8)
Experi-mental 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)

3. Results

3.1 Standard comparison

The performances of the six methods at the two sites were evaluated with 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 ADRI increased, while Fit-FC and FSDAF always generated better results for all the images in the time series than the other four methods.

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

		UBDF	LMGM	STARFM	Fit-FC	OPDL	FSDAF
CIA	RMSE	0.1533	0.1816	0.1292	0.0816	0.131	0.1006
	r	0.7606	0.7717	0.8883	0.8979	0.8555	0.8758
GWY	RMSE	0.1125	0.1133	0.0754	0.0643	0.0718	0.0669
	r	0.8681	0.8726	0.9196	0.9226	0.9072	0.9175
Tianjin	RMSE	0.1346	0.1296	0.0926	0.0788	0.0843	0.0797
	r	0.8769	0.8748	0.9376	0.9409	0.9329	0.9385

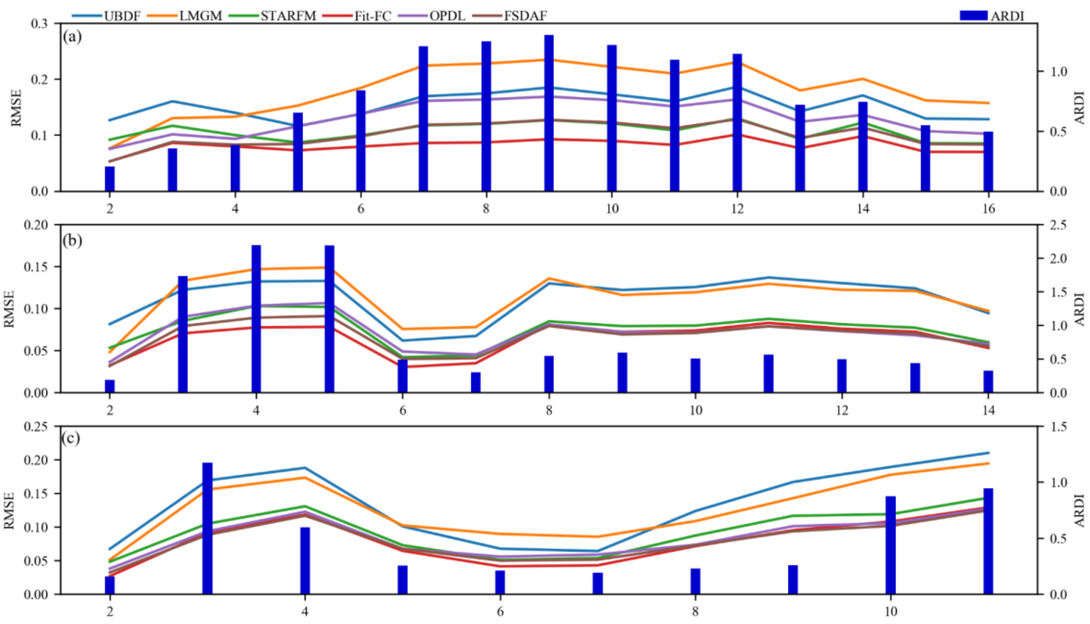

Fig. 8. The relationship between the prediction accuracy and the temporal variation at the two sites: (a) CIA; (b) GWY; (c) Tianjin. Image number is the number of the predicted image in the image time series.

3.2 Geometric misregistration

The performances of the six methods with the simulated misregistration errors between coarse and fine images are showed in this section. The extent of misregistration was quantitatively measured as the shifting pixel distance. From visual comparison of NDVI fusion results (Fig. 9), there are little distortions of the results fused by Fit-FC under eight pixel shifting. However, the results fused by other five fusion methods are obviously different from the reference results. 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 most slowly. 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 drops along in addition to the shifting distance increases. The results of reflectance (Fig. A1, Fig. A4, and Fig. A7) are similar to those of NDVI.[ZJ1]

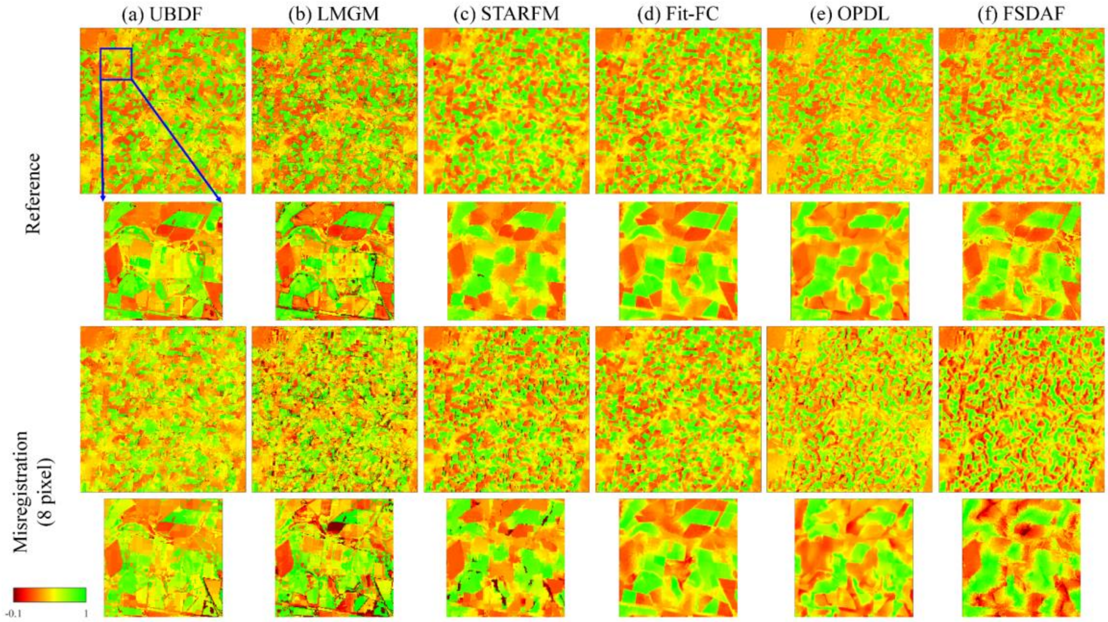

Fig. 9. Using fusion results of the CIA site on February 13, 2002, visual comparison of NDVI results without geometric errors (e.g., reference) and with misregistration (eight pixel shifting) by six methods: (a) UBDF; (b) LMGM; (c) STARFM; (d) Fit-FC; (e) OPDL; (f) FSDAF.

