

Article

An Effective High Spatiotemporal Resolution NDVI Fusion Model Based on Histogram Clustering

Xuegang Xing^{1,2}, Changzhen Yan^{1,3,*}, Yanyan Jia⁴, Haowei Jia^{1,2}, Junfeng Lu¹ and Guangjie Luo⁵

¹ Key Laboratory of Desert and Desertification, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China; xingxuegang@lzb.ac.cn (X.X.); haoweijia@lzb.ac.cn (H.J.); lujunfeng@nieer.ac.cn (J.L.)

² University of Chinese Academy of Sciences, Beijing 100049, China

³ National Earth System Science Data Center, Beijing 100020, China

⁴ Department of Architectural Engineering, Binzhou University, Binzhou 256603, China; jiayanyan@bzu.edu.cn

⁵ Guizhou Provincial Key Laboratory of Geographic State Monitoring of Watershed, Guizhou Education University, Guiyang 550018, China; luoguangjie@gznc.edu.cn

* Correspondence: yancz@lzb.ac.cn

Received: 24 September 2020; Accepted: 15 November 2020; Published: 17 November 2020



Abstract: The normalized difference vegetation index (NDVI) is a powerful tool for understanding past vegetation, monitoring the current state, and predicting its future. Due to technological and budget limitations, the existing global NDVI time-series data cannot simultaneously meet the needs of high spatial and temporal resolution. This study proposes a high spatiotemporal resolution NDVI fusion model based on histogram clustering (NDVI_FMHC), which uses a new spatiotemporal fusion framework to predict phenological and shape changes. Meanwhile, this model also uses four strategies to reduce error, including the construction of an overdetermined linear mixed model, multiscale prediction, residual distribution, and Gaussian filtering. Five groups of real MODIS_NDVI and Landsat_NDVI datasets were used to verify the predictive performance of the NDVI_FMHC. The results indicate that NDVI_FMHC has higher accuracy and robustness in forest areas ($r = 0.9488$ and $ADD = 0.0229$) and cultivated land areas ($r = 0.9493$ and $ADD = 0.0605$), while the prediction effect is relatively weak in areas subject to shape changes, such as flooded areas ($r = 0.8450$ and $ADD = 0.0968$), urban areas ($r = 0.8855$ and $ADD = 0.0756$), and fire areas ($r = 0.8417$ and $ADD = 0.0749$). Compared with ESTARFM, NDVI_LMGM, and FSDAF, NDVI_FMHC has the highest prediction accuracy, the best spatial detail retention, and the strongest ability to capture shape changes. Therefore, the NDVI_FMHC can obtain NDVI time-series data with high spatiotemporal resolution, which can be used to realize long-term land surface dynamic process research in a complex environment.

Keywords: normalized difference vegetation index (NDVI); spatiotemporal fusion framework; histogram clustering; MODIS; Landsat

1. Introduction

The normalized difference vegetation index (NDVI) is one of the commonly used indicators to detect and indicate the status and dynamics of vegetation cover. NDVI time-series data are derived from a wide range of sources, such as MODIS, AVHRR, SeaWiFS, ASTER, Landsat (TM (the Thematic Mapper), ETM+ (the Enhanced Thematic Mapper Plus), OLI (the Operational Land Imager)), and Sentinel-2 MSI, which have been widely investigated and applied from regional to global scale [1–9]. Due to technological and budget limitations, the current NDVI datasets cannot

simultaneously meet the needs of high spatial and temporal resolution [10–12]. In recent years, although the widespread application of UAV (Unmanned Aerial Vehicle) technology and the successive launch of new satellite systems (for example, Sentinel-2) provide a valuable supplement for traditional satellites, we are still lacking NDVI time-series data with high spatiotemporal resolution [12–14]. Therefore, spatiotemporal fusion is of great significance for the study of long-term land surface dynamic processes in a complex environment.

When using spatiotemporal fusion technology to produce NDVI data with high spatiotemporal resolution, there are two blending strategies: blend-then-index (BI) and index-then blend (IB). Research has shown that the IB strategy will become the main strategy for producing NDVI data with high spatial resolution. The IB strategy has three advantages [15–17]: (1) The IB strategy only fuses one band, and the calculation cost is low. (2) NDVI data can eliminate most of the effects related to instrument calibration, solar angle, terrain, cloud shadow, and atmospheric conditions, and enhance the response to vegetation so that the IB strategy can better reduce noise. (3) Because the BI strategy needs to fuse the two bands required for NDVI calculation, there is more error transmission. In addition, the short-term change of NDVI data can be considered linear, and it is reasonable and feasible for the IB strategy to use NDVI data instead of reflectance [18].

At present, there are many spatiotemporal fusion models that can obtain NDVI time-series data with high spatiotemporal resolution [12,19]. Where the land cover remains unchanged and there is less landscape disturbance (hereafter called “phenological change”), several models can accurately and effectively predict phenological changes. For example, the spatial and temporal adaptive reflectance fusion model (STARFM) [10], spatiotemporal adaptive algorithm for mapping reflectance change (STAARCH) [20], enhanced STARFM (ESTARFM) [11], spatiotemporal integrated temperature fusion model (STITFM) [21], and spatial and temporal reflectance unmixing model (STRUM) [22]. However, for land cover change with shape changes (hereafter called “shape change”), such as during urbanization, deforestation, reforestation, wildfires, and floods, the above models struggle to capture and predict these changes. To overcome these difficulties, researchers have successively proposed a series of models, such as the unmixing-based spatial-temporal reflectance fusion model (U-STFM) [23], flexible spatiotemporal data fusion (FSDAF) model [24], sparse-representation-based spatiotemporal reflectance fusion model (SPSTFM) [25], hierarchical spatiotemporal adaptive fusion model (HSTAFM) [26], spatiotemporal fusion network (StfNet) [17], prediction smooth reflectance fusion model (PSRFM) [27], and robust adaptive spatial and temporal fusion model (RASTFM) [28]. Although these BI models produce competitive results, the model based on IB is more suitable for NDVI fusion [18].

Currently, some spatiotemporal fusion models have been designed to produce NDVI data with high spatiotemporal resolution based on the IB strategy, such as the LAC-GAC NDVI integration [29], the NDVI linear mixing growth model (NDVI-LMGM) [18], the NDVI Bayesian spatiotemporal fusion model (NDVI-BSFM) [30], the improved flexible spatiotemporal data fusion (IFSDAF) [31], and spatial-temporal fraction map fusion model (STFMF) [32]. Although the above models can capture the temporal changes and maintain the spatial details of NDVI data, they still have some limitations: (1) The land cover classification maps (hereafter called “classification map”) are the auxiliary data or prior knowledge needed by most spatiotemporal fusion models. At present, there are two ways to input the classification map, one is to input the classification map directly [31], and the other is to generate the classification map automatically according to the unsupervised classification method [24,33]. Because the first method will increase workload and be affected by human factors, it is an efficient method to automatically generate a classification map using unsupervised classification method. Meanwhile, the classification map input of most spatiotemporal fusion models is obtained from the high-resolution NDVI at base date, such as NDVI-BSFM and IFSDAF, and the low-resolution NDVI at the prediction date are not involved. However, the input data for the classification should include the coarse-resolution observation on the prediction date, because it is the only observation that contains information of the surface changes on the prediction date [28]. (2) The spatiotemporal fusion

methods all adopt the window strategy to realize the prediction. Due to the grid effect of the window and the residual distribution error, the final prediction result will contain block effects, such as the ESTARFM [11]. (3) Two or more pairs of cloudless basic NDVI data are needed, such as NDVI-LMGM and STFMF. Although NDVI-LMGM can complete NDVI data fusion by reusing one pair of images, the prediction accuracy is low [31].

