Front. Earth Sci. 2020, 14(3): 601–614
https://doi.org/10.1007/s11707-019-0800-x

RESEARCH ARTICLE

A modified flexible spatiotemporal data fusion model

Jia TANG[1], Jingyu ZENG[1], Li ZHANG[2], Rongrong ZHANG[1], Jinghan LI[3], Xingrong LI[4], Jie ZOU[5], Yue Zeng[1], Zhanghua Xu[1], Qianfeng WANG (✉)[1,6], Qing ZHANG (✉)[2]

1 Fujian Provincial Key Laboratory of Remote Sensing of Soil Erosion and Disaster Protection, College of Environment and Resources, Fuzhou University, Fuzhou 350116, China
2 Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China
3 College of Geography and Tourism, Anhui Normal University, Wuhu 241000, China
4 College of Geomatics, Shandong University of Science and Technology, Qingdao 266590, China
5 Spatial Information Research Center of Fujian Province, Fuzhou University, Fuzhou 350116, China
6 Joint Global Change Research Institute, Pacific Northwest National Laboratory and University of Maryland, College Park, MD 20740, USA

Abstract Remote sensing spatiotemporal fusion models blend multi-source images of different spatial resolutions to create synthetic images with high resolution and frequency, contributing to time series research where high quality observations are not available with sufficient frequency. However, existing models are vulnerable to spatial heterogeneity and land cover changes, which are frequent in human-dominated regions. To obtain quality time series of satellite images in a human-dominated region, this study developed the Modified Flexible Spatial-temporal Data Fusion (MFSDAF) approach based on the Flexible Spatial-temporal Data Fusion (FSDAF) model by using the enhanced linear regression (ELR). Multiple experiments of various land cover change scenarios were conducted based on both actual and simulated satellite images, respectively. The proposed MFSDAF model was validated by using the correlation coefficient (*r*), relative root mean square error (RRMSE), and structural similarity (SSIM), and was then compared with the Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model (ESTARFM) and FSDAF models. Results show that in the presence of significant land cover change, MFSDAF showed a maximum increase in *r*, RRMSE, and SSIM of 0.0313, 0.0109 and 0.049, respectively, compared to FSDAF, while ESTARFM performed best with less temporal difference of in the input images. In conditions of stable landscape changes, the three performance statistics indicated a small advantage of MFSDAF over FSDAF, but were 0.0286, 0.0102, 0.0317 higher than for ESTARFM, respectively. MFSDAF showed greater accuracy of capturing subtle changes and created high-precision images from both actual and simulated satellite images.

Keywords MFSDAF, enhanced linear regression, land cover change, heterogeneous, time-series

1 Introduction

Remote sensing images with high spatial resolution and temporal frequency are essential for many applications, such as land cover mapping, vegetation phenology monitoring, and retrieval of physicochemical parameters (Townshend et al., 2012; Zhang et al., 2013; Roy et al., 2014; Weng et al., 2014; Zhang et al., 2014). There are two methods for obtaining time series of high/medium-spatial resolution remote sensing images. The first is through direct Earth Observations with satellites with both high spatial and temporal resolution (Wu et al., 2018; Zhao et al., 2018), such as Quick Bird, RapidEye, Worldview, and Sentinel-2. The second is by applying spatiotemporal fusion models to multi-source remote sensing images (MODIS, Landsat, Sentinel, GF, etc.) and then generating synthetic images with high spatiotemporal resolution (Gao et al., 2006; Zhu et al., 2010; Xie et al., 2016; Wang et al., 2017a; Cui et al., 2018; Wang and Atkinson, 2018; Wu et al., 2018). This approach takes advantage of different satellite sensors and overcomes the trade-offs between their spatial and temporal resolutions (Xue et al., 2017). The first method still faces challenges in acquiring specific images. Due to both persistent cloud cover and shadow contamination found in certain geographic regions, new satellites are limited in the frequency with which they can acquire successive cloud-free images (Wang and Atkinson,

Received July 2, 2019; accepted October 8, 2019

E-mails: wangqianfeng@fzu.edu.cn (Qianfeng WANG); zhangqing01@radi.ac.cn (Qing ZHANG)

2018). In addition, the availability of historical images with high spatial resolution is limited. Moreover, most images with high spatiotemporal resolution can only be obtained through commercial satellites making them cost prohibitive to most researchers. A more cost effective method for creating synthetic images has been achieved however by blending multi-source remote sensing images, thus increasing the availability of images with various spatial and temporal resolutions (Wang and Huang, 2018). Fully mining this information is of great significance for a number of applications and requires a robust remote sensing spatial-temporal fusion model.

The spatiotemporal fusion of remote sensing images aims to fully use both the spatial structure information provided by images with high spatial resolution, but low frequency (hereafter referred to as higher-resolution images), and the temporal change information offered by images with high temporal frequency, but coarse spatial resolution (hereafter referred to as lower-resolution images), to synthesize images with both high spatial resolution and temporal frequency (hereafter referred to as fusion images) (Zhao et al., 2018). Spatiotemporal fusion models can be divided into three categories: weighted function-based models, unmixing-based models, and learning-based models (Zhu et al., 2016). Among the weighted function-based models, the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) (Gao et al., 2006) has been widely used for the spatiotemporal fusion of remote sensing data (Emelyanova et al., 2013), and has spawned the development of many improved models, including the enhanced STARFM (ESTARFM) (Zhu et al., 2010), the integrated STARFM (Wang et al., 2014a), and the Spatial-temporal Integrated Temperature Fusion Model (Wu et al., 2015). However, these models generally ignore any land cover changes taking place during the fusion process. To solve this problem, new approaches and models have been developed, including an approach that optimizes the input images of the model through the medium-resolution images (Wang et al., 2017b), the Fit-FC algorithm (Wang and Atkinson, 2018), a rigorously weighted spatial-temporal fusion model (Wang and Huang, 2017), and the Robust Adaptive Spatial and Temporal Fusion Model (Zhao et al., 2018).

Two models unmixing-based models commonly used for heterogeneous areas, including an improved STARFM (Xie et al., 2016) and an improved Spatial and Temporal Data Fusion Approach (Wu et al., 2016). Models applicable to land cover change include the unmixing-based Spatial-temporal Reflectance Fusion Model (Huang and Zhang, 2014), the FSDAF approach (Zhu et al., 2016), and an integrated framework of spatiotemporal temperatures blending (Quan et al., 2018).

Recently developed learning-based models can also take gradual and abrupt changes in land cover into account during the fusion process. These models include the sparse-representation-based Spatial-temporal Reflectance Fusion Model (Song and Huang, 2013), the Hierarchical Spatial-temporal Adaptive Fusion Model (Chen et al., 2017), the extreme learning machine model (Xun et al., 2017), in addition to deep learning approaches (Das and Ghosh, 2016).

Both advantages and disadvantages have been observed in all three categories of spatiotemporal fusion models. Learning-based models are more complex and, hence, require abundant computer system resources. Among the weighted function-based and unmixing-based models, ESTARFM and FSDAF are more commonly used. ESTARFM has been shown to be effective in heterogeneous areas (Knauer et al., 2016) while FSDAF is more efficient with gradual changes and land cover type changes (Zhu et al., 2016). However, ESTARFM and FSDAF are only suitable for areas covered by natural vegetation or regions with little disturbance from human activity where the extent of change is recognized through lower-resolution images. Therefore, it is necessary to develop a new fusion model that can be applied to human-dominated regions with better performance and accuracy.

At present, linear regression models have been widely utilized in spatiotemporal fusion methods due to their simple theoretical basis. Some studies have assumed that the fitting coefficients of the linear models remain constant while other studies tried to calculate the coefficients via regression equations. For example, the STARFM and ESTARFM regarded the coefficient as a constant (Gao et al., 2006; Zhu et al., 2010). The spatial and temporal nonlocal filter-based fusion model and the Fit-FC algorithm solved the regression equations using the least square method and regression model fitting, respectively (Cheng et al., 2017; Wang and Atkinson 2018). However, this assumption resulted in inevitable errors when the temporal change of the landscape took place between the acquisitions of the fused images (Ping et al., 2018). To avoid these errors, Ping et al. (2018) proposed an Enhanced Linear Spatio-temporal Fusion Method which treats the coefficients as variables. However, the performance of the model is reduced in the presence of land cover change, which limits its application in human-dominated regions where both phenological and land cover change are frequent.