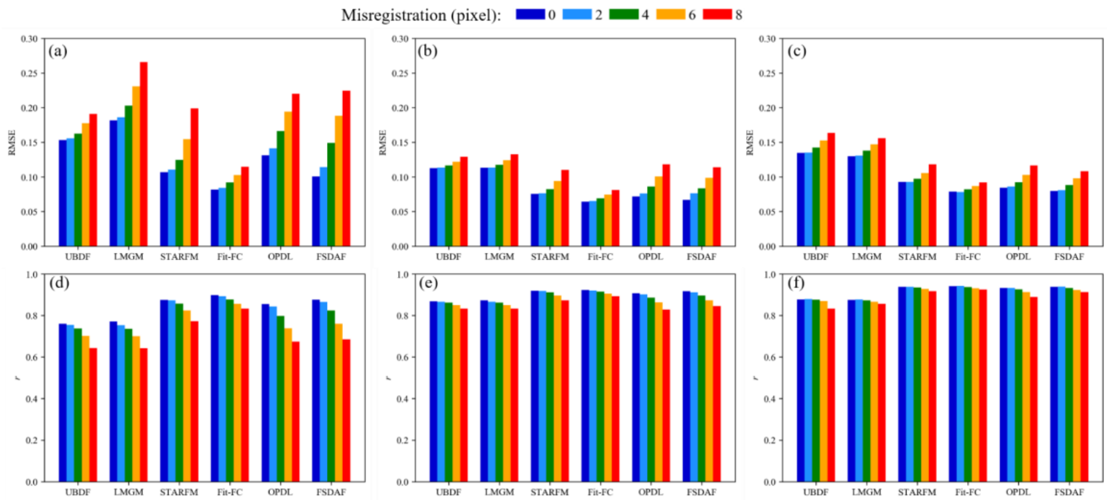

Fig. 10. Quantitative comparison of NDVI fusion results under different levels of geometric errors from 0 to 8 (misregistration pixel). (a) RMSE in CIA; (b) RMSE in GWY; (c) RMSE in Tianjin; (d) r in CIA; (e) r in GWY; (f) r in Tianjin.

3.3 Radiometric inconsistency

Fig. 11 shows the robustness of different methods to the radiometric inconsistency between sensors (i.e., linear stretches of QuickBird-ASTR2). There are significant distortions in the results fused by UBDF, STARFM, and Fit-FC. And the results evaluated by quantitative indices are in good agreement with that of visual comparison. When the fusion methods were applied to the datasets with small radiometric inconsistency (Table 1), such as TM-MODIS, TM-AVHRR, and ETM+-MODIS, they all produced accurate results. However, when there were larger radiometric inconsistencies, like POLDER-ASTR2, and QuickBird-ASTR2, UBDF, STARFM, and Fit-FC showed larger errors than the other methods. In contrast, LMGM, OPDL and FSDAF are more robust to the radiometric inconsistency between two sensors. As for the results of reflectance (Fig. A2, Fig. A5, and Fig. A8), the sensitivity of fusion methods to radiometric errors is consistent with that of NDVI.

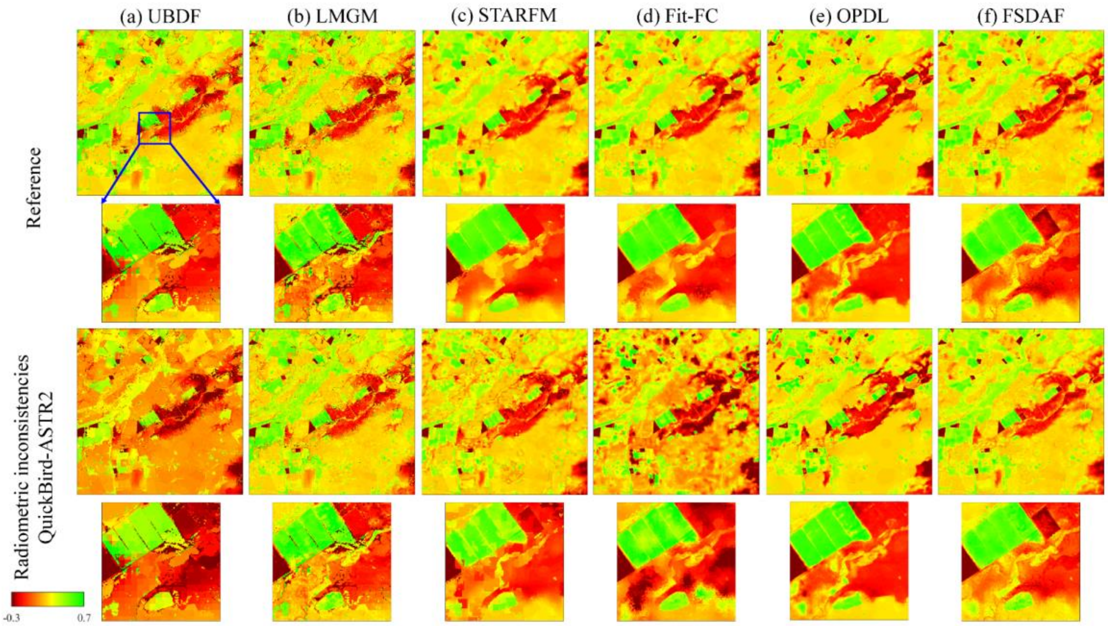

Fig. 11. Using fusion results of the GWY site on December 12, 2004, visual comparison

of 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; (f) FSDAF.

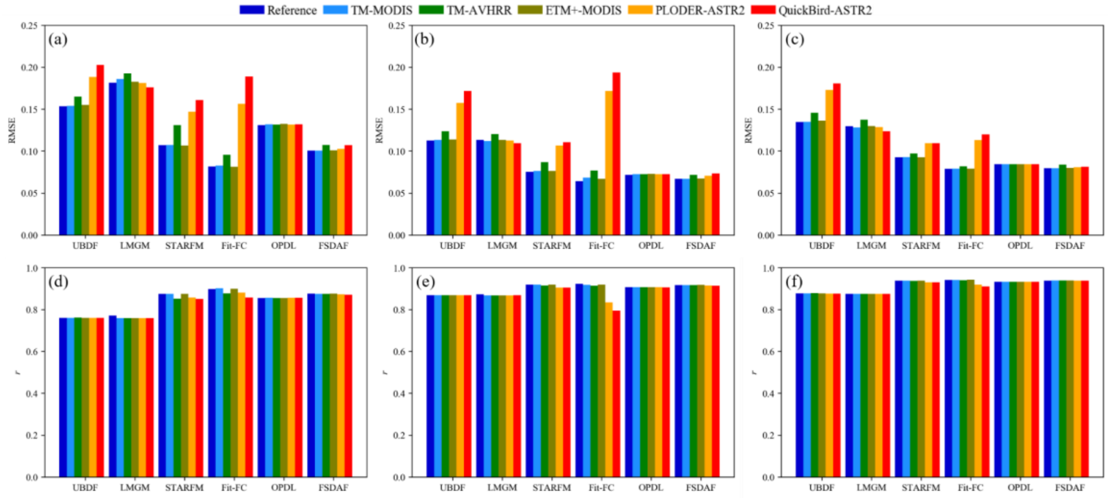

Fig. 12. Quantitative comparison of NDVI fusion results under different levels of radiometric inconsistencies. (a) RMSE in CIA; (b) RMSE in GWY; (c) RMSE in Tianjin; (d) *r* in CIA; (e) *r* in GWY; (f) *r* in Tianjin. Reference means that there is no radiometric inconsistency.