The majority of remote sensing data pixels are composed of several ground objects, and the spectral characteristics of most pixels are also composed of spectral values of multiple features. By consequence, the reflectance of pixels can be expressed as a function of the spectral characteristics of the endmembers and their area percentage (abundance). Assuming that the same ground object has the same spectral characteristics and linear additivity, the pixel between low-resolution and high-resolution data can be linked together with linear mixing theory [10,11]. In recent years, the vegetation index has also been used to predict the time change by linear mixing theory [18,34]. This study proposes a high spatiotemporal resolution NDVI fusion model based on histogram clustering (NDVI_FMHC) to generate high spatiotemporal resolution NDVI by using a linear mixing theory. In summary, the main contributions of this study are the following:

- (1) The NDVI_FMHC constructs an effective fusion framework using two kinds of classification maps which are automatically generated by histogram clustering method. Due to the classification map containing coarse-resolution information of the surface changes on the prediction date, the NDVI_FMHC forms an important contribution to improve the prediction accuracy of phenological and shape changes.
- (2) To reduce error, the NDVI_FMHC approach uses four strategies, namely, the construction of an overdetermined linear mixed model, multiscale prediction, residual distribution, and Gaussian filtering.
- (3) We have designed a friendly and concise software interface for the NDVI_FMHC, which will be shared with interested users. The NDVI_FMHC software can directly generate NDVI time-series data with a high spatial resolution by using one pair of cloudless basic NDVI data but with less computational costs and prior knowledge.

2. Method

Before introducing the method, a brief description of the symbols used in the model is provided to facilitate the reader's understanding (Table 1).

Table 1. Model symbols.

Symbol Name	Symbol Brief Description
t_0	Base date
t_p	Prediction date
HS_{t_0}	Landsat_NDVI data at t_0
LS_{t_0}	MODIS_NDVI data at t_0
LS_{t_p}	MODIS_NDVI data at t_p
HS_{t_0}	Landsat resolution data generated from MODIS_NDVI data sampling at t_0
HS_{t_p}	Landsat resolution data generated from MODIS_NDVI data sampling at t_p
HS_{tp1}	Predicted Landsat_NDVI data based on classification map 1
HS_{tp2}	Predicted Landsat_NDVI data based on classification map 2

2.1. Spatiotemporal Fusion Framework

As the input data of the spatiotemporal fusion model, the quality of the classification map is crucial to the accuracy and robustness of the fusion results. Most models assume that the type of land cover cannot change during spatiotemporal fusion; however, this assumption is unreasonable in some situations, such as floods and fires. This study uses hierarchical clustering to provide land cover for NDVI data fusion through local histogram features and generates two kinds of classification

maps, in which classification map 1 is used for predicting HS_{tp1} , which is directly generated from HS_{t0} , and classification map 2 is used for predicting HS_{tp2} , which is generated from HS_{tp1} and HS_{t0} . An effective high spatiotemporal NDVI data fusion framework based on histogram feature clustering is constructed using the above two classification maps.

Based on the HS_{t0} , LS_{t0} and LS_{tp} , the NDVI_FMHC approach constructs an effective fusion framework using two kinds of classification maps to predicting high-resolution NDVI data (HS_{tp2}). HS_{t0} , LS_{t0} , and LS_{tp} all need to be preprocessed with the same geometric correction and atmospheric correction. The spatiotemporal fusion framework consists of five steps: ① Classification map 1 is generated by hierarchical clustering according to the local histogram features based on HS_{t0} ; ② Combined with classification map 1, LS_{t0} and LS_{tp} are up-sampled using a linear mixing model to obtain the HS_{u0} and HS_{up} with the Landsat resolution as HS_{t0} ; ③ Based on HS_{t0} , HS_{u0} , and HS_{up} , using classification map 1, the first high spatial resolution NDVI data at tp (HS_{tp1}) are predicted by the linear mixing model; ④ Combined with HS_{t0} and HS_{tp1} , classification map 2 is generated by hierarchical clustering according to the local histogram features; ⑤ Based on HS_{t0} , HS_{u0} , and HS_{up} , using classification map 2, the second high spatial resolution NDVI data at tp (HS_{tp2}) are predicted through the linear mixing model. In the above five steps, the HS_{t0} , HS_{u0} , and HS_{up} , HS_{tp1} , and HS_{tp2} are all processed by Gaussian filtering.

2.2. Feature Description of Land Cover Classification

The NDVI_FMHC automatically generated a classification map by the hierarchical clustering method, in which a local histogram is used as the feature to classify pixels. In the first three steps of the NDVI_FMHC, due to the lack of high spatial resolution information at tp , a 3×3 moving window based on only HS_{t0} is used to build classification map 1 produced by a 16-dimensional local histogram feature. After obtaining HS_{tp1} , classification map 2 is generated based on HS_{t0} and HS_{tp1} , in which each moving window at the corresponding position generates a 16-dimensional local histogram, and then expands to a 32-dimensional local histogram. Therefore, classification map 2 contains both phenological information without shape change and land cover change information with the shape change at $t0$ and tp , which helps improve the accuracy of the fusion result.

2.3. Estimation of NDVI Change with a Linear Mixture Model

Although the variation in NDVI over time is complex, it satisfies the linear change hypothesis in the short term [18]. The NDVI increment of the coarse pixel (x_1, y_1) is defined as

$$\Delta\text{NDVI}_C(x_1, y_1) = \text{NDVI}_{C,tp}(x_1, y_1) - \text{NDVI}_{C,t0}(x_1, y_1) \quad (1)$$

According to linear mixing theory, the NDVI of the coarse pixel is a linear superposition of the fine pixel, from which the relationship between the short-term change of the coarse pixel and the fine pixel NDVI can be obtained, as follows:

$$\begin{bmatrix} \Delta\text{NDVI}_C(x_1, y_1) \\ \vdots \\ \Delta\text{NDVI}_C(x_i, y_i) \\ \vdots \\ \Delta\text{NDVI}_C(x_n, y_n) \end{bmatrix} = \begin{bmatrix} f_1(x_1, y_1) & f_2(x_1, y_1) & \cdots & f_l(x_1, y_1) \\ \vdots & \vdots & & \vdots \\ f_1(x_i, y_i) & f_2(x_i, y_i) & \cdots & f_l(x_i, y_i) \\ \vdots & \vdots & & \vdots \\ f_2(x_n, y_n) & f_1(x_n, y_n) & \cdots & f_l(x_n, y_n) \end{bmatrix} \begin{bmatrix} \Delta\text{NDVI}_{F,1} \\ \vdots \\ \Delta\text{NDVI}_{F,C} \\ \vdots \\ \Delta\text{NDVI}_{F,l} \end{bmatrix} \quad (2)$$

In Equation (2), l is the count of classes and $\Delta\text{NDVI}_C(x_i, y_i)$ is the increment of the coarse pixel NDVI from $t0$ to tp . It should be noted that the coarse pixel here is not the original low-resolution data pixel, but a new coarse pixel composed of a certain number of fine pixels based on HS_{u0} and HS_{up} (detailed in Section 2.5). $f_l(x_i, y_i)$ is the abundance of the l category in the coarse pixels (x_i, y_i) . When obtaining HS_{u0} , HS_{up} , and HS_{tp1} , the abundance is calculated from classification map 1,

and when predicting HS_{tp2} , the abundance is calculated from classification map 2. $\Delta NDVI_{F,C}$ is the increment of the fine pixel NDVI of category 1 from $t0$ to tp . The least-squares method is used to solve Equation (2) to obtain the optimal solution of $\Delta NDVI_{F,C}$.

2.4. Residuals and Its Allocation Formed by Predicting Time Changes

Previous studies [24] have shown that if there is no land cover change from $t0$ to tp , the value of NDVI fine pixels can be accurately estimated at tp , while the prediction is less accurate where land cover change has occurred and large within-class variability exists. In addition, it is difficult to construct a positive definite linear mixing model due to the non-equilibrium of ground feature distribution. To eliminate the influence of multiple solutions under positive definite equations, a large window is usually used to construct the equations of the overdetermined linear mixing model. After the optimal solution is obtained by the least-squares method, although the residual error is very small, it is still not zero. The residual estimation method of the NDVI_FMHC for the prediction and true value of fine pixels NDVI is consistent with that of FSDAF [24], and the residual distribution method is consistent with that of IFSDAF [31,32].

2.5. Gaussian Filtering

First, the remote sensing data inevitably contains noise due to the influence of sensor material properties and the working environment. Second, while solving the linear mixed model using the least squares method, additional noise is generated. Therefore, The NDVI_FMHC uses Gaussian filtering to remove the noise of HS_{t0} , HS_{u0} , HS_{up} , HS_{tp1} , and HS_{tp2} . The formula is as follows:

$$G(x_i, y_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{NDVI(x_i, y_i)^2}{2\sigma^2}} \quad (3)$$

Gaussian filtering is mainly affected by two parameters when removing noise, which are Gaussian kernel size and standard deviation (σ). Gaussian kernel size is a 3×3 fine pixel window, and σ is determined through multiple experiments. In addition, $G(x_i, y_i)$ is the NDVI(x_i, y_i) value of the fine pixels at (x_i, y_i) after Gaussian filtering.