In this study, we developed a modified FSDAF model (MFSDAF) by combining the idea of the enhanced linear regression model (ELR) with the FSDAF approach, with the goal of producing accurate predictions in human-dominated regions with frequent land cover change. Surface reflectance data of Landsat 8 and MODIS were used, and multiple experiments were conducted to compare our proposed method with ESTARFM and FSDAF. By combining the advantages of a linear regression model with FSDAF's approach, the MFSDAF model can be applied for dynamic monitoring of time series in human-dominated regions and provides a quality data source for the spatiotemporal evolution analysis of drought (Wang et al., 2014b; Wang et al., 2015a),

evapotranspiration (Wang et al., 2018a), and net primary productivity (Wang et al., 2018b). It can also support the precise evaluation of crop yield (Wang et al., 2017c) and regional detecting of change points (Wang et al., 2019), which enrich the scope of application of this spatial-temporal fusion method.

2 Study area and data sets

2.1 Study area

We tested our spatiotemporal fusion model over the Xiong'an New Area, the deputy capital of China, located in the center of Hebei province (38°43′N–39°10′N, 115°38′E–116°20′E) (Fig. 1). The area is dominated by low-altitude plains and depressions, and is mainly covered by developed and agricultural lands. Due to variable environmental conditions and agricultural planting patterns (Wang et al., 2015b), there is a wide variance in the vegetation growth rate in the area. Vegetation cover and phenology are susceptible to seasonal variations and human activities (Xu et al., 2017). Dynamically monitoring vegetation is of great significance for ecological conservation and urban vegetation planning in the New Area. In this study, we selected an 18 km×18 km section of the Xiong'an New Area and defined different scenarios of land cover change using vegetation phenology information.

2.2 Data sets

This study was based on the Operational Land Imager (OLI) surface reflectance from the Landsat 8 satellite and MOD09A1 surface reflectance from the Terra satellite. OLI surface reflectance have 30 m spatial resolution and 16-day revisit time. Atmospherically corrected OLI surface reflectance can be downloaded from the United States Geological Survey (USGS) website. The MOD09A1 images are freely available from Level-1 and the Atmosphere Archive and the Distribution System Distributed Active Center with a 500 m spatial resolution and an 8-day temporal resolution. The study area included 7 OLI scenes with less than 10% cloud cover and 46 high quality MOD09A1 scenes for the study area (Fig. 2). In the fusion preprocessing stage, we transformed the sinusoidal projection of the original MODIS images to Albers projection and set the spatial resolution to 450 m through the MODIS Reprojection Tool. We then resampled all MOD09A1 images to OLI images by the nearest neighbor method, and then cropped all images to cover the same extent of the study area. To eliminate the image time gap between OLI and MOD09A1 images, we scaled up the OLI images via the nearest neighbor interpolation to obtain simulated MOD09A1 images. The simulated images ensure an effective comparison and evaluation of the model itself. In addition, we classified the acquisition times of the OLI images according to the phenological curve of vegetation to define different scenarios of land cover change and extracted the normalized difference vegetation index (NDVI) based on actual MOD09A1 images. Based on prior knowledge of the local agricultural planting, the yearly phenological curve of vegetation displayed a double peak due to the double-cropping pattern of wheat and summer maize. Thus, we piecewise fitted the phenological curve using a logistic model (Zhang et al., 2003; Zhang, 2015). The first increasing greenness phase is shown in Fig. 3(a) and the second decreasing greenness phase is

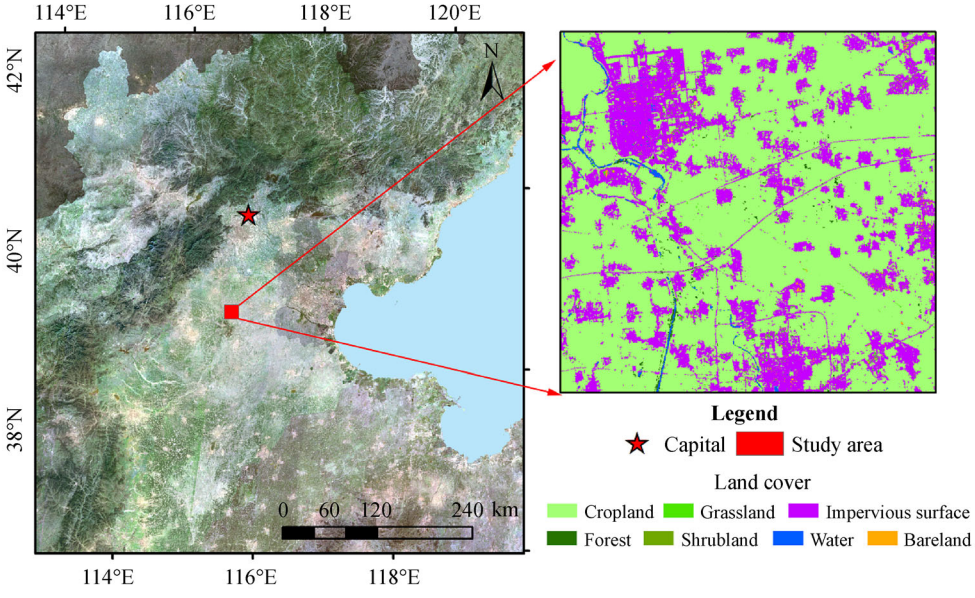


Fig. 1 Location of the study area and the 30 m land cover map provided by Tsinghua University (Available at Tsinghua University website).

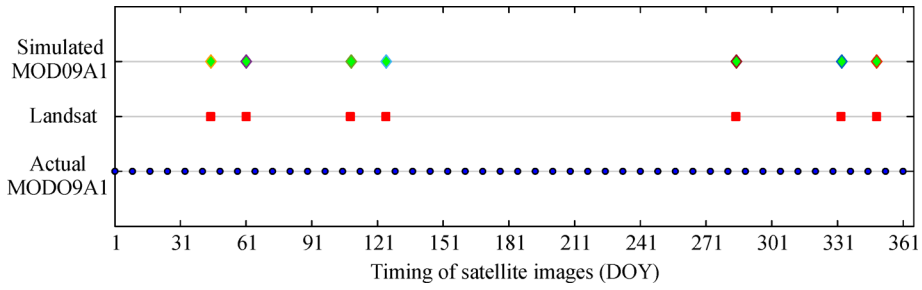


Fig. 2 Acquisition dates of OLI and MOD09A1 images available in 2016 for this study.

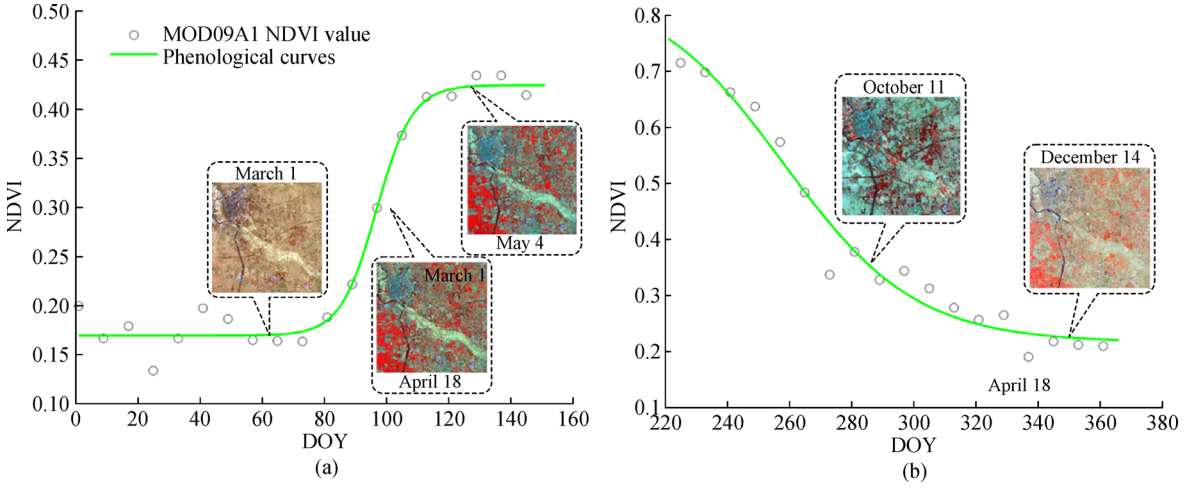


Fig. 3 Phenological curves extracted by actual MOD09A1 NDVI images from 2016 by using a logistic model. (a) First increasing greenness phase; and (b) second decreasing greenness phase; the inset images are the actual OLI images.

shown in Fig. 3(b). From this phenological curve, we were able to classify the different phenological stages of the OLI images.

3 Methodology

3.1 MFSDAF model

The MFSDAF introduced the ELR model on the basis of the FSDAF approach, in which the input images included one pair of higher-resolution and lower-resolution images acquired at the base date and one lower-resolution image acquired at the prediction date. The output included a higher-resolution image of the prediction date. In MFSDAF, we first used the higher-resolution image of the base date to search the similar pixels, then utilized all the input images to obtain the spatially predicted high resolution image at the prediction date $F_{t_2}^{\mathrm{RP}}$ based on ELR. The original FSDAF approach was then used to acquire the final prediction. The flowchart of the MSDAF model is shown in Fig. 4, with a description of the primary steps used stated in the following sections.