3.4 Spatial resolution ratio

Fig. 13 presents the accuracies of the six fusion methods in the scenarios of different spatial resolution ratios between 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 like CIA. In contrast, UBDF, LMGM, FSDAF, and Fit-FC are somehow less sensitive to the spatial resolution ratio. And the results of reflectance is similar to those of NDVI (Fig. A3, Fig. A6, and Fig. A9).

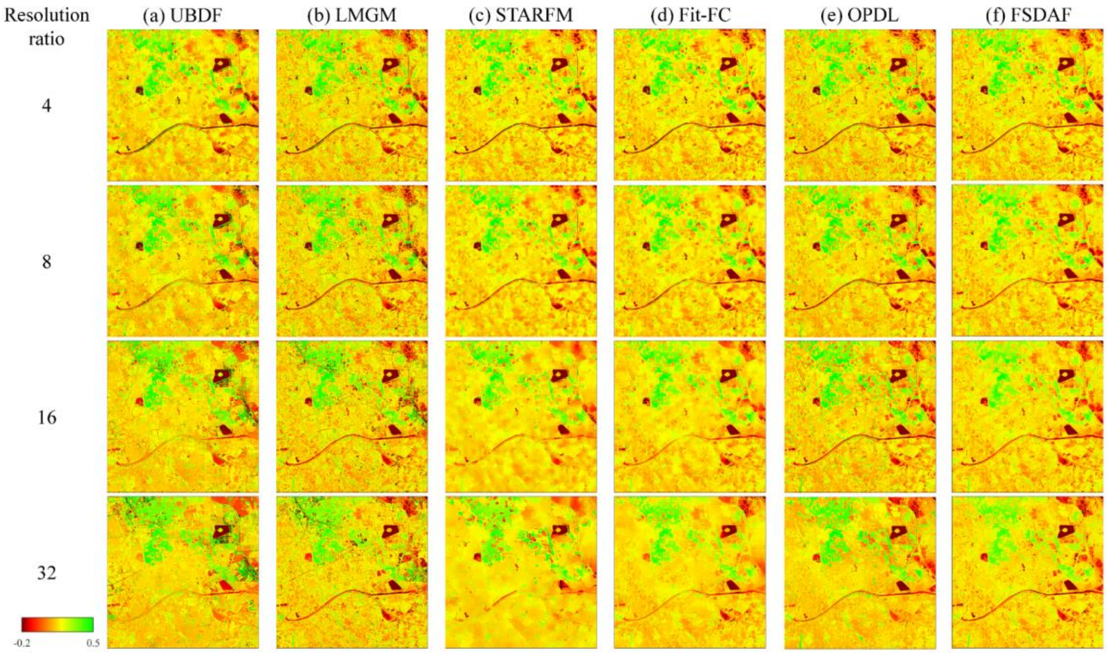

Fig. 13. Using NDVI fusion results of the Tianjin site on December 25, 2014, visual comparison under different levels of spatial resolution ratio from 4 to 32. by six methods: (a) UBDF; (b) LMGM; (c) STARFM; (d) Fit-FC; (e) OPDL; (f) FSDAF.

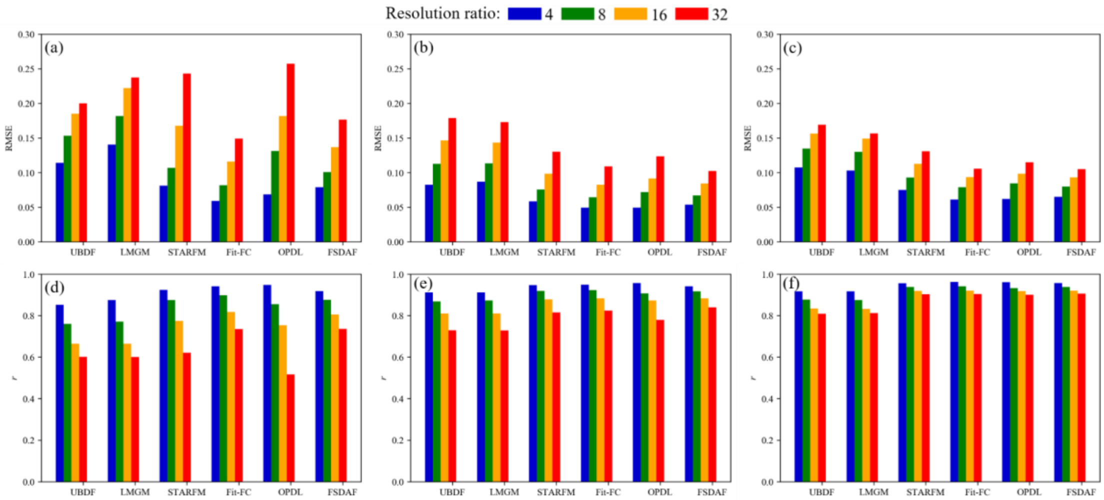

Fig. 14. Quantitative comparison of NDVI fusion results under different levels of spatial resolution ratio from 4 to 32. (a) RMSE in CIA; (b) RMSE in GWY; (c) RMSE in Tianjin; (d) r in CIA; (e) r in GWY; (f) r in Tianjin.

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 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. [ZJ2]For the experiment based on actual MODIS data, Fit-FC performed best in CIA and GWY sites whereas OPDL performed best in Tianjin site. FSDAF ranked second in GWY and Tianjin sites, whereas UBDF ranked second in CIA site. These results imply the complexity of the fusion experiments based on actual data. However, in overall, Fit-FC and FSDAF still worth recommendation considering their relative stability.

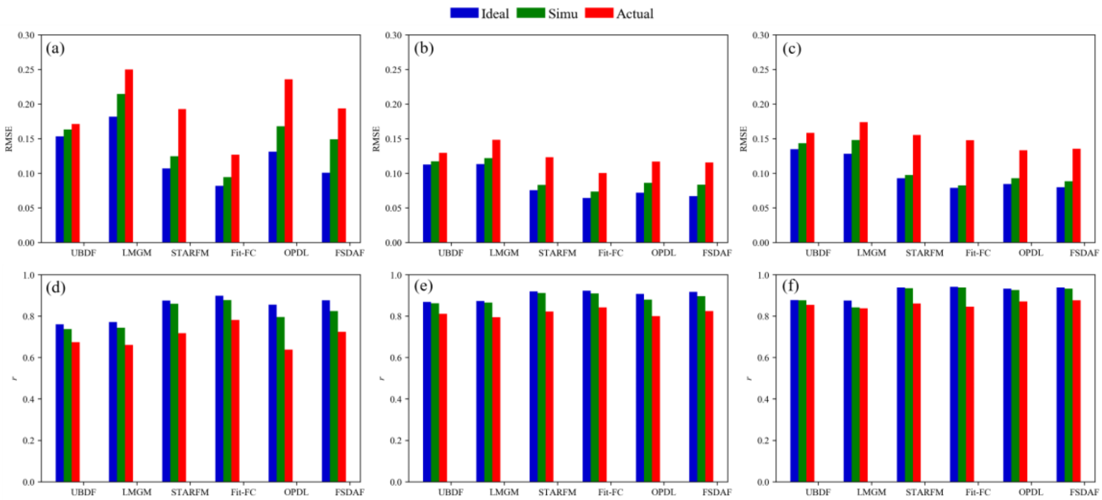

Fig. 15. Quantitative comparison of NDVI fusion results based on actual MODIS images. (a) RMSE in CIA; (b) RMSE in GWY; (c) RMSE in Tianjin; (d) *r* in CIA; (e) *r* in GWY; (f) *r* in Tianjin. Ideal, Simu, and Actual imply that the input coarse images are

simulated ideally without any errors, simulated with geometric and radiometric errors, and are the actual MODIS images, respectively.