2.6. Block Effect Elimination and Final Prediction

The spatiotemporal fusion methods all adopt the window strategy to realize the prediction. Due to the grid effect of the window and the residual distribution error, the final prediction result will contain block effects, similar to the ESTARFM algorithm [11]. To eliminate the block effect, the NDVI_FMHC uses multiple different numbers of fine pixels to form new coarse pixels based on HS_{u0} and HS_{up} , and uses the arithmetic average of several prediction results as the final prediction result (hereafter called “multiscale prediction”). Specifically, the NDVI_FMHC introduces a new parameter (coarse pixel width) to determine the size of a new coarse pixel. If the input value of the coarse pixel width is k ($k \geq 3$), the size of the new coarse pixel has three values, and the resolutions are $k \times 30$ m, $(k + 2) \times 30$ m, and $(k + 4) \times 30$ m (fine pixel size is 30 m). Values above the three coarse pixel width are used to make three predictions respectively, obtaining three parallel HS_{tp2} , which can be marked as HS_{tp2}^1 , HS_{tp2}^2 , and HS_{tp2}^3 . For instance, when the value of k is 4 and the fine pixel size is 30 m, HS_{tp2}^1 is predicted using the new coarse pixel of 120×120 m according to the framework of the NDVI_FMHC (Figure 1), then HS_{tp2}^2 and HS_{tp2}^3 are predicted using the new coarse pixel of 180×180 m and 240×240 m, respectively. The final prediction result of Landsat_NDVI is as follows:

$$\text{Landsat_NDVI}^P = \frac{\sum_1^3 HS_{tp2}^i}{3} \quad (4)$$

Landsat_NDVI^P is the final prediction result.

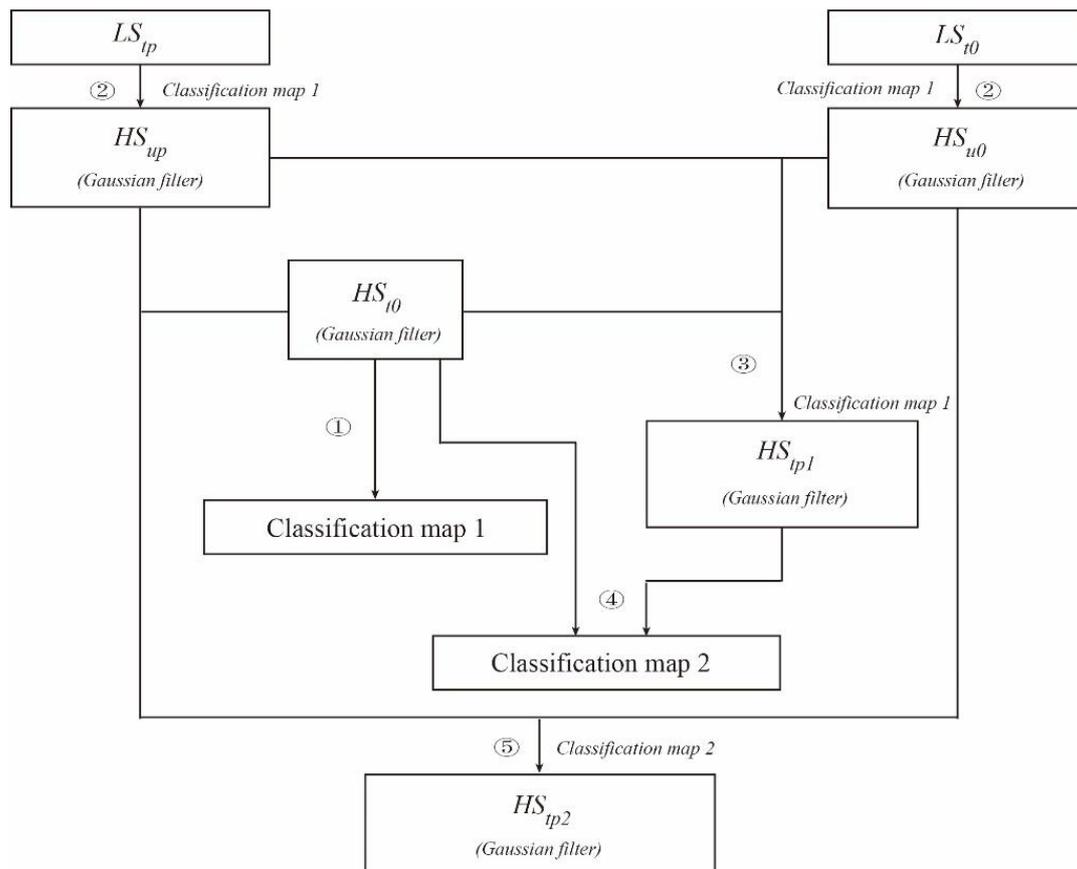


Figure 1. Framework of the NDVI_FMHC. ①–⑤ represent the five steps of the NDVI_FMHC.

3. Data

3.1. Test Dataset Preprocessing

To evaluate the performance of the NDVI_FMHC, five group datasets of which the high-resolution (Landsat TM\ETM+\OLI) and low-resolution (MOD09GQ) datasets are used to test the prediction ability with different land cover characteristics. All the data were downloaded from the USGS Earth Explorer website (<https://earthexplorer.usgs.gov/>). After FLAASH atmospheric correction for Landsat reflectance data, the Landsat_NDVI is calculated based on the near-infrared and red bands with 30 m spatial resolution. After pre-processing by the MODIS Reprojection Tool Swath (MRT Swath) software for MODIS reflectance data, MODIS_NDVI is calculated based on the near-infrared band and red band with 250 m spatial resolution and one-day temporal resolution and resampled to 30 m using bilinear interpolation. Both testing dates adopted the Universal Transverse Mercator (UTM) coordinate system under the WGS-84 datum, in which the coverage area is 576 km² from the first to fourth group dataset, while the fifth dataset is 4444.64 km².

3.2. Study Area and Data

The first site is located in Saskatoon, Canada (54° N, 106° W), which has been used to test many spatiotemporal fusion methods [10,11,25,35,36]. The type of land cover in this area is mainly forest, such as spruce, pine, and aspen, supplemented by swamps and sparse vegetation patches. The test datasets were obtained on 11 July (tp) and 12 August 2001 ($t0$), using MODIS_NDVI on 11 July 2001 for prediction (Figure 2). During the prediction period, land cover change is very small, but the growing season is short, and the phenology changes significantly [10]. Therefore, this site can be considered a forest area with phenological changes.

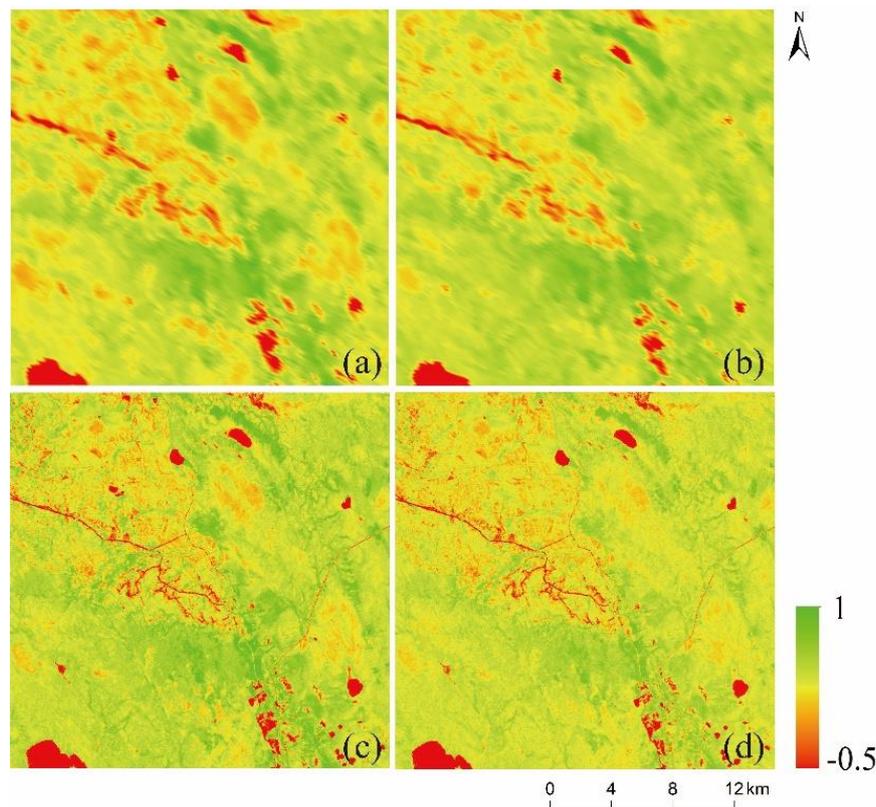


Figure 2. Testing datasets of forest areas. (a,b). MODIS_NDVI on 11 July and 12 August 2001; (c,d). Landsat_NDVI on 11 July and 12 August 2001.