3.1.1 Original FSDAF approach

FSDAF is a type of unmixing-based fusion model, which first decomposes the "pure" pixel of the lower-resolution image to obtain the higher-resolution image of temporally predicted ($F_{t_2}^{\mathrm{TP}}$).The image resolution is then increased to obtain the residual of each lower-resolution pixel at the predicted date (R). The residuals are then allocated by utilizing a distribution weight function, which generates the final target image after reducing the block effects using similar pixels in the neighborhood. The distribution weight function can be generated through the homogeneity index (HI) as follows:

$$E_{\mathrm{ho}} = F_{t_2}^{\mathrm{SP}} - F_{t_2}^{\mathrm{TP}}, \tag{1}$$

$$CW = E_{\mathrm{ho}} \times HI + R \times (1 - HI), \tag{2}$$

$$W = \mathrm{normalize}(CW), \tag{3}$$

where $F_{t_2}^{\mathrm{SP}}$ is the spatially predicted higher-resolution image obtained by thin plate spline interpolation. More information can be found in Zhu et al. (2016).

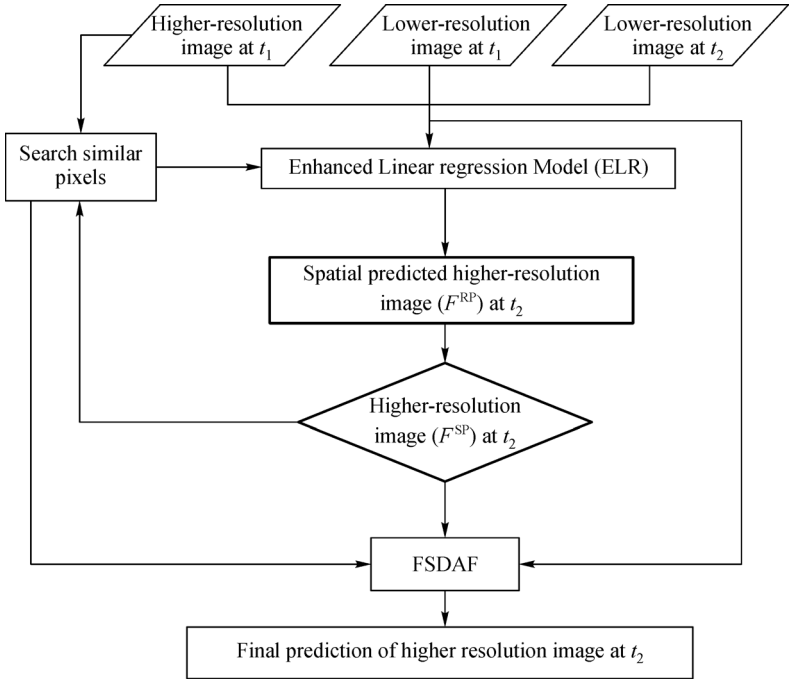


Fig. 4 Flowchart of the MFSDAF algorithm.

3.1.2 Enhanced linear regression

In FSDAF, the interpolation produced the higher-resolution image of the prediction date, but with the potential of introducing a large bias in heterogeneous regions with small scale land cover changes caused by human activities. The over-smoothing effect caused by the interpolation is not enough to be adjusted through the distribution of the residuals. To solve this problem, we replaced the interpolation with the ELR utilized by Ping et al. (2018). The theoretical basis for this is detailed below.

In homogeneous landscapes, the difference in reflectance acquired by different sensors during the same time period and region is primarily rooted in the systematic biases of the sensors. These biases are typically relatively stable (Zhu et al., 2010); hence, the relationship between the reflectance of the higher- and lower-resolution images can beare described as follows:

$$C(x_i, y_j, t_1, B) = a \times F(x_i, y_j, t_1, B) + b, \tag{4}$$

where F and C denote the reflectance of the higher-resolution images and the corresponding lower-resolution images after resampling, respectively; (x_i, y_i) indicates the coordinates of a given pixel; t_1 denotes the base date; a is the slope of the linear model for band B; b is the residual caused by the systematic biases; and a and b are related to atmospheric conditions, solar angle, and altitude. Hence, we consider that a and b remain the same over short time periods. The linear model at t_2 can be expressed as:

$$C(x_i, y_j, t_2, B) = a \times F(x_i, y_j, t_2, B) + b. \tag{5}$$

By combining Eqs. (4) and (5), we obtain:

$$\begin{aligned} F(x_i, y_j, t_2, B) &= \frac{C(x_i, y_j, t_2, B) - b}{a} \\ &= \frac{C(x_i, y_j, t_2, B) - b}{C(x_i, y_j, t_1, B) - b} \times F(x_i, y_j, t_1, B) \\ &= \frac{C(x_i, y_j, t_2, B) - C(x_i, y_j, t_1, B)}{C(x_i, y_j, t_1, B) - b} \\ &\quad \times F(x_i, y_j, t_1, B) + F(x_i, y_j, t_1, B). \end{aligned} \tag{6}$$

According to Eq. (6), an accurately calculation of the reflectance of the higher resolution image at t_2 depends on the simulation of b. Since the influence of solar geometry and sensor bandwidth on each endmember is relatively stable in one coarse pixel (Gao et al., 2006), we can assume that each has identical system residuals, concluding that b is based on the unmixing-theory (Eq. (7));

$$\begin{aligned} &b(x_i, y_j, t_1, B) \\ &\quad = C(X, Y, t_1, B) - \frac{1}{m} \sum_{(i,j) \in (X, Y)}^{m} F(x_i, y_j, t_1, B), \end{aligned} \tag{7}$$

where $C(X, Y, t_1, B)$ is the reflectance of the lower-resolution image (before resampling), and m denotes the number of endmembers inside a given lower-resolution pixel. In this study, we can set m equal to 15×15.

Due to the large discrepancies in mixed-pixel effects

observed in heterogeneous, we tried to obtain the reflectance of higher-resolution images by incorporating a moving window to choose similar neighboring pixels. Thus, Eq. (6) should be modified to incorporate the moving window as follows:

$$F_{t_2}^{RP} = F(x_i, y_j, t_2, B) = \sum_{i=1}^{w}\sum_{j=1}^{w} W_{i,j} \times \left(\frac{C(x_i, y_j, t_2, B) - C(x_i, y_j, t_1, B)}{C(x_i, y_j, t_1, B) - b(x_i, y_j, t_1, B)} \times F(x_i, y_j, t_1, B)) + F(x_i, y_j, t_1, B) \right), \quad (8)$$

where w denotes the moving window size, and $W_{i,j}$ refers to the weight of the similar pixel (x_i, y_j). The selection of similar pixels is based on the principle of minimizing the spectral difference between each pixel and the central pixel in the moving window following Eq. (9):

$$D = \sqrt{\sum_{B=1}^{\text{num\_band}} [F(x_i, y_j, t_1, B) - C(x_i, y_j, t_1, B)]^2 / \text{num\_band}}, \quad (9)$$

where num_band is the total number of bands. The weight of each similar pixel depends on the distance from the center pixel, which can be obtained using the following equations:

$$d_{i,j} = 1 + \sqrt{(x_i - x_{w/2})^2 - (x_i - x_{w/2})^2} / (w/2), \quad (10)$$

$$W_{i,j} = (1/d_{i,j}) / \sum (1/d_{i,j}), \quad (11)$$

where $d_{i,j}$ refers to the spatial distance between the pixel (x_i, y_i) and the central pixel $(x_{w/2}, y_{w/2})$, and $W_{i,j}$ refers to the normalized weight.

By incorporating the above steps, another spatial predicted higher-resolution image $F_{t_2}^{\text{RP}}$ is obtained, which is used to replace $F_{t_2}^{\text{SP}}$.

3.2 Design of multiple experiments

To test the robustness and accuracy of MFSDAF, two different land cover change scenarios were constructed for the following experiments: in scenario 1, the time of the base date and the prediction date were in the same growing phase, while in scenario 2, the time of these dates were in different growing phases (Table 1). Since Zhu et al. (2016) used the simulated lower-resolution images in the original FSDAF, the fusion accuracy using actual lower-resolution images is still unknown. To verify the suitability of the MFSDAF model for actual lower-resolution images and if the accuracy is superior to that of the FSDAF model, we used both actual and simulated lower-resolution images for the experiments. Similarly, since the ESTARFM needs two pairs of higher-resolution and lower-resolution images (one pair before and one after the prediction date), and the temporal difference of the images from the base date to the prediction date has a direct impact on the accuracy (Knauer et al., 2016), we controlled the consistency between the input and output images in each model as much as possible for model intercomparison. In addition, we set another image pair for ESTARFM with the smallest temporal differences. Based on the inset images in Fig. 3, the Landsat image of March 1 was regarded as the higher-resolution image before the growing phase, images of April 18 and May 4 as those during the growing phase, and images of October 11 and December 14 as those at the end of the growing phase. The images from March 1 to May 4 could represent conditions of temporal differences during the entire growth cycle in a human-dominated region, including changes for both the same as well as different growing phases. We have thus concluded that these images are sufficient for the experiments in the two scenarios. Figure 5 shows the NIR-red-green RGB composite of these images.