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 input data.

For the convenience of comparison, the different fusion methods (except OPDL) were grouped into three types, origin weighting (Eq. (18)), increment weighting (Eq. (19)), and regression weighting methods here (Eq. (20)):

$$\hat{F}_2(x,y) = \sum_i w_i C_{2i}\,, \tag{18}$$

$$\hat{F}_2(x,y) = F_1(x,y) + \sum_i w_i \left(C_{2i} - C_{1i}\right), \tag{19}$$

$$\hat{F}_2(x,y) = \sum_i w_i \left(a \times F_{1i} + b + r_i\right), \tag{20}$$

where $F_1(x_i, y_i)$, $C_1(x_i, y_i)$, $C_2(x_i, y_i)$, and $r(x_i, y_i)$ are denoted as F_{1i}, C_{1i}, C_{2i}, and r_i for simplification, respectively. UBDF is a typical origin weighting method. As shown in Eq. (18), the fused result is calculated by weighting different coarse pixels acquired at t_2; the w_i is calculated by $\left[\mathbf{f}^{\mathrm{T}}(\mathbf{M})\mathbf{f}(\mathbf{M})\right]^{-1}\mathbf{f}^{\mathrm{T}}(\mathbf{M})$ for UBDF. LMGM, STARFM, and FSDAF belong to increment weighting methods. As shown in Eq. (19), 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. (20), 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 convenient theoretical analysis, Eq. (28) was further simplified by replacing r_i with $C_{2i} - aC_{1i} - b$

$$\begin{aligned}\hat{F}_2(x,y) &= \sum_i w_i \left(a \times F_{1i} + b + C_{2i} - aC_{1i} - b\right) \\ &= \sum_i w_i \left(a\left(F_{1i} - C_{1i}\right) + C_{2i}\right)\end{aligned} \quad . \tag{21}$$

4.1 Propagation of geometric errors to fusion results

As the fine images were considered as the 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 pixel induced at t_1 and t_2 are assumed as random variables of 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 ($\delta \hat{F}_2^G$) induced by geometric misregistration could be estimated based on the error propagation equation. For UBDF, the fusion uncertainty of UBDF induced by geometric errors could be derived as:

$$\mathrm{std}\left(\delta \hat{F}_2^G(x,y)\right) = \sqrt{\sum_i w_i^2 \,\mathrm{var}\left(\delta C_2\right)}\,, \tag{22}$$

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

$$\begin{aligned}\mathrm{std}\left(\delta\hat{F}_2^G(x,y)\right) &= \sqrt{\sum_i \mathrm{var}\left(w_i\delta\left(C_2-C_1\right)\right)} \\ &= \sqrt{\sum_i w_i^2\left(\mathrm{var}\left(\delta C_2\right)+\mathrm{var}\left(\delta C_1\right)-2\,\mathrm{cov}(\delta C_1,\delta C_2)\right)},\end{aligned} \tag{23}$$

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 $\mathrm{cov}(\delta C_1,\delta C_2)$ approaches zero. Thus, $\mathrm{var}\left(\delta\left(C_2-C_1\right)\right)$ is larger than $\mathrm{var}\left(\delta C_2\right)$ 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 fusion error of Fit-FC could be also derived as:

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

As a is the regression coefficient between C_1 and C_2, $\mathrm{var}\left(\delta\left(aC_{1i}\right)\right)$ is strongly correlated with $\delta \mathrm{C}_2$. Thus, $\mathrm{var}\left(\delta\left(C_{2i}-aC_{1i}\right)\right)$ is smaller than $\mathrm{var}\left(\delta C_2\right)$ because of error compensation, which could also be shown in Fig. 16. Therefore, Fit-FC is the most robust method for geometric error.

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

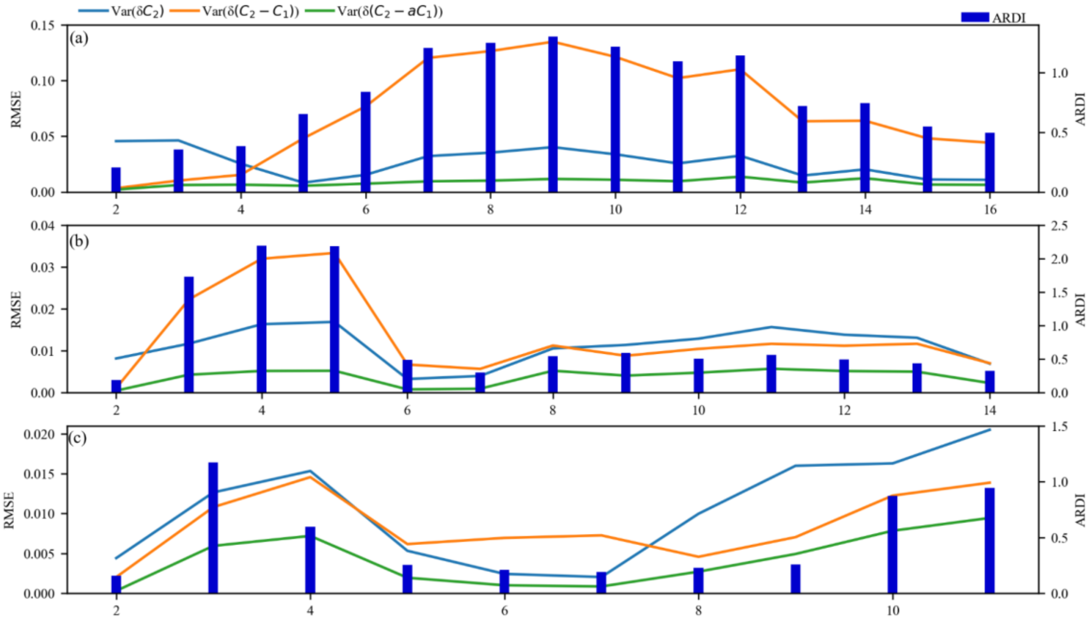

Fig. 16. Comparison of variances of three weighting terms in the two sites: (a) CIA; (b) GWY; (c) Tianjin, image number is the number of the predicted image in the image time series. Fine images were shifted 8 pixels before aggregation.