The second site is located in southern New South Wales, Australia (29°S, 150°E), and the main type of land cover is irrigation fields, and a small number of drylands and forests are present [37]. The test datasets were obtained on 5 July (t_0) and 22 August 2004 (t_p), using MODIS_NDVI on 22 August 2004 for prediction (Figure 3). In this area, crop planting or harvesting is the main activity, leading to drastic short-term changes in non-shape land cover. Owing to the crop type differences in each patch, this site has the characteristics of high spatial heterogeneity. Therefore, this site is an example of a cultivated land area with high spatial heterogeneity and mainly occurs as phenological change and land cover change without shape changes (hereafter called “non-shape change”).

The third site is located in northern New South Wales, Australia (29° S, 149° E), which is used in many spatiotemporal fusion methods [12,27], and the main types of land covers are cultivated land, water, and forest. The test datasets were obtained on 26 November (t_0) and 12 December 2004 (t_p), using MODIS_NDVI on 12 December 2004 for prediction (Figure 4). A large flood event occurred in December 2004, causing a large amount of land to be inundated, and the type of land cover changed drastically. Therefore, this site is an example of a flood area with shape changes caused by floods.

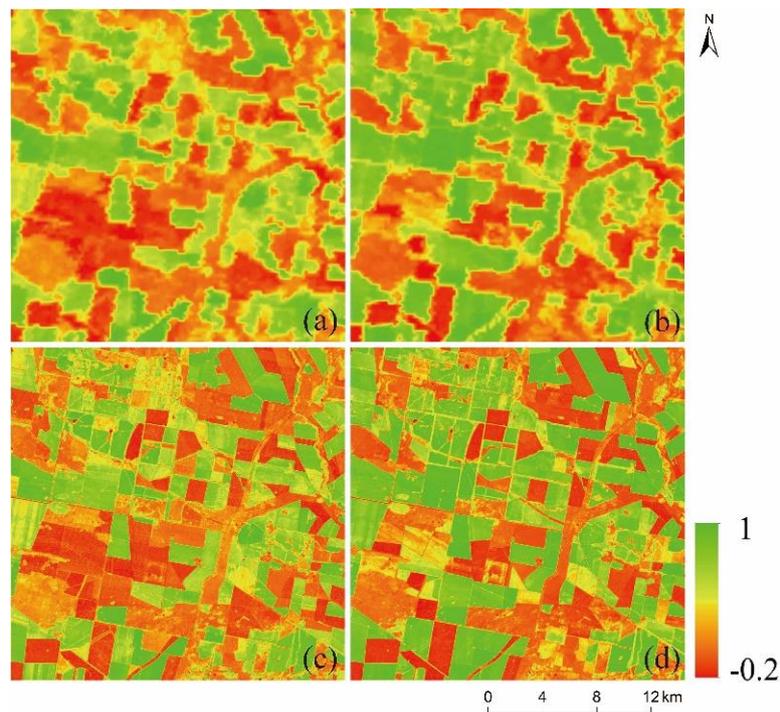


Figure 3. Testing datasets of cultivated land areas. (a,b). MODIS_NDVI on 5 July and 22 August 2004; (c,d). Landsat_NDVI on 5 July and 22 August 2004.

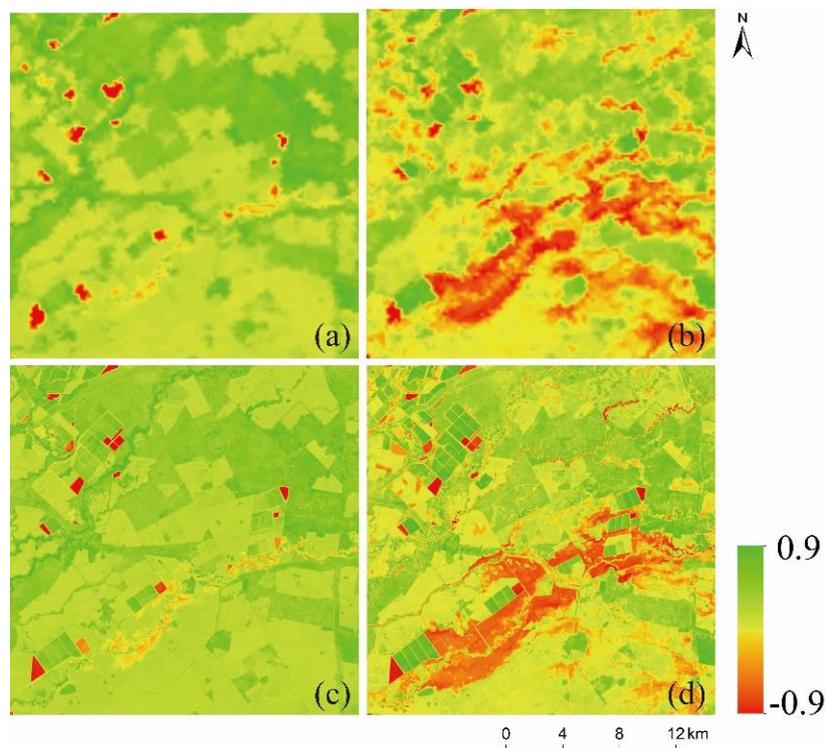


Figure 4. Testing datasets of flood areas. (a,b). MODIS_NDVI on 26 November and 12 December 2004; (c,d). Landsat_NDVI on 26 November and 12 December 2004.

The fourth site is located in Shenzhen, China (22° N, 114° E), and has been used to test many fusion methods [28,38], and the types of land covers are mainly urban land, water, and forest. The test datasets were obtained on 1 November 2000 (t_0) and 7 November 2002 (t_p), using MODIS_NDVI on

7 November 2002 for prediction (Figure 5). Because Shenzhen developed rapidly, the city has expanded vigorously from 2000 to 2002, causing a large area of vegetation to be transformed into urban areas. Therefore, this site is an example of an urban area with shape changes caused by urban expansion.

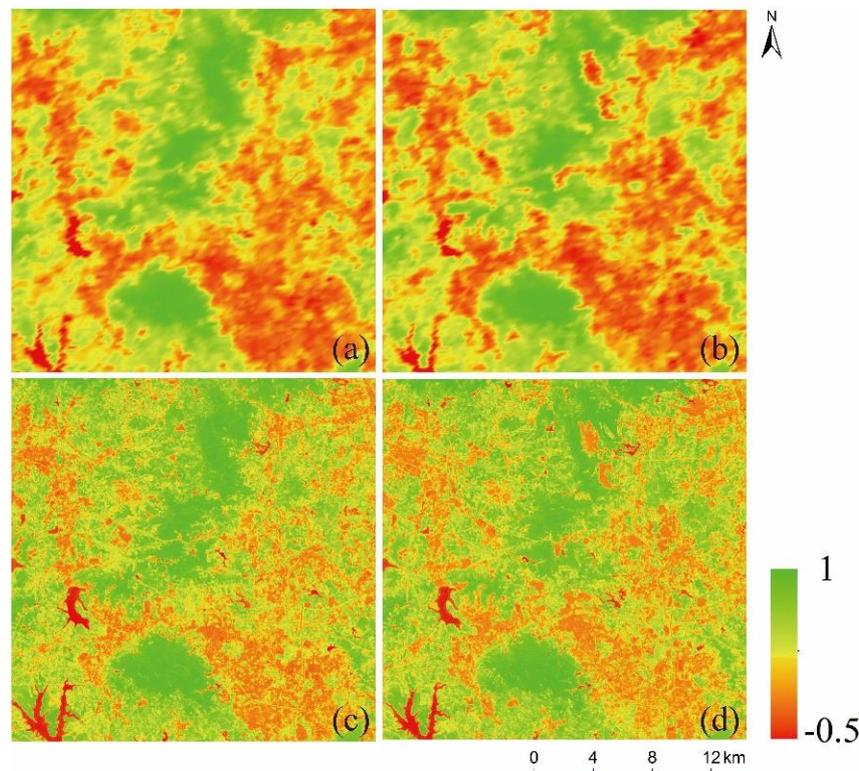


Figure 5. Testing datasets of urban areas. (a,b). MODIS_NDVI on 1 November 2000 and 7 November 2002; (c,d). Landsat_NDVI on 1 November 2000 and 7 November 2002.