3.3 Accuracy evaluation indicators

The correlation coefficient (r), the relative root mean square error (*RRMSE*), and the structural similarity (*SSIM*) (Eqs. (12)–(14)) were used to quantitatively evaluate the fusion accuracy of the three models based on pixel values between fusion images and actual images (Wang et al., 2004; Watts et al., 2011; Walker et al., 2012). The r and *RRMSE* are two commonly used accuracy evaluation indicators, and *SSIM* is a visual assessment index used to reflect the similarity of the overall structure between actual and fusion images (Wang et al., 2004). Values of *SSIM* range from 0 to 1, and a value of *SSIM* closer to 1 indicates

Table 1 Design details of the experiments for FSDAF and MFSDAF testing

	Experiment	Base date		Prediction date (Blend)	Test date (Landsat)
		Landsat	MODIS		
Different growing phases	1	2016/03/01	2016/03/01	2016/04/18	2016/04/18
	2	2016/03/01	2016/02/26	2016/04/14	2016/04/18
Same growing phase	3	2016/04/18	2016/04/18	2016/05/04	2016/05/04
	4	2016/04/18	2016/04/14	2016/04/30	2016/05/04

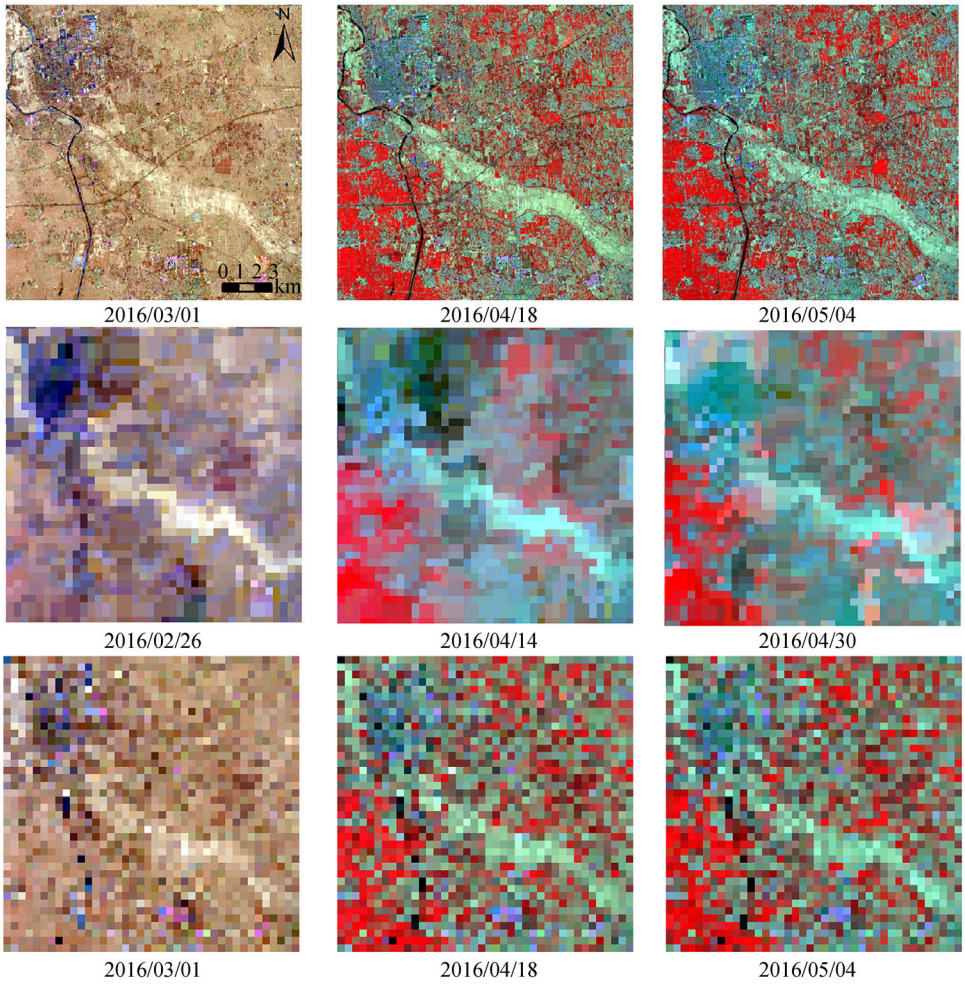


Fig. 5 Images used for four experiments. The first, second, and third rows represent Landsat, actual MOD09A1 and simulated MOD09A1 images, respectively.

greater similarity (He et al., 2017).

$$R = \frac{\mathrm{cov}(X_{\mathrm{obs}}, X_{\mathrm{pre}})}{\sqrt{\mathrm{D}(X_{\mathrm{obs}})}\sqrt{\mathrm{D}(X_{\mathrm{pre}})}}, \tag{12}$$

$$RRMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_{\mathrm{obs},i} - X_{\mathrm{pre},i})^2}{n}} / \overline{X_{\mathrm{obs}}}, \tag{13}$$

$$SSIM = \frac{\left(2\overline{X_{\mathrm{obs}}}\,\overline{X_{\mathrm{pre}}} + C_1\right)\left(2\mathrm{cov}(X_{\mathrm{obs}}, X_{\mathrm{pre}}) + C_2\right)}{\left(\overline{X_{\mathrm{obs}}}^2 + \overline{X_{\mathrm{pre}}}^2 + C_1\right)\left(\mathrm{D}(X_{\mathrm{obs}}) + \mathrm{D}(X_{\mathrm{pre}}) + C_2\right)}, \tag{14}$$

where X_{obs} and X_{pre} denote pixel values of actual and fusion images, $\overline{X_{\mathrm{obs}}}$ and $\overline{X_{\mathrm{pre}}}$ are the corresponding mean values, C is a constant approaching 0, and cov and D denote covariance and variance respectively.

4 Results

4.1 Comparison of the three models for different growing phases

Experiments 1 and 2 were conducted with simulated and actual MOD09A1 images, respectively, to compare the accuracy of the three models under different growing phases. The accuracy results (Table 2) show that fusion results based on actual MOD09A1 images had a lower accuracy compared to those based on simulated MOD09A1 images, primarily due to the systematic errors between different sensors. MFSDAF performed better than FSDAF, with a maximum increase in *r*, *RRMSE*, and *SSIM* of 0.0313, 0.0109, and 0.049, respectively, decreased by 0.0061, 0.0014 and 0.0057 in the NIR band in experiment 1. However, we found greater accuracy in red and NIR bands in ESTARFM compared to MFSDAF, with noted differences in the red band (0.2159, 0.1367, and 0.2322, respectively) for *r*, *RRMSE*, and *SSIM* of the two models.

Table 2 Accuracy assessment of the three models applied to different growing phases

Experiment	Band	ESTARFM			FSDAF			MFSDAF		
		R	*RRMSE*	*SSIM*	*R*	*RRMSE*	*SSIM*	*R*	*RRMSE*	*SSIM*
1	Blue	0.7223	0.2930	0.7150	0.7020	0.2545	0.6810	0.7244	0.2476	0.7122
	Red	0.9429	0.1281	0.9394	0.7063	0.2723	0.6800	0.7270	0.2648	0.7072
	NIR	0.9323	0.0675	0.9314	0.7635	0.1212	0.7527	0.7574	0.1226	0.7470
2	Blue	0.5335	0.3904	0.5287	0.5919	0.3457	0.5187	0.6103	0.3402	0.5542
	Red	0.8676	0.2133	0.8251	0.6421	0.3048	0.5534	0.6734	0.2939	0.6024
	NIR	0.8220	0.1806	0.7710	0.6322	0.1930	0.6242	0.6335	0.1927	0.6246

A possible reason for this is that another input, higher-resolution image (May 4) of ESTARFM could provide additional information on land cover change under conditions of dramatic changes within the landscape. The OLI sensor has 9 bands, while the MODIS sensor has 36 bands, which caused the narrowest blue spectral range of MODIS contrasted to OLI (Table 3). The difference of bandwidth between the two sensors was strongly related to the intensity in spectral response, causing the largest discrepancy and error accumulation in the ESTARFM requiring two pairs of input images. Therefore, we can see from Table 2 that MFSDAF had a better performance in the blue band compared to ESTARFM. To further compare the three models, we drew scatter plots of the reflectance of the fusion image and the actual image obtained in experiment 1 (Fig. 6). The scatter plots of the actual and predicted values from ESTARFM appear more concentrated around the 1:1 line than those of the other two models, while MFSDAF has a better fit than FSDAF, as indicated by the correlation coefficient.