4.2 Propagation of radiometric error to fused result

As radiometric inconsistency is usually a systematic error, linear stretch was used to express radiometric inconsistency. Thus, the fusion error of UBDF induced by radiometric inconsistency could be derived as:

$$\begin{aligned}\Delta\hat{F}_2^R(x,y) &= \sum_i w_i(\alpha C_{2i}+\beta)-\sum_i w_i C_{2i} \\ &= \sum_i w_i\left((\alpha-1)C_{2i}+\beta\right)\end{aligned}, \tag{25}$$

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

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

$$\begin{aligned}\Delta\hat{F}_2^R(x,y) &= \sum_i w_i\left[(\alpha C_{2i}+\beta)-(\alpha C_{1i}+\beta)\right]-\sum_i w_i(C_{2i}-C_{1i}) \\ &= \sum_i w_i(\alpha-1)(C_{2i}-C_{1i})\end{aligned} \quad . \tag{26}$$

Therefore, the intercept term (β) is removed in the term of ΔC. Theoretically, these three methods are less sensitive to radiometric inconsistency compared to UBDF. However, STARFM shows high sensitivity to radiometric inconsistence in the experiments (Fig. 11 and Fig. 12), which is somehow inconsistent with above theoretical analysis. It 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. (11)), which is sensitive to radiometric inconsistence. If the weight calculation in original STARFM is modifed as that in Fig-FC, the modified STARFM will be also robust to radiometric inconsistency as the theoretical analysis (Fig. 17).

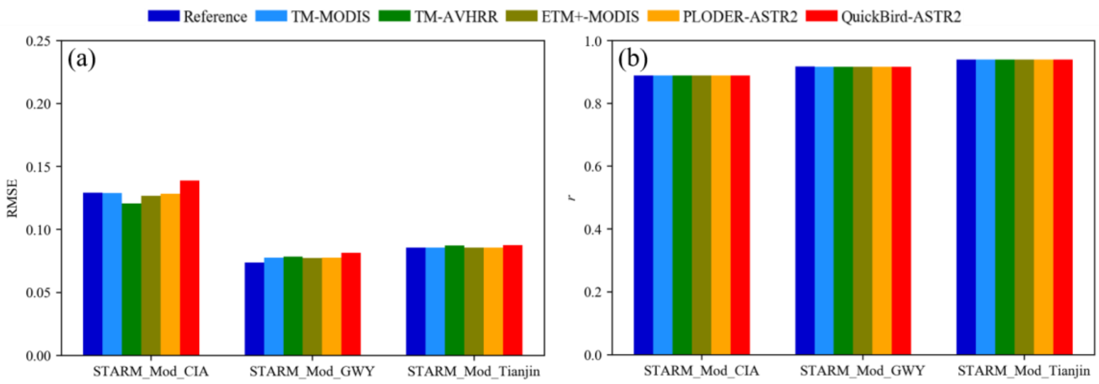

Fig. 17. Quantitative comparison of NDVI fusion results of STARFM_Mod_Site (i.e., results of modified STARFM in the different site) under different levels of radiometric inconsistencies. (a) RMSE; (b) *r*. Reference means that there is no radiometric inconsistency.

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

$$
\begin{aligned}
\Delta \hat{F}_2^R(x,y) &= \sum_i w_i \left(a\left(F_{1i} - \alpha C_{1i} - \beta\right) + \alpha C_{2i} + \beta \right) - \sum_i w_i \left(a\left(F_{1i} - C_{1i}\right) + C_{2i} \right) \\
&= \sum_i W_i \left((\alpha - 1)\left(C_{2i} - aC_{1i}\right) + (1 - a)\beta \right)
\end{aligned} \quad . \tag{27}
$$

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 intercalibration of coarse and fine images.

4.3 Influence of spatial resolution ratio on spatiotemporal fusion

The spatial resolution ratio of sensors determines the information gap between coarse and fine images acquired at the same time. In other words, with the spatial resolution ratio increasing, coarse pixels contain more fine pixels and, thus, become more mixed; this is a similar effect as the increase of spatial heterogeneity. Thus, those methods that perform relatively better in heterogonous 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. It is because only two land cover types (i.e., vegetation and non-vegetation) need to be considered in 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 in 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 NDVI data than reflectance data.

4.4 Method selection and guidance for future method design

The above comparison and analyses can guide the selection of suitable spatiotemporal fusion methods in applications. Other than 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 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 applications. Similar to a recent comparative study (Liu et al., 2019), 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 summed up in Table 7. The most recommended algorithm is Fit-FC, which can produce accurate results with high efficiency for NDVI fusion. However, 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 geometric misregistration can be well corrected.

Table 7 The pros and cons of six typical fusion methods under comparison of different influential factors (worst: 1, good: 2~4, best: 5). Due to the dominant of 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

This study can also give guidance for the future development of spatiotemporal fusion methods. Previous developments of spatiotemporal fusion methods were generally designed without the consideration of inevitable geometric and radiometric errors. For example, increment weighting (Eq. (19)) is commonly used in a large group of fusion methods (e.g., STARFM, LMGM, and FSDAF) as it can keep good spatial details and reduce the radiometric inconsistency of sensors to some extent. However, the above analysis indicates that it would be highly sensitive to geometric error. In contrast, the regression model 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 has not completely considered all the influential factors. The geometric and radiometric errors were simply simulated by pixel shifting and linear transformation in this study. However, there are more complicated errors between sensors, including complex geometric errors from imagery scaling, rotation, and skewing (Dai and Khorram, 1998; Toutin, 2004) and radiometric inconsistency caused by nonlinear distortion, such as analogous bands between sensors with different spectral response functions (SRFs), radiometric resolution difference and the angle effect that solar-sensor geometry bidirectional reflectance distribution function (BRDF) 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 of the spatiotemporal fusion methods due to limitations of the source code availability and heavy works. Notwithstanding the representative methods as much as possible that we selected, the better methods are probably missed. An organization of programming contest with a standard dataset and assessment protocol could be a solution to engaging more algorithm developers and a fair comparison of different spatiotemporal fusion methods in near future.

5. Conclusions

Besides the spatiotemporal variations of 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 applications. These influencing factors could affect different fusion methods to different degrees. The simulation experiment and the theoretical analysis showed that Fit-FC achieved the best performances 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 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. 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 to 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.