The fifth site is located in Kangaroo Island, Australia (36° S, 137° E), in which the main types of land cover are cultivated land, forest, grasslands, and water. The test datasets were acquired on 22 January 2019 (t_0) and 10 February 2020 (t_p), using MODIS_NDVI on 10 February 2020 for prediction (Figure 6). A major forest fire broke out in Kangaroo Island during the New Year of 2020, causing a large area of forest, grassland, and cultivated land to become bare soil. Therefore, this site is an example of a fire area with land cover from multi- to single type.

3.3. Quality Evaluation Index

For these five sites, ESTARFM [11], NDVI-LMGM [18], and FSDAF [24] are also applied to the same datasets for comparison, to verify that the NDVI_FMHC has a good ability to predict non-shape and shape changes. To ensure fairness, these algorithms all use the default parameters given by the authors. In this study, the Pearson correlation coefficient (r), root mean square error (RMSE), average absolute difference (AAD), average error (AE), and structure similarity (SSIM) [39] are used as quantitative measures to evaluate the spectral and structural similarity between the predicted and original NDVI.

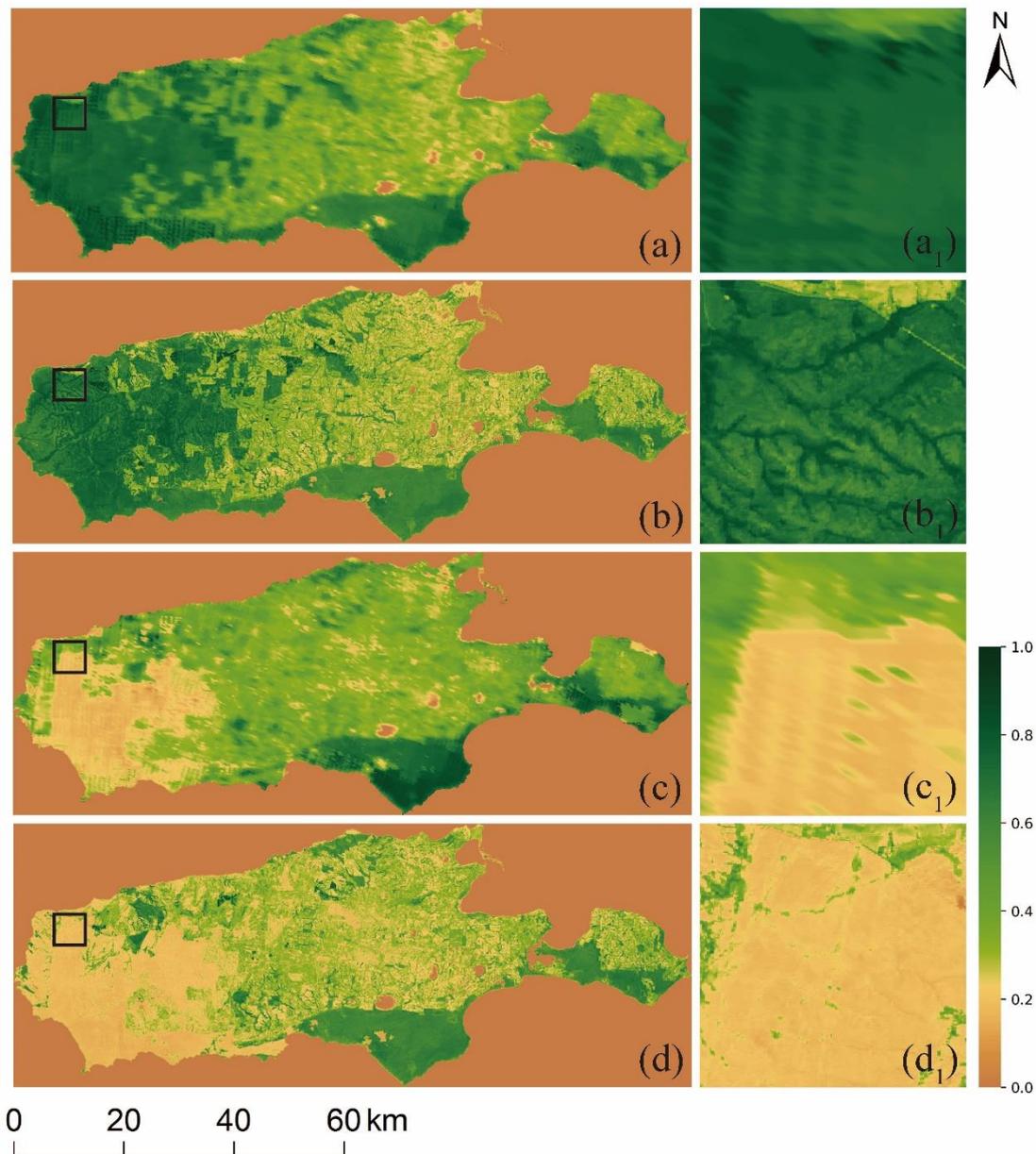


Figure 6. Testing datasets of fire areas. **(a,a₁)**. MODIS_NDVI and inset map on 19 January 2019; **(b,b₁)**. Landsat_NDVI and inset map on 19 January 2019; **(c,c₁)**. MODIS_NDVI and inset map on 10 February 2020; **(d,d₁)**. Landsat_NDVI and inset map on 10 February 2020.

4. Analysis Results

4.1. Phenological Changes in Forest Area

The original and predicted Landsat_NDVI of the forest study area on 11 July 2001 by four spatiotemporal fusion models are very similar, indicating that all four models can capture the phenology change and success prediction of Landsat_NDVI (Figures 7 and 8). The main reason is that the correlation and spatial structure between the above original and Landsat_NDVI on 12 August 2001 are very good with lower prediction difficulty ($r = 0.9377$ and $SSIM = 0.9436$, Table 2). By analyzing the quantitative indicators, we find that r and $SSIM$ of the NDVI_FMHC are higher than those of ESTARFM, NDVI-LMGM, and FSDAF, and AAD (0.0229), AE (0.0000), and RMSE (0.0373) of the NDVI_FMHC are the smallest (Table 2). In addition, the scatter of all four models is close to the 1:1 line; however, the scatterplots of FSDAF and NDVI_FMHC are more concentrated than those of ESTARFM

and NDVI-LMGM (Figure S1). Therefore, in the case of only phenological changes in the forest area, the FSDAF and NDVI_FMHC have an effective prediction, in which the NDVI_FMHC has the least difference and the best correlation between original and predicted Landsat_NDVI in the forest area.

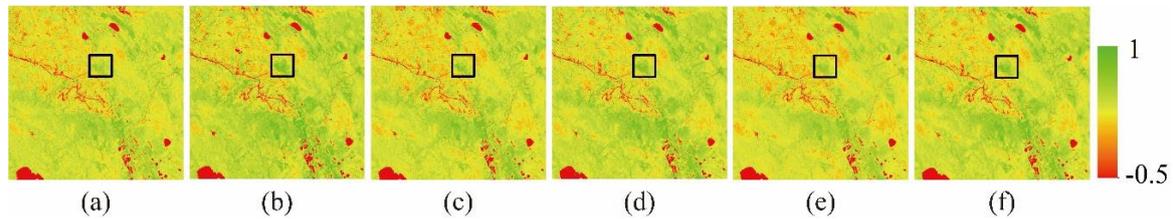


Figure 7. Visual comparison of the forest area. (a,b). Original Landsat_NDVI on 12 August and 11 July 2001; (c–f). Predicted Landsat_NDVI by ESTARFM, NDVI-LMGM, FSDAF and NDVI_FMHC.