4.2 Comparison of the three models for the same growing phase

During the same growing phase, the results based on simulated MOD09A1 images (experiment 3) had higher accuracy than those generated by actual images (experiment 4), showing a similar trend to the fusion of different growing phases (Table 4). However, the accuracy of experiments 3 and 4 were significantly greater than those of experiments 1 and 2, respectively, which may be explained by the relatively stable landscape change from April 18 to May 4. Except for the NIR band in experiment 4, the accuracies of all bands were greater for MFSDAF, followed by FSDAF and then ESTARFM. The average values of *r*, *RRMSE*, and *SSIM* of MFSDAF were 0.0159, 0.0036, 0.0338 higher, respectively, than for FSDAF for all bands, and 0.0286, 0.0102, 0.0317 higher, respectively, than for ESTARFM for five out six bands. For the NIR band of experiment 4, the accuracy was greater for ESTARFM, followed by MFSDAF and then FSDAF due to vegetation sensitivity to NIR in the growing phase, resulting in a larger system residual. Given there was only one pair of images for input, the influence was greater on the MFSDAF and FSDAF models. The models' performance was also evaluated with scatter plots based on experiment 3 (Fig. 7), showing that the MFSDAF scatter plots of all three bands are closer to the 1:1 line than those of the other two models.

In experiment 4, the image fused by MFSDAF showed more spatial details in comparison to the other two models (see the areas marked by ellipses, circles, and rectangles in Fig. 8). As shown in the enlarged areas in Fig. 9, the fused images generated by FSDAF and ESTARFM are blurry lacking spatial details while MFSDAF preserved the fine spatial details of land cover.

4.3 Comparison of predicted image by MFSDAF with actual image

Taking the Landsat image of May 4 as an example, we concluded that fusion results were accurate when the temporal differences from the base date to predicted date were gradual (Fig. 10). The scatter plots of the three bands are all concentrated around the 1:1 line with *r* up to 0.9470. The concentration is greater in the scatter plot based on the simulated MOD09A1 image (Figs. 11(d)–11(f)), than that of the actual MOD09A1 image (Figs. 11 (a)–11(c)). By zooming into the areas marked by rectangles in Fig. 10 (Fig. 12), we find that MFSDAF can synthesize high-precision real images either with actual or simulated MOD 09A1 images, which fits the actual image to a high degree.

Table 3 Spectral properties of the OLI and MOD09A1 images

Band	MOD09A1		OLI	
	Wavelength range/nm	Wavelength	Wavelength range/nm	Wavelength
Blue	459–479	20	450–515	65
Red	620–670	50	630–680	50
NIR	841–876	35	845–885	40

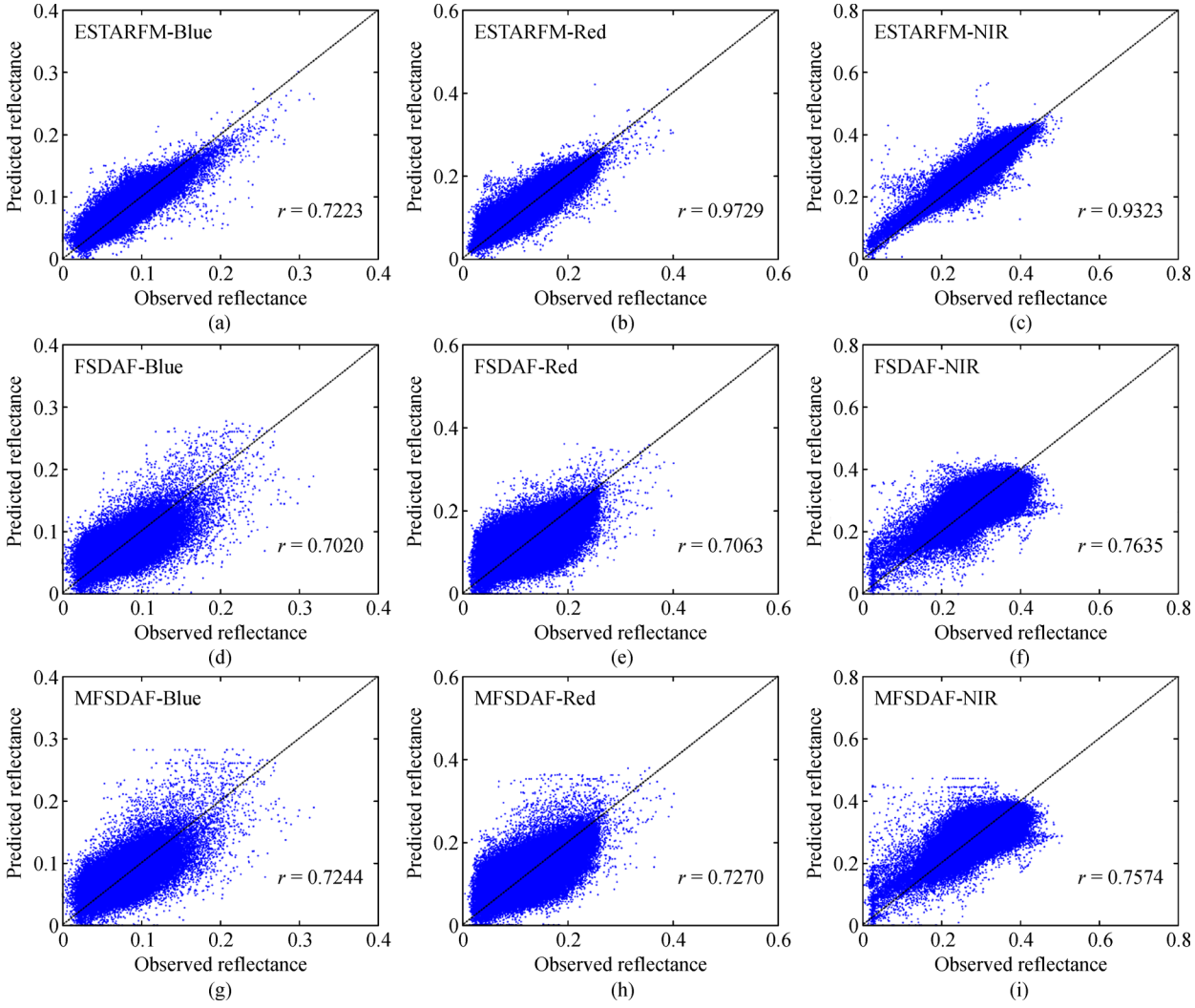


Fig. 6 Scatter plots of the actual and predicted values of the three bands for experiment 1.

Table 4 Accuracy assessment of the three models applied to the same growing phases

Experiment	Band	ESTARFM			FSDAF			MFSDAF		
		R	*RRMSE*	*SSIM*	*R*	*RRMSE*	*SSIM*	*R*	*RRMSE*	*SSIM*
3	Blue	0.9083	0.1533	0.9055	0.9126	0.1489	0.9102	0.9204	0.1456	0.9208
	Red	0.9332	0.1384	0.9320	0.9404	0.1303	0.9384	0.9470	0.1255	0.9471
	NIR	0.9151	0.0907	0.9132	0.9245	0.0678	0.8221	0.9284	0.0668	0.9285
4	Blue	0.8351	0.2006	0.8214	0.8629	0.2111	0.8587	0.8820	0.2060	0.8772
	Red	0.8623	0.1947	0.8567	0.9037	0.1879	0.8996	0.9190	0.1830	0.9139
	NIR	0.8061	0.1336	0.8018	0.6212	0.1764	0.6166	0.6638	0.1742	0.6609

5 Discussion

5.1 Advantages of MFSDAF

Our proposed MFSDAF optimized the original FSDAF approach by using the ELR model. We applied the MFSDAF in a human-dominated region where both phenological change and land cover change occurred during the entire vegetation growing cycle. To test the robustness of MFSDAF, we designed two different scenarios of land cover change to determine the accuracy and applicability of the image-based time series research.