Acknowledgement

This study is supported by National Natural Science Foundation of China (No. 67441830108 and No. 41871224). [ZJ3]

Appendix

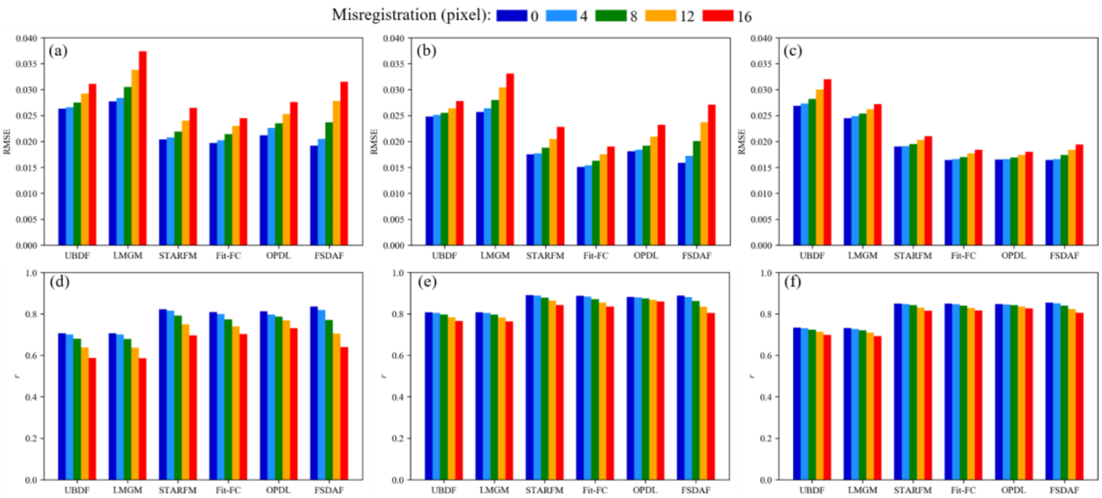

Fig. A1. Quantitative comparison of green band fusion results under different levels of geometric errors from 0 to 16 (misregistration pixel). (a) RMSE in CIA; (b) RMSE in GWY; (c) r in CIA; (d) r in GWY.

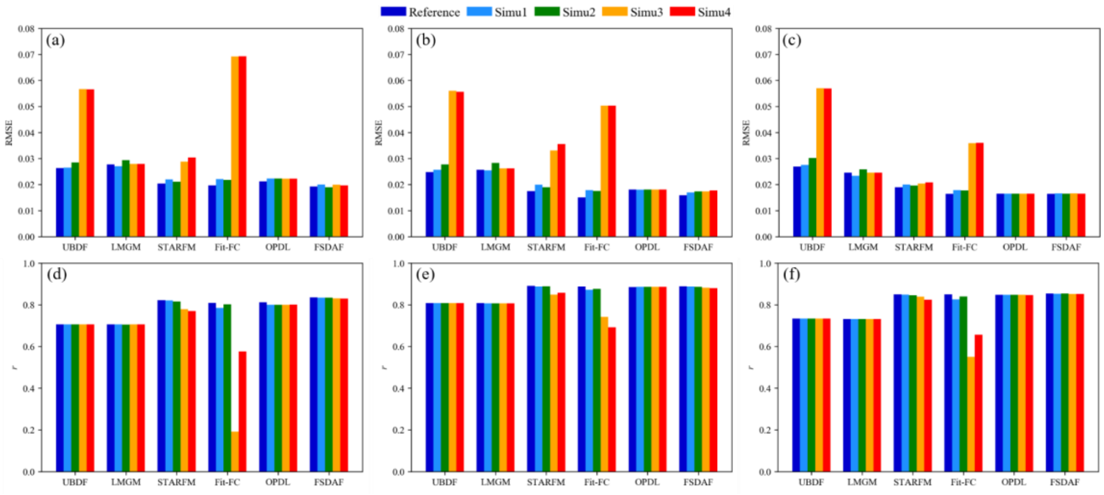

Fig. A2. Quantitative comparison of green band fusion results under different levels of radiometric inconsistencies. (a) RMSE in CIA; (b) RMSE in GWY; (c) r in CIA; (d) r in GWY. Reference means that there is no radiometric inconsistency.

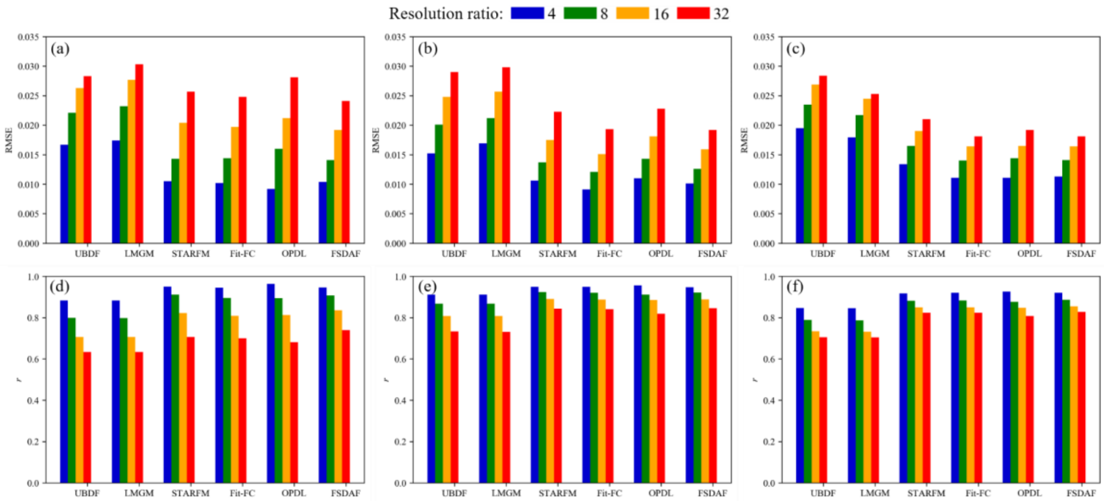

Fig. A3. Quantitative comparison of green band fusion results under different levels of spatial resolution ratio from 4 to 32. (a) RMSE in CIA; (b) RMSE in GWY; (c) r in CIA; (d) r in GWY.

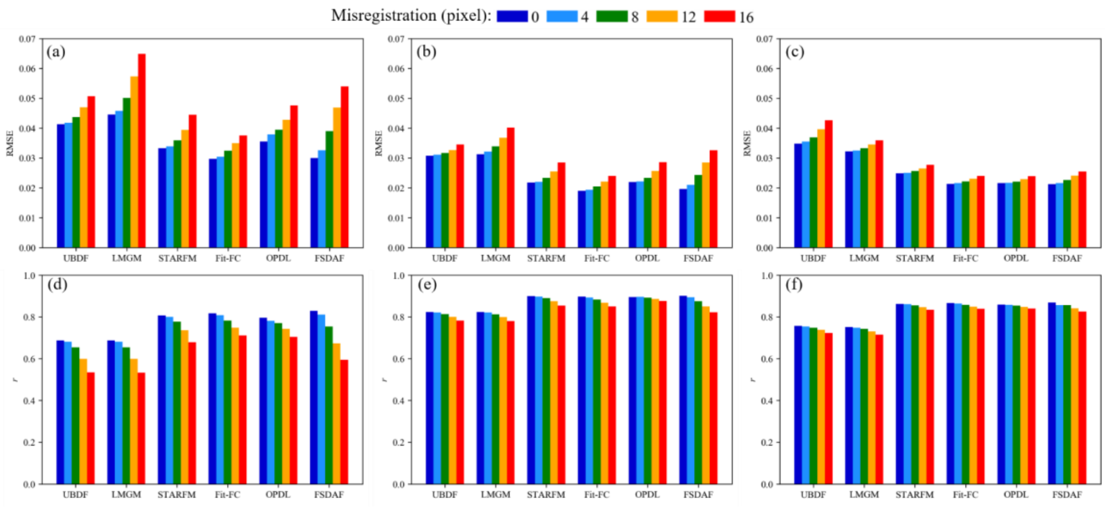

Fig. A4. Quantitative comparison of red band fusion results under different levels of geometric errors from 0 to 16 (misregistration pixel). (a) RMSE in CIA; (b) RMSE in GWY; (c) r in CIA; (d) r in GWY.