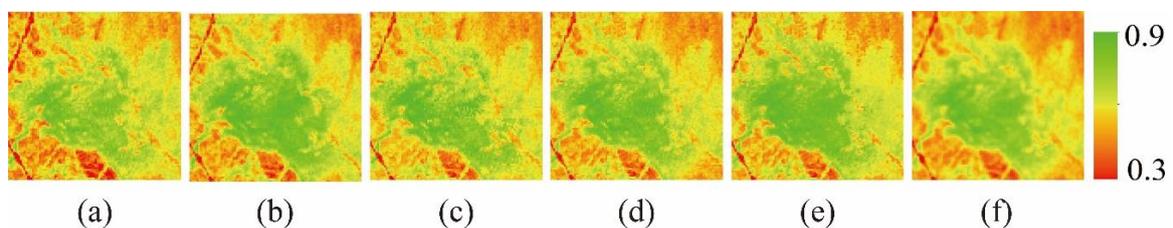


Figure 8. Visual comparison of inset map in the forest area. (a,b). Original Landsat_NDVI inset map on 12 August and 11 July 2001; (c–f). Predicted Landsat_NDVI inset map by ESTARFM, NDVI-LMGM, FSDAF and NDVI_FMHC.

Table 2. Accuracy comparison of four fusion methods in forest area.

Name\Indicator	Landsat 12 August 2001	ESTARFM	NDVI-LMGM	FSDAF	NDVI_FMHC
r	0.9377	0.9363	0.8996	0.9405	0.9488
AAD	0.0346	0.0258	0.0321	0.0247	0.0229
AE	0.0263	0.0006	0.0002	−0.0002	0.0000
RMSE	0.0496	0.0417	0.0536	0.0408	0.0373
SSIM	0.9436	0.9436	0.9100	0.9470	0.9548

4.2. Non-Shape Change in Cultivated Land Area

For the non-shape change, the prediction effect of the NDVI_FMHC is better than that of the other three models (Figure 9). Four models can capture and successfully predict the change in land cover over a large area (elliptical region), however, the prediction effect is poor in small areas (rectangular region), especially ESTARFM (Figure 10). Due to the low correlation and spatial structural similarity between original Landsat_NDVI on 5 July 2004 and Landsat_NDVI on 22 August 2004 ($r = 0.8240$ and $SSIM = 0.8123$, Table 3), it is difficult to predict the change in cultivated land areas, especially for the small area. Compared with the NDVI_FMHC, the values of AAD, AE, and RMSE of ESTARFM, NDVI-LMGM, and FSDAF are higher 3.14% to 11.61%, 4.41% to 22.18%, and 4.26% to 17.91%, while r and SSIM are lower, 0.42% to 2.08% and 0.36% to 2.05%, respectively (Table 3). In addition, all scatter of FSDAF and NDVI_FMHC is closer to the 1:1 line than those of ESTARFM and NDVI-LMGM, indicating that the predicted values of FSDAF and NDVI_FMHC are closer to the actual value (Figure S2). Thus, compared with the ESTARFM, NDVI-LMGM, and FSDAF, the NDVI_FMHC has the best spatial correlation and structural similarity with the original Landsat_NDVI and the smallest prediction error in cultivated land area.

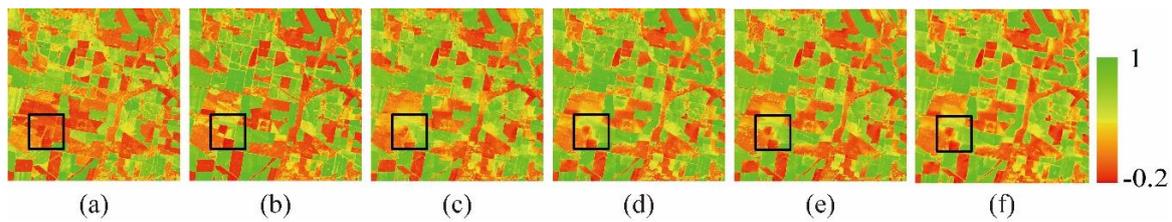


Figure 9. Visual comparison of cultivated land area. (a,b). Original Landsat_NDVI on 5 July and 22 August 2004; (c–f). Predicted Landsat_NDVI by ESTARFM, NDVI-LMGM, FSDAF and NDVI_FMHC.

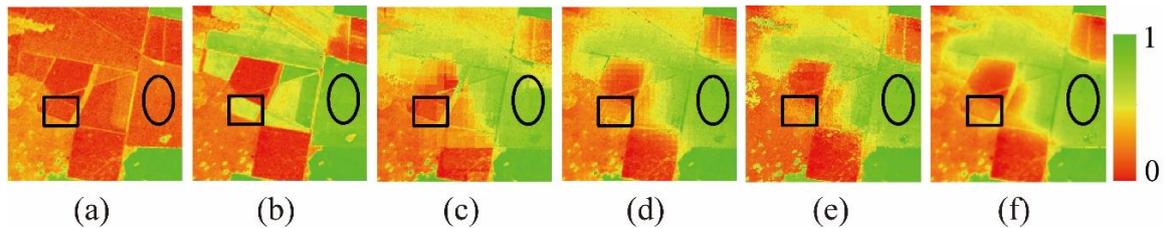


Figure 10. Visual comparison of inset map in cultivated land area. (a,b). Original Landsat_NDVI inset map on 5 July and 22 August 2004; (c–f). Predicted Landsat_NDVI inset map by ESTARFM, NDVI-LMGM, FSDAF, and NDVI_FMHC.

Table 3. Accuracy comparison of four fusion methods in cultivated land area.

Name\Indicator	Landsat 22 August 2004	ESTARFM	NDVI-LMGM	FSDAF	NDVI_FMHC
r	0.8240	0.9295	0.9388	0.9453	0.9493
AAD	0.1129	0.0676	0.0639	0.0624	0.0605
AE	−0.0653	0.0250	0.0215	0.0213	0.0204
RMSE	0.1657	0.1025	0.0954	0.0906	0.0869
SSIM	0.8123	0.9299	0.9396	0.9459	0.9494

4.3. Shape Change in the Flood Area

The prediction results of the four models are not ideal, and it is difficult to accurately capture the shape changes of flood action (Figures 11 and 12). The main reasons are as follows: first, the correlation and spatial structural similarity between Landsat_NDVI on 26 November and 12 December 2004 are very low with a large shape change range formed by the flood ($r = 0.6625$ and $SSIM = 0.5643$, Table 4); second, MODIS and Landsat images were collected at an interval of approximately 0.5 h, and the flood occurred very quickly, leading to different inundation ranges between MODIS_NDVI and Landsat_NDVI on 12 December 2004 (Figure 4b,d); third, NDVI is very sensitive to surface vegetation coverage and less sensitive to the depth of water and turbidity. Overall, the NDVI_FMHC is superior to the other three models because ESTARFM and NDVI-LMGM have obvious block effects, FSDAF has poor continuity, and ESTARFM has the largest difference with the actual flood range (Figure 12). Compared with NDVI_FMHC, AAD, and RMSE of NDVI_FMHC are lower than those for ESTARFM, NDVI-LMGM, and FSDAF, while r and SSIM are higher (Table 4). In addition, all scatter of the NDVI_FMHC is closest to the 1:1 line, and ESTARFM has the largest dispersion and the lowest prediction accuracy (Figure S3). Thus, although the four models are not ideal in the flood area, the Landsat_NDVI predicted by the NDVI_FMHC is still better than that by ESTARFM and FSDAF, while better than NDVI_LMGM by some but not all measures.