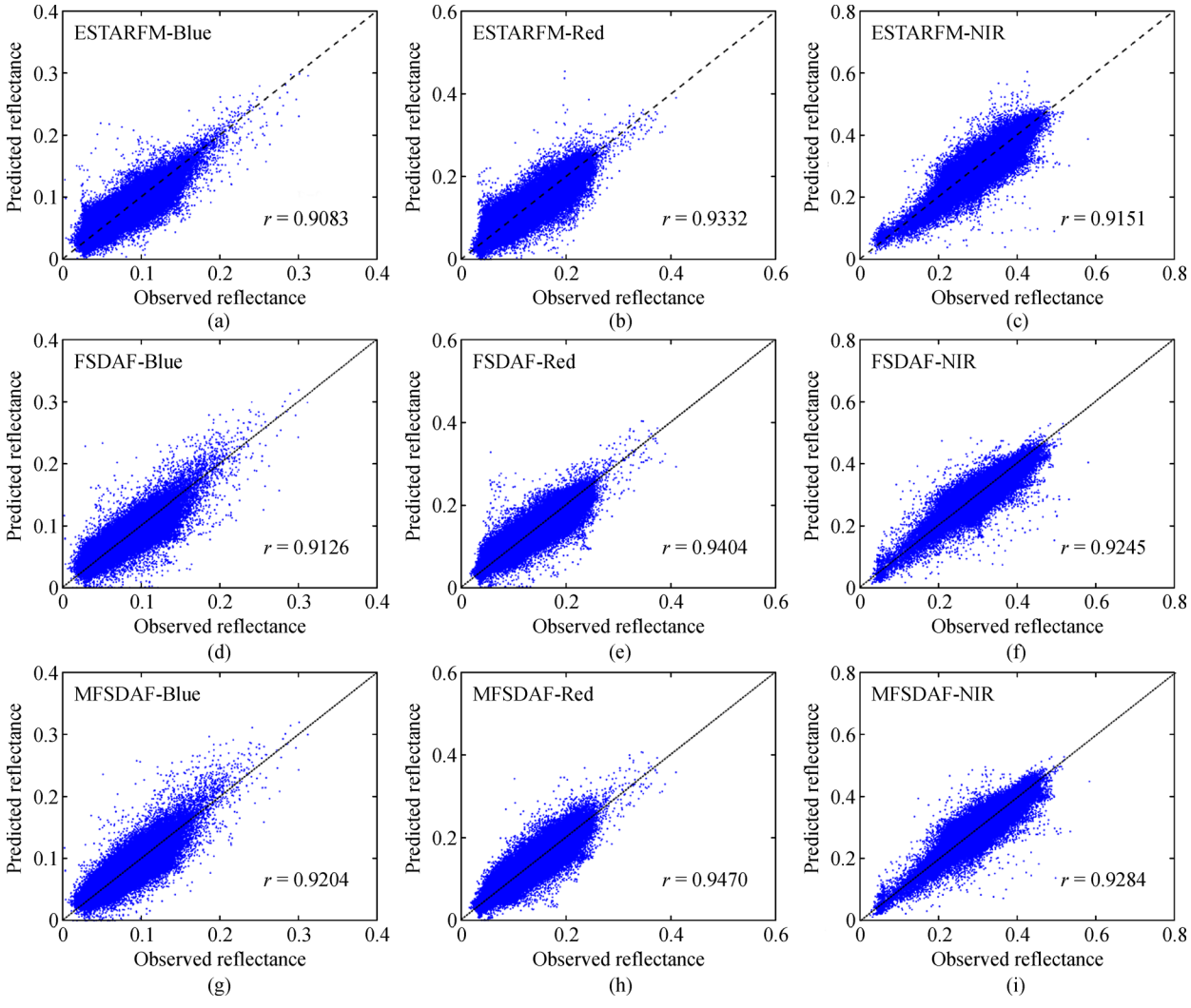


Fig. 7 Scatter plots of the actual and predicted values of the three bands for experiment 3.

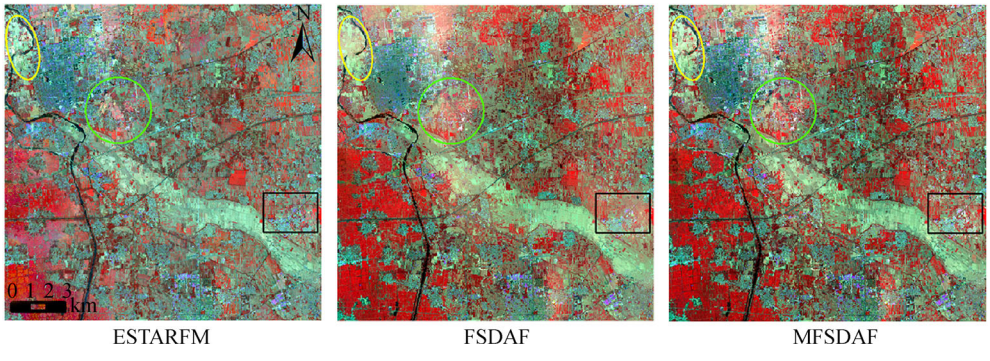


Fig. 8 Comparison of predicted images on May 4, 2016 based on three models.

In previous studies, the accuracy validation was only conducted for the fusion results under gradual land cover change (Zhu et al., 2016). Compared with the other two models used in this study, MFSDAF and FSDAF only needed minimal input, while ESTARFM required two pairs of images as input. In our 4 experiments, MFSDAF showed a better performance both in terms of accuracy and spatial detail recognition in the fusion images compared to FSDAF. Although ESTARFM achieved greater accuracy when the input and prediction images were in different

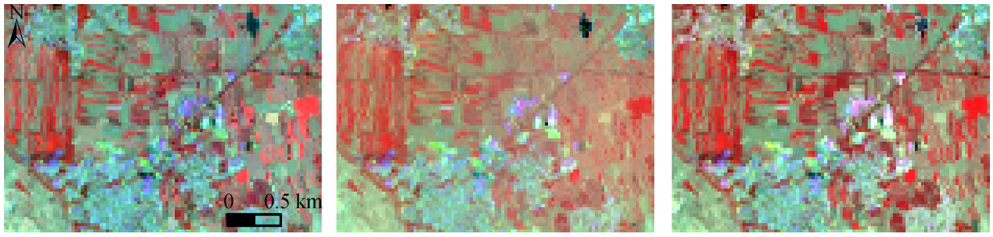


Fig. 9 Zoomed in images of the areas marked by rectangles in Fig. 8.

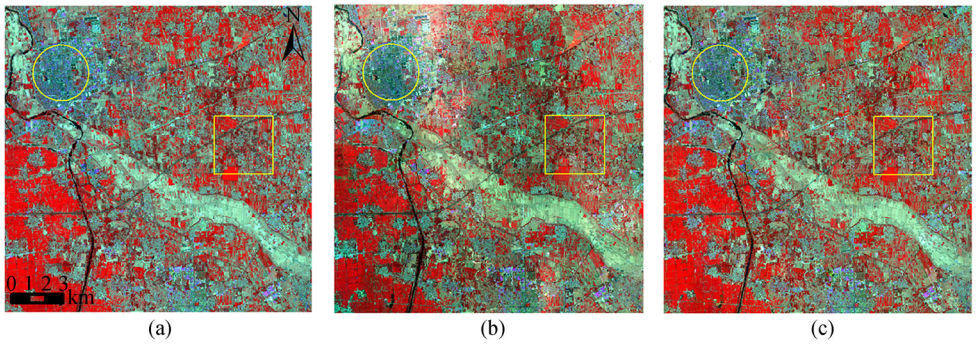


Fig. 10 Landsat image of May 04, 2016. (a) Actual image; (b) predicted image by MFSDAF in experiment 4 (c) predicted image by MFSDAF in experiment 3.