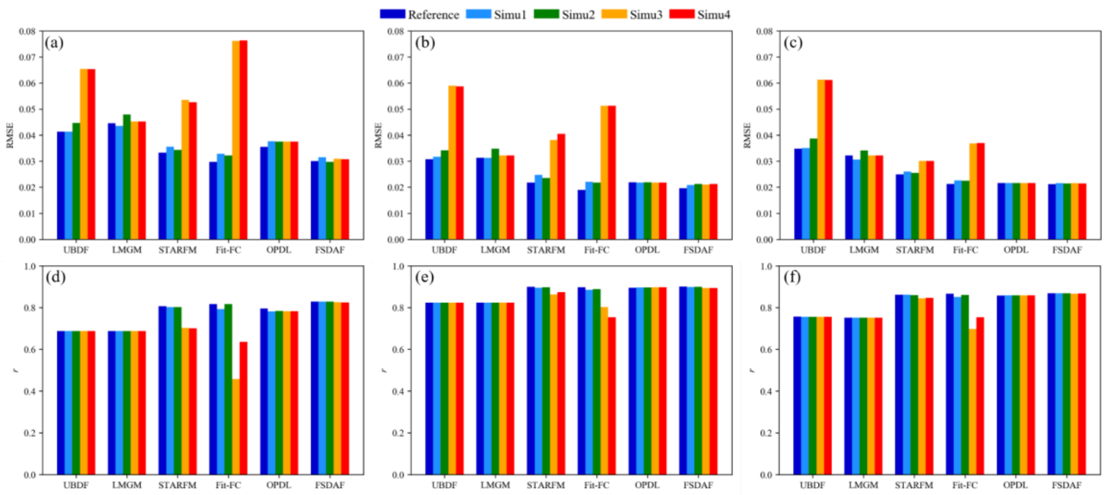

Fig. A5. Quantitative comparison of red band fusion results under different levels of radiometric inconsistencies. (a) RMSE in CIA; (b) RMSE in GWY; (c) r in CIA; (d) r in GWY. Reference means that there is no radiometric inconsistency.

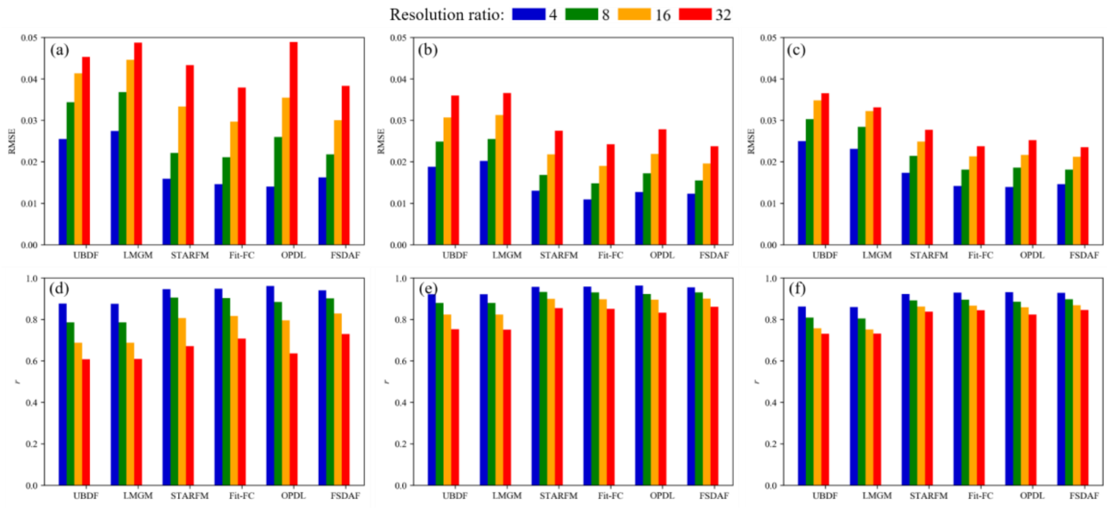

Fig. A6. Quantitative comparison of red band fusion results under different levels of spatial resolution ratio from 4 to 32. (a) RMSE in CIA; (b) RMSE in GWY; (c) *r* in CIA; (d) *r* in GWY.

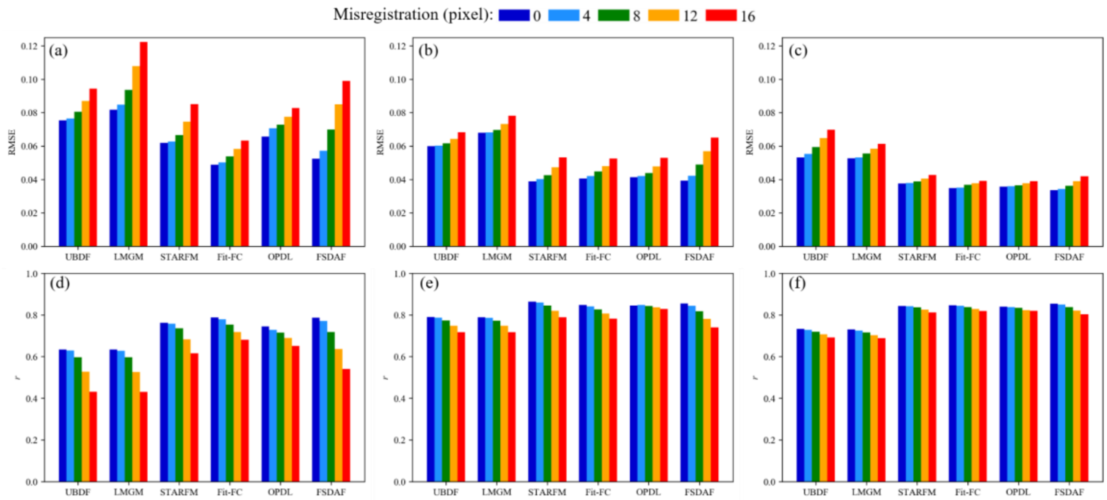

Fig. A7. Quantitative comparison of NIR band fusion results under different levels of geometric errors from 0 to 16 (misregistration pixel). (a) RMSE in CIA; (b) RMSE in GWY; (c) *r* in CIA; (d) *r* in GWY.