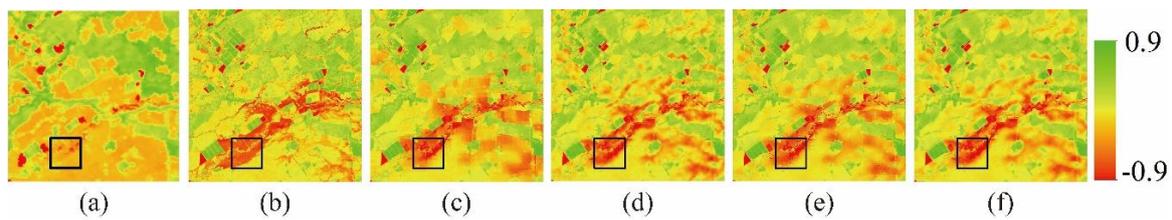


Figure 11. Visual comparison of the flood area. (a,b). The original Landsat_NDVI on 26 November and 12 December 2004; (c–f). Predicted Landsat_NDVI by ESTARFM, NDVI-LMGM, FSDAF and NDVI_FMHC.

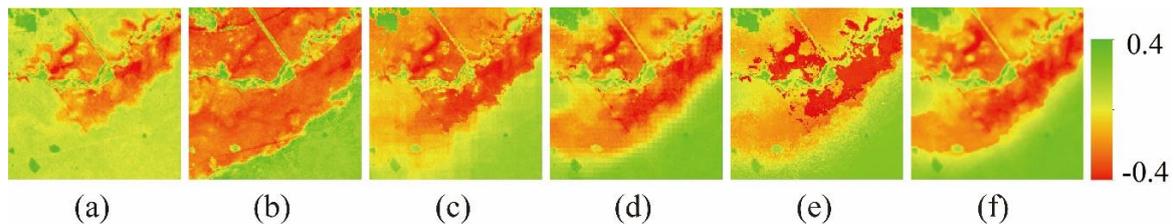


Figure 12. Visual comparison of inset map in the flood area. (a,b). Original Landsat_NDVI inset map 26 November and 12 December 2004; (c–f). Predicted Landsat_NDVI inset map by ESTARFM, NDVI-LMGM, FSDAF, and NDVI_FMHC.

Table 4. Accuracy comparison of four fusion methods in the flood area.

Name\Indicator	Landsat 12 December 2004	ESTARFM	NDVI-LMGM	FSDAF	NDVI_FMHC
r	0.6625	0.8027	0.8365	0.8311	0.8450
AAD	0.1222	0.1052	0.0960	0.1003	0.0968
AE	−0.0064	−0.0468	−0.0457	−0.0455	−0.0458
RMSE	0.1646	0.1367	0.1265	0.1284	0.1245
SSIM	0.5643	0.7710	0.8080	0.7953	0.8072

4.4. Shape Change in Urban Area

The prediction result of the NDVI_FMHC is the closest to the original Landsat_NDVI, followed by NDVI-LMGM and FSDAF, and ESTARFM is the worst (Figure 13). Meanwhile, the four models can only capture large-scale shape changes (small rectangular region) but have difficulty predicting small-scale shape changes (elliptical region) because the partial area produced by urbanization is too small and not recorded in the predicted MODIS_NDVI (Figure 14). Compared with the NDVI_FMHC, the value of AAD, AE, and RMSE of ESTARFM, NDVI-LMGM, and FSDAF are higher 3.50% to 7.38%, 0.56% to 13.84%, and 5.35% to 10.19%, and r and SSIM are lower 1.53% to 2.99% and 1.37% to 2.89%, respectively (Table 5). In addition, all scatter of the NDVI_FMHC is closest to the 1:1 line, indicating that the predicted value of this model is closest to the actual value (Figure S4).

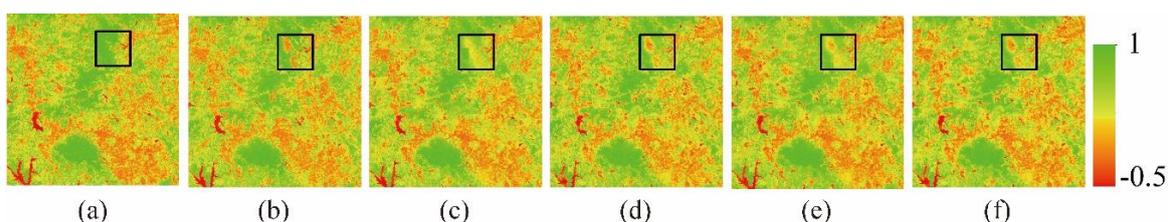


Figure 13. Visual comparison of urban area. (a,b). Original Landsat_NDVI on 1 November 2000 and 7 November 2002; (c–f). Predicted Landsat_NDVI by ESTARFM, NDVI-LMGM, FSDAF, and NDVI_FMHC.

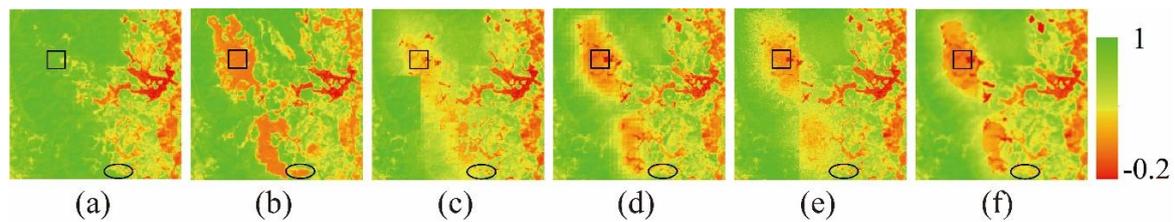


Figure 14. Visual comparison of inset map in urban area. (a,b). Original Landsat_NDVI inset map 1 November 2000 and 7 November 2002; (c–f). Predicted Landsat_NDVI inset map by ESTARFM, NDVI-LMGM, FSDAF, and NDVI_FMHC.

Table 5. Accuracy comparison of four fusion methods in urban area.

Name\Indicator	Landsat 12 August 2001	ESTARFM	NDVI-LMGM	FSDAF	NDVI_FMHC
r	0.8224	0.8590	0.8691	0.8719	0.8855
AAD	0.0868	0.0811	0.0807	0.0782	0.0756
AE	0.0234	0.009	0.0108	0.0108	0.0109
RMSE	0.1338	0.1154	0.1132	0.1103	0.1047
SSIM	0.8271	0.8594	0.8728	0.8711	0.8850

4.5. Shape Change in the Fire Area

Four models can capture shape changes formed by the fire and successfully predict them, while the texture of predicted and original Landsat_NDVI are not consistent from the inset map, and the prediction effect is not good (Figure 15). First, the correlation and spatial structural similarity between original Landsat_NDVI on 19 January 2019 and 10 February 2020 are extremely low ($r = 0.2792$ and $SSIM = 0.2812$), indicating that a large area of forest, grassland, and cultivated land become bare soil due to the fire. Second, the predicted Landsat_NDVI still retains the texture information of basic Landsat_NDVI, leading to poor local effect. The five indicators of ESTARFM, FSDAF, and NDVI_FMHC are basically consistent; however, NDVI-LMGM has lower r (0.8333) and SSIM (0.8274) and larger AAD (0.0780) and RMSE (0.1092), indicating that ESTARFM, FSDAF, and NDVI_FMHC have basically similar effects in predicting shape change of the fire area (Table 6). In addition, the scatter plots of ESTARFM, FSDAF, and NDVI_FMHC are basically consistent, and the prediction results are similar (Supplemental Figure S5). Therefore, for the multiple types of land cover converted into a single type due to the fire, although four models can capture the overall shape change, the local prediction effect is not ideal because the prediction results still retain the texture information of the original Landsat_NDVI.

Table 6. Accuracy comparison of four fusion methods in fire area.