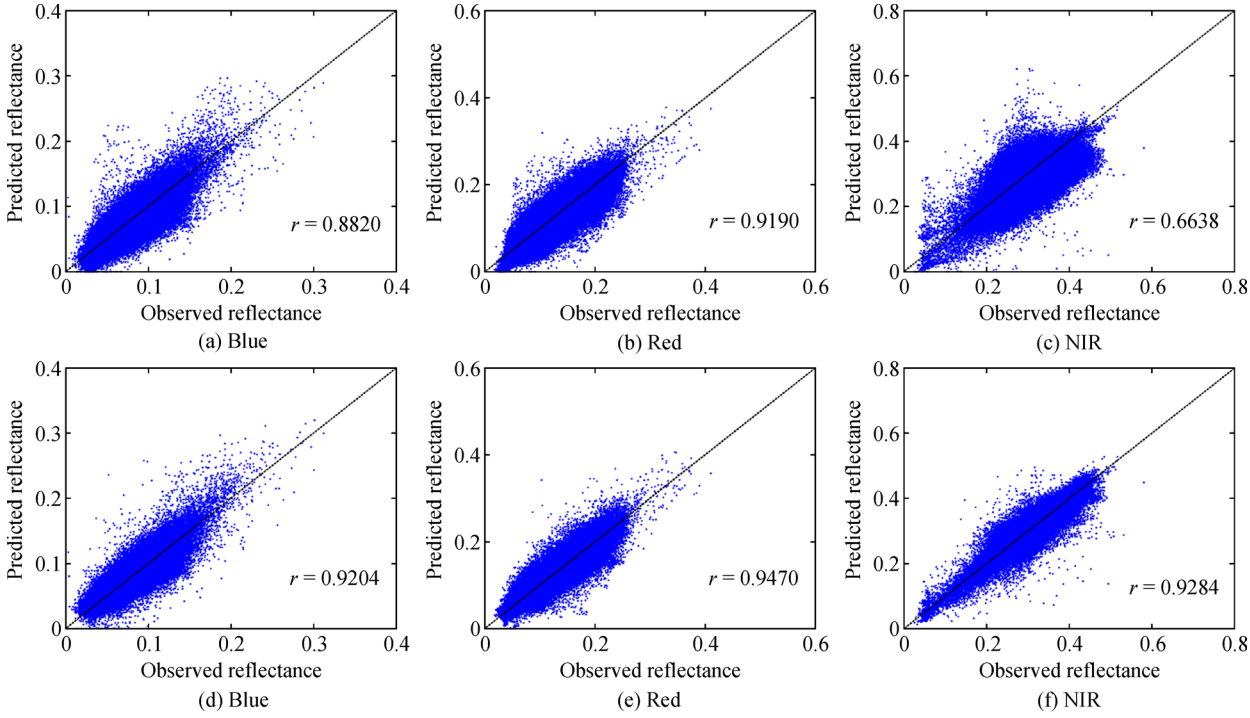
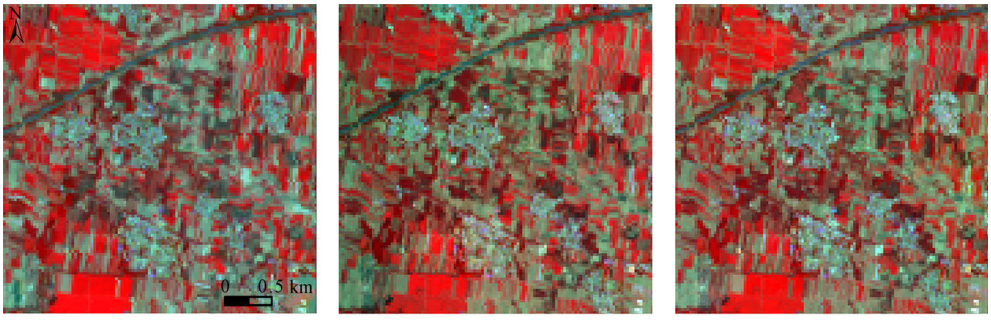


Fig. 11 Scatter plots of the actual and predicted values based on MFSDAF on May 04, 2016. (a)–(c) based on actual MOD09A1 images; (d)–(f) based on simulated MOD09A1 images.



(a) (b) (c)

Fig. 12 Zoomed in images of the areas marked by rectangles in Fig. 10. (a) Actual image; (b) predicted image by MFSDAF in experiment 4; (c) predicted image by MFSDAF in experiment 3.

Table 5 Accuracy assessment of ESTARFM based on different images combinations as input

Prediction date (Blend)	Band	*R*	*RRMSE*	*SSIM*	*R*	*RRMSE*	*SSIM*
2016/04/18	Base date		(2016/03/01, 2016/05/04)			(2016/03/01, 2016/10/11)	
	Blue	0.7223	0.2930	0.7150	0.3837	0.3918	0.3875
	Red	0.9429	0.1281	0.9394	0.5776	0.3157	0.5347
	NIR	0.9323	0.0675	0.9314	0.5672	0.1627	0.5632
2016/05/04	Base date		(2016/04/18, 2016/10/11)			(2016/03/01, 2016/10/11)	
	Blue	0.9083	0.1533	0.9055	0.2995	0.4881	0.2961
	Red	0.9332	0.1384	0.9320	0.5571	0.3222	0.5157
	NIR	0.9151	0.0907	0.9132	0.5182	0.2201	0.5137

growth phases (experiments 1 and 2), the accuracy of ESTARFM was easily affected by the temporal differences of the input images from the base date to the prediction date (Knauer et al., 2016). To test the robustness of ESTARFM, additional experiments under different land cover change were conducted by using different images combinations as input. By comparing the results (Table 5), we noticed that a sharp drop in accuracy occurred when a change in land cover took place, indicating that land cover change has a direct influence on the performance of ESTARFM. Thus, the base and prediction date of high spatial resolution images must be controlled in the same growth phase to ensure the fusion images can be further used. However, this could be challenging given the difficulty in obtaining high spatial resolution images due to a satellite's relatively long revisiting time and cloud contamination. Therefore, the time series applications based on ESTARFM need further development. In summary, MFSDAF showed the greatest potential for generating quality time series fusion images with guaranteed accuracy under different changes of land cover.

5.2 Limitations of MFSDAF

Although extensive efforts were undertaken to improve and validate our proposed models, there are still some limitations that will need to be addressed in the future. First, the accuracy of MFSDAF varied in our four experiments, with the best result obtained for the fusion image blended by simulated images under the same growth phase. Thus, the MFSDAF could synthesize images with high precision under the same growth phase scenario while simulating images with acceptable precision under different growth phase scenarios. The quantitative analysis of the accuracy fluctuation, which is induced by different land cover changes, is missing in this study resulting in the need to develop new scientific approaches in future work. Second, the ELR employed in this study tended to assume the pixel values of images to be related to linear change, which would go against the reality in some human-dominated regions. Therefore, based on our research, a more complex regression model may be needed. Third, although MFSDAF could be suitably applied to actual satellite images, the influence of cloud contamination on accuracy requires further clarification. Finally, although it is theoretically possible to blend other satellite images in addition to Landsat or MODIS using our proposed model, the actual work has still not been carried out. Thus, further additional research that focuses on the comparison of other satellites needs to be conducted (Cui et al., 2018).

6 Conclusions

In this study, we proposed a modified spatiotemporal fusion model based on the FSDAF approach, and designed multiple experiments using both simulated and actual lower-resolution images under two different scenarios of land cover change. The robustness and accuracy of our proposed MFSDAF model were evaluated by multiple indicators and compared with those of the ESTARFM and FSDAF. The results showed the following:

1) For different growing phases, the accuracy of MFSDAF was higher than that of FSDAF, with *r*, *RRMSE*, and *SSIM* increasing by as much as 0.0313, 0.0109 and 0.049, respectively. The result of ESTARFM was better than that of MFSDAF in the red and NIR bands with less temporal difference of input images, while the reverse was observed in the blue band.

2) For images from the same growing phase, the overall accuracy of the three models was higher than that obtained for images in different phases. The average value of *r*, *RRMSE*, and *SSIM* of MFSDAF was 0.0159, 0.0036, 0.0338 higher, respectively, than for FSDAF for all bands. The average value of *r*, *RRMSE*, and *SSIM* of MFSDAF was 0.0286, 0.0102, 0.0317 higher, respectively, than for ESTARFM, except for the NIR band where the fusion accuracy was greater for ESTARFM, followed, in order, by that of MFSDAF and FSDAF.

3) Compared with ESTARFM and FSDAF, MFSDAF requires the minimal inputs and could capture the subtle changes in land cover in a human-dominated region. Furthermore, the image predicted by MFSDAF based on either actual actual or simulated MOD09A1 images fits the actual image to a high degree and can be used for research of time series.

Acknowledgements This research received financial support by the National Natural Science Foundation of China (Grant Nos. 41601562 and 41761014), the National Key Research and Development Program of China (No. 2017YFC1502404), the China Institute of Water Resources and Hydropower Research Team Construction and Talent Development Project (No. JZ0145B752017), the Research Project for Young Teachers of Fujian Province (No. JAT160085).