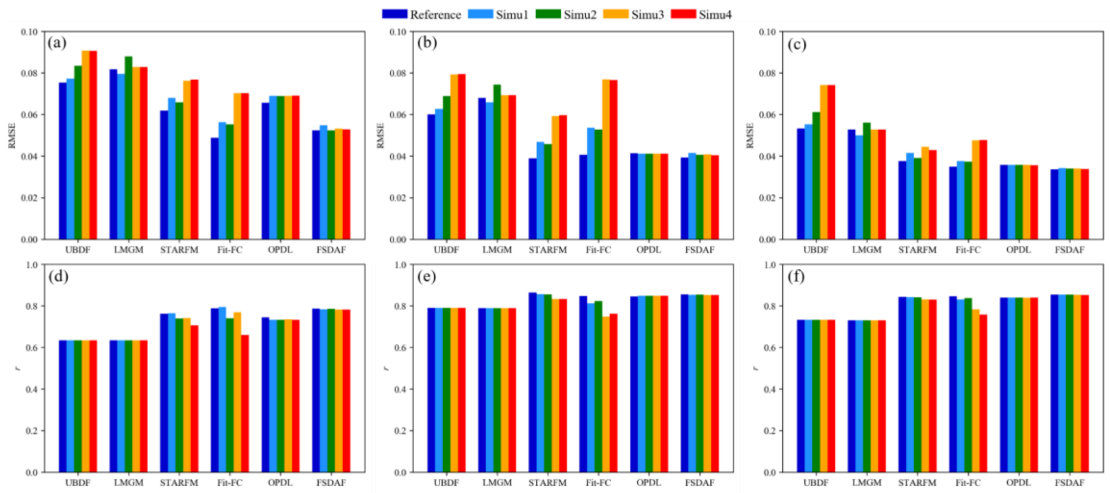

Fig. A8. Quantitative comparison of NIR band fusion results under different levels of radiometric inconsistencies. (a) RMSE in CIA; (b) RMSE in GWY; (c) *r* in CIA; (d) *r* in GWY. Reference means that there is no radiometric inconsistency.

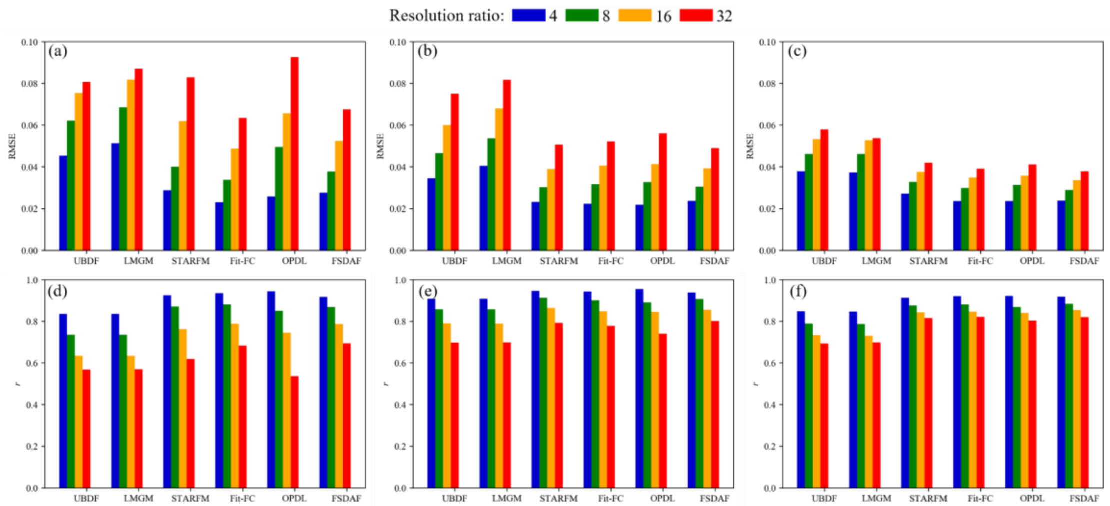

Fig. A9. Quantitative comparison of NIR band fusion results under different levels of spatial resolution ratio from 4 to 32. (a) RMSE in CIA; (b) RMSE in GWY; (c) *r* in CIA; (d) *r* in GWY.

Reference

Aharon, M., Elad, M., Bruckstein, A., 2006. K-SVD: An algorithm for designing over-complete dictionaries for sparse representation. IEEE Trans. Signal Process. 54, 4311-4322. https://doi.org/10.1109/TSP.2006.881199

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

Belgiu, M., Stein, A., 2019. Spatiotemporal image fusion in remote sensing. Remote sensing. 11, 818.

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. Agr. Forest 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.

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.

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.

Chen, B., Huang, B., Xu, B., 2015. Comparison of spatiotemporal fusion models: A review. Remote Sens. 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, 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.

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.

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.

Davis, G., Mallat, S., Avellaneda, M., 1997. Adaptive greedy approximations. Constructive Approx. 13, 57-98. https://doi.org/10.1007/BF02678430

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.

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.

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.

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.

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. 6, 11518-11532. https://doi.org/10.3390/rs61111518

Jun, L., Yunfei, L., Lin, H., Jin, C., Antonio, P., 2020. Spatio-Temporal Fusion for Remote Sensing Data: An overview and New Benchmark. Sci. China Inf. Sci.

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. 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, 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.

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, Northwest China. Agr. Forest Meteorol. 244, 82-97. https://doi.org/10.1016/j.agrformet.2017.05.023

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. 9, 1125. https://doi.org/10.3390/rs9111125

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. 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, M., Ke, Y., Yin, Q., Chen, X., Im, J., 2019. Comparison of Five Spatio-Temporal

Satellite Image Fusion Models over Landscapes with Various Spatial Heterogeneity and Temporal Variation. Remote Sens. 11, 2612. https://doi.org/10.3390/rs11222612

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.

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

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.

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. 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.

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

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.

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.

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.

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.

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

Townshend, J. R., Justice, C. O., Gurney, C., McManus, J., 1992. The impact of misregistration on change detection. IEEE Trans. Geosci. Remote Sens. 30, 1054-1060.

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

Walker, J. J., De Beurs, K. M., Wynne, R. H., Gao, F., 2012. Evaluation of Landsat and MODIS data fusion products for analysis of dryland forest phenology. Remote Sens. Environ. 117, 381-393. https://doi.org/10.1016/j.rse.2011.10.014

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.

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 Sensing. 8, 520.

Yang, J., Wright, J., Huang, T. S., Ma, Y., 2010. Image super-resolution via sparse representation. IEEE Trans. Image Process. 19, 2861-2873. https://doi.org/10.1109/TIP.2010.2050625

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.

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. 8, 10. https://doi.org/10.3390/rs8010010

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

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 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., Cai, F., Tian, J., Williams, T., 2018. Spatiotemporal fusion of multisource remote sensing data: literature survey, taxonomy, principles, applications, and future directions. Remote Sens. 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.

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

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.

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.

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.

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.

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.

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.

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.

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.

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 Sensing. 8, 520.

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.