References

Chen B, Huang B, Xu B (2017). A hierarchical spatiotemporal adaptive fusion model using one image pair. Int J Digit Earth, 10(6): 639–655

Cheng Q, Liu H Q, Shen H F, Wu P H, Zhang L P (2017). A spatial and temporal nonlocal filter-based data fusion method. IEEE Trans Geosci Remote Sens, 55(8): 4476–4488

Cui J T, Zhang X, Luo M Y (2018). Combining linear pixel unmixing and STARFM for spatiotemporal fusion of Gaofen-1 wide field of view imagery and MODIS imagery. Remote Sens, 10(7): 1047

Das M, Ghosh S K (2016). Deep-STEP: a deep learning approach for spatiotemporal prediction of remote sensing data. IEEE Geosci Remote S, 13(12): 1984–1988

Emelyanova I V, McVicar T R, Van Niel T G, Li L T, van Dijk A I J M (2013). Assessing the accuracy of blending Landsat-MODIS surface reflectances in two landscapes with contrasting spatial and temporal dynamics: a framework for algorithm selection. Remote Sens Environ, 133(12): 193–209

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 T Geosci Remote, 44(8): 2207–2218

He C, Zhang Z, Xiong D, Du J, Liao M (2017). Spatio-temporal series remote sensing image prediction based on multi-dictionary Bayesian Fusion. ISPRS Int J Geoinf, 6(11): 374

Huang B, Zhang H (2014). Spatio-temporal reflectance fusion via unmixing: accounting for both phenological and land-cover changes. Int J Remote Sens, 35(16): 6213–6233

Knauer K, Gessner U, Fensholt R, Kuenzer C (2016). An ESTARFM fusion framework for the generation of large-scale time series in cloud-prone and heterogeneous landscapes. Remote Sens, 8(5): 425

Ping B, Meng Y S, Su F Z (2018). An enhanced linear spatio-temporal fusion method for blending landsat and MODIS data to synthesize landsat-like imagery. Remote Sens, 10(6): 881

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

Roy D P, Wulder M A, Loveland T R, C E W, Allen R G, Anderson M C, Helder D, Irons J R, Johnson D M, Kennedy R, Scambos T A, Schaaf C B, Schott J R, Sheng Y, Vermote E F, Belward A S, Bindschadler R, Cohen W B, Gao F, Hipple J D, Hostert P, Huntington J, Justice C O, Kilic A, Kovalskyy V, Lee Z P, Lymburner L, Masek J G, McCorkel J, Shuai Y, Trezza R, Vogelmann J, Wynne R H, Zhu Z (2014). Landsat-8: science and product vision for terrestrial global change research. Remote Sens Environ, 145: 154–172

Song H, Huang B (2013). Spatiotemporal satellite image fusion through one-pair image learning. IEEE Trans Geosci Remote Sens, 51(4): 1883–1896

Townshend J R, Masek J G, Huang C, Vermote E F, Gao F, Channan S, Sexton J O, Feng M, Narasimhan R, Kim D, Song K, Song D, Song X P, Noojipady P, Tan B, Hansen M C, Li M, Wolfe R E (2012). Global characterization and monitoring of forest cover using Landsat data: opportunities and challenges. Int J Digit Earth, 5(5): 373–397

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

Wang H, Pan X, Luo J, Luo Z, Chang C, Shen Y (2015b). Using remote sensing to analyze spatiotemporal variations in crop planting in the North China Plain. Chin J Eco Agric, 23(9): 1199–1209

Wang J, Huang B (2018). A spatiotemporal satellite image fusion model with autoregressive error correction (AREC). Int J Remote Sens, 39 (20): 1–26

Wang J, Huang B (2017). A rigorously-weighted spatiotemporal Fusion model with uncertainty analysis. Remote Sens, 9(10): 990

Wang P, Gao F, Masek J G (2014a). Operational data fusion framework for building frequent landsat-like imagery. IEEE Trans Geosci Remote Sens, 52(11): 7353–7365

Wang Q, Atkinson P M (2018). Spatio-temporal fusion for daily Sentinel-2 images. Remote Sens Environ, 204: 31–42

Wang Q M, Blackburn G A, Onojeghuo A O, Dash J, Zhou L, Zhang Y,

Atkinson P M (2017a). Fusion of Landsat 8 OLI and Sentinel-2 MSI data. IEEE Trans Geosci Remote Sens, 55(7): 3885–3899

Wang Q F, Shi P, Lei T, Geng G, Liu J, Mo X, Li X, Zhou H, Wu J (2015a). The alleviating trend of drought in the Huang-Huai-Hai Plain of China based on the daily SPEI. Int J Biometeorol, 35(13): 3760–3769

Wang Q F, Tang J, Zeng J Y, Qu Y P, Zhang Q, Shui W, Wang W L, Yi L, Leng S (2018a). Spatial-temporal evolution of vegetation evapotranspiration in Hebei Province, China. J Integr Agric, 17(9): 2107–2117

Wang Q F, Tang J, Zeng J Y, Leng S, Shui W (2019). Regional detecting of multiple change points and workable application for precipitation by maximum likelihood approach. Arab J Geosci, 12(23): 745

Wang Q F,Wu J,Lei T,He B,Wu Z,Liu M,Mo X,Geng G,Li X,Zhou H, Liu D (2014b). Temporal-spatial characteristics of severe drought events and their impact on agriculture on a global scale. Quatern int, 349: 10–21

Wang Q F, Wu J, Li X, Zhou H, Yang J, Geng G, An X, Liu L, Tang Z (2017c). A comprehensively quantitative method of evaluating the impact of drought on crop yield using daily multi-scale SPEI and crop growth process model. Int J Biometeorol, 61(4): 685–699

Wang Q F, Zeng J Y, Leng S, Fan B X, Tang J, Jiang C, Huang Y, Zhang Q, Qu Y P, Wang W L, Shui W (2018b). The effects of air temperature and precipitation on the net primary productivity in China during the early 21st century. Front Earth Sci, 12(4): 818–833

Wang Q M, Zhang Y, Onojeghuo A O, Zhu X, Atkinson P M (2017b). Enhancing spatio-temporal fusion of MODIS and landsat data by incorporating 250 m MODIS data. IEEE J Stars, 10(9): 1–8

Wang Z, Bovik A C, Sheikh H R, Simoncelli E P (2004). Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process, 13(4): 600–612

Watts J D, Powell S L, Lawrence R L, Hilker T (2011). Improved classification of conservation tillage adoption using high temporal and synthetic satellite imagery. Remote Sens Environ, 115(1): 66–75

Weng Q, Fu P, Gao F (2014). Generating daily land surface temperature at landsat resolution by fusing landsat and MODIS data. Remote Sens Environ, 145(8): 55–67

Wu M Q, Wu C Y, Huang W J, Niu Z, Wang C Y, Li W, Hao P Y (2016). An improved high spatial and temporal data fusion approach for combining landsat and MODIS data to generate daily synthetic Landsat imagery. Inf Fusion, 31: 14–25

Wu M, Yang C, Song X, Hoffmann W C, Huang W, Niu Z, Wang C, Li W, Yu B (2018). Monitoring cotton root rot by synthetic Sentinel-2 NDVI time series using improved spatial and temporal data fusion. Sci Rep, 8(1): 2016

Wu P, Shen H, Zhang L, Göttsche F M (2015). Integrated fusion of multi-scale polar-orbiting and geostationary satellite observations for the mapping of high spatial and temporal resolution land surface temperature. Remote Sens Environ, 156: 169–181

Xie D, Zhang J, Zhu X, Pan Y, Liu H, Yuan Z, Yun Y (2016). An improved STARFM with help of an unmixing-based method to generate high spatial and temporal resolution remote sensing data in complex heterogeneous regions. Sensors (Basel), 16(2): 207

Xu H, Shi T, Wang M, Lin Z (2017). Land cover changes in the Xiong' an New Area and a prediction of ecological response to forthcoming regional planning. Acta Ecol Sin, 37(19): 6289–6301

Xue J, Leung Y, Fung T (2017). A bayesian data fusion approach to spatio-temporal fusion of remotely sensed images. Remote Sens, 9 (12): 1310

Xun L, Deng C, Wang S, Huang G B, Zhao B, Lauren P (2017). Fast and accurate spatiotemporal fusion based upon extreme learning machine. IEEE Geosci Remote S, 13(12): 2039–2043

Zhang H, Chen J M, Huang B, Song H, Li Y (2014). Reconstructing seasonal variation of landsat vegetation index related to leaf area index by fusing with MODIS data. IEEE J Stars, 7(3): 950–960

Zhang W, Li A, Jin H, Bian J, Zhang Z, Lei G, Qin Z, Huang C (2013). An enhanced spatial and temporal data fusion model for fusing landsat and MODIS surface reflectance to generate high temporal landsat-like data. Remote Sens, 5(10): 5346–5368

Zhang X Y (2015). Reconstruction of a complete global time series of daily vegetation index trajectory from long-term AVHRR data. Remote Sens Environ, 156: 457–472

Zhang X Y, Friedl M A, Schaaf C B, Strahler A H, Hodges J C F, Gao F, Reed B C, Huete A (2003). Monitoring vegetation phenology using MODIS. Remote Sens Environ, 84(3): 471–475

Zhao Y, Huang B, Song H (2018). A robust adaptive spatial and temporal image fusion model for complex land surface changes. Remote Sens Environ, 208: 42–62

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(11): 2610–2623

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