Name\Indicator	Landsat 10 February 2020	ESTARFM	NDVI-LMGM	FSDAF	NDVI_FMHC
r	0.2792	0.8489	0.8333	0.8363	0.8417
AAD	0.1785	0.0738	0.0780	0.0733	0.0749
AE	0.1600	0.0290	0.02259	0.0281	0.0257
RMSE	0.2772	0.1019	0.1092	0.1021	0.1035
SSIM	0.2812	0.8463	0.8274	0.8389	0.8399

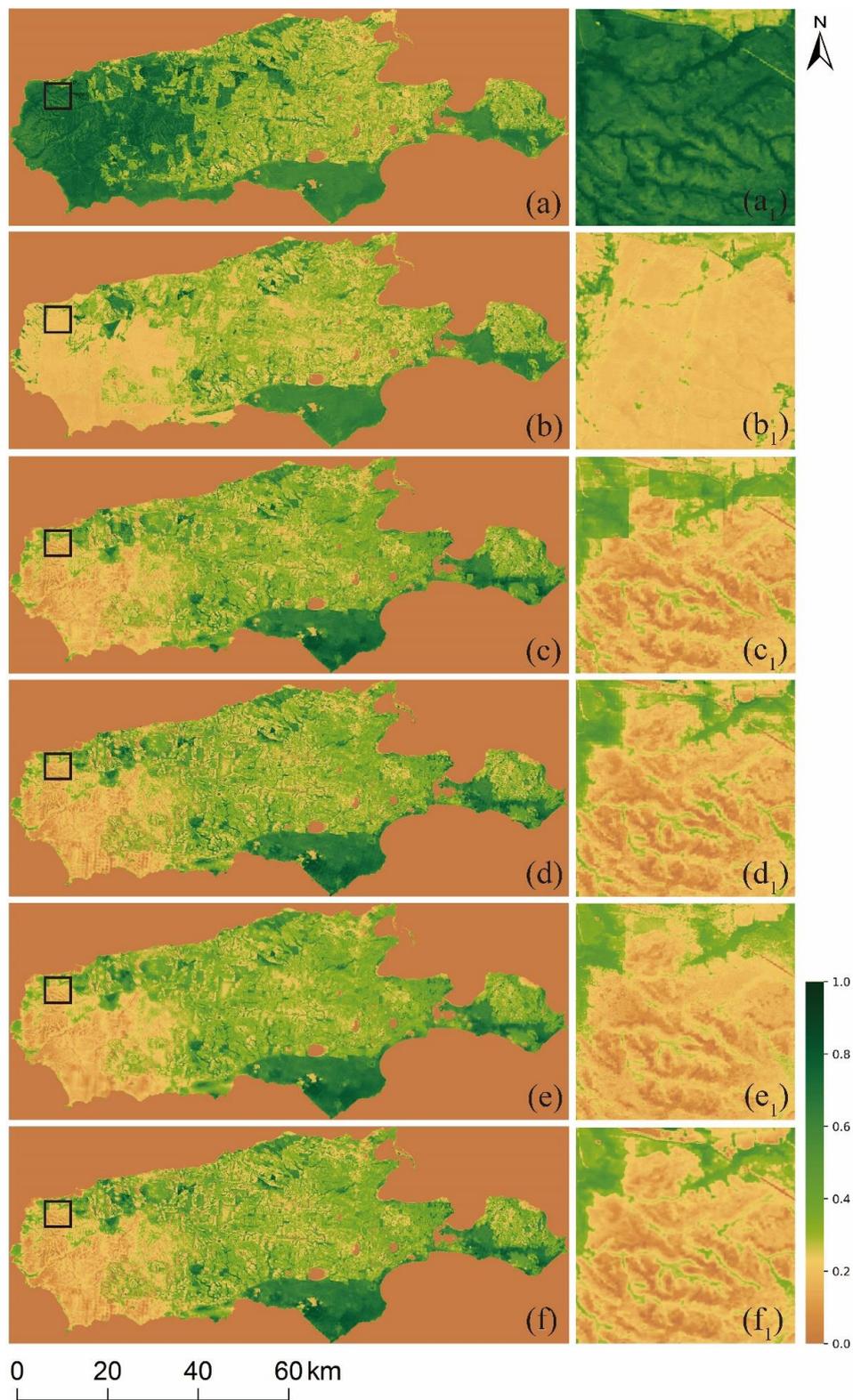


Figure 15. Visual comparison of the fire area. (a,b) and (a₁,b₁). Original Landsat_NDVI and inset map on 22 January 2019 and 10 February 2020; Predicted Landsat_NDVI (c–f) and its inset map (c₁–f₁) by ESTARFM, NDVI-LMGM, FSDAF and NDVI_FMHC.

5. Discussion

The study integrates multi-source remote sensing data (MODIS_NDVI and Landsat_NDVI) to obtain NDVI time-series data with high spatiotemporal resolution and study the long-term dynamic processes of the surface environment. In general, the most common natural changes on the land surface are phenological changes, simultaneously accompanied by dramatic land cover changes. Due to water levels rising or falling, fire hazards, floods, urban expansion, and deforestation, shape changes of land cover also frequently occur. During spatiotemporal fusion, the prediction of phenological changes is relatively easy, while the prediction of these shape changes is extremely difficult. Therefore, it is of great significance to understand the accurate, automatic, and robust prediction of complex shape changes in various landscapes for comprehensive monitoring of the surface dynamics.

5.1. Prediction Advantages of the NDVI_FMHC

The changing pattern of NDVI is determined by biotic factors (e.g., vegetation type) and environmental factors (soil, temperature, and precipitation). However, environmental factors play a more significant role than biotic factors, and in a small area, NDVI change can be assumed to be similar within the same type of land cover. Therefore, NDVI-LMGM assumes that within a short period, NDVI values of adjacent pixels of the same land cover exhibit the same linear changes [18]. If only phenological changes occurred, this assumption would be reasonable, but the prediction results will have a large error when shape changes and existing variations occur within the class. The FSDAF [24] has designed a new weighting function reallocating residuals to areas with shape changes reducing errors. The NDVI_FMHC proposes a new spatiotemporal fusion framework to solve this problem based on classification map 1 and classification map 2, in which classification map 1 is obtained by HS_{t0} and classification map 2 is obtained by combining HS_{t0} and HS_{tp1} . Compared with classification map 1, classification map 2 further refines the land cover types and its changing characteristics, which can distinguish shape changes that did and did not occur during the prediction period. Therefore, the abundance calculated by classification map 2 is more reasonable, and the time variation of the terminal element solved by the linear mixture model has better robustness and higher accuracy.

It is necessary to select a suitable classification method to classify and obtain an accurate classification map, which is the basis of the spatiotemporal fusion model. Currently, some spatiotemporal fusion models select similar pixels based on the threshold of spectral distance to obtain classification maps, such as STARFM and ESTARFM [11,24]. Because the threshold is empirical, it is difficult to accurately capture the boundaries of different land cover types owing to different objects with the same spectrum, especially NDVI. There are also some spatiotemporal fusion models using ISODATA [18,24] or the K-Means method [2] to perform unsupervised classification to obtain classification maps. However, the ISODATA method needs many parameters, and the input value is not easy to determine, and K-Means is easily affected by the initial value. Hierarchical clustering is also widely used in image segmentation and classification owing to ease of use and defining similarity rules [40,41], and the NDVI_FMHC selects this method. There are many clustering methods for image classification, each of which has its own advantages and disadvantages. We have designed two features for applying clustering classification methods, including a 16-dimensional local histogram based on HS_{t0} , and a 32-dimensional local histogram expanded by two 16-dimensional local histograms based on HS_{t0} and HS_{tp1} , which not only contains the information of land cover but also reflects the change at two moments.

Due to the grid effect of the window and error of residual distribution, the regional errors of shape changes and variations within the class are large, leading to the block effect of the prediction image. Therefore, it is very important to reduce the error between the true and predicted value for various spatiotemporal fusion models. The NDVI_FMHC uses four strategies to reduce error. First, the overdetermined equations of the linear mixed model are constructed with a large window and the optimal solution is obtained by the least-squares method which can reduce errors. Second, three times

parallel HS_{tp2} prediction results are obtained to calculate the arithmetic average. It should be noted that the multiscale prediction is to fuse high and low frequency information to reduce the residual between the prediction and true value. Third, local variability of temporal change caused by land cover conversions or within-class differences is modelled well through the distribution of residuals. Fourth, Gaussian filter is applied to suppress the noise of HS_{t0} , HS_{u0} , HS_{up} , HS_{tp1} , and HS_{tp2} . Although the filtering makes the prediction image much blurrier and thus reduces the actual spatial resolution, a smaller standard deviation of Gaussian filter has little influence on the blurriness and can reduce the singular points of the predicted result.

Low computational efficiency is a key factor restricting the widespread application of spatiotemporal fusion models [12]. In terms of the tested fire area (4444.64 km²), NDVI_FMHC and NDVI-LMGM only require 4.10 min and 2.58 min, while ESTARFM and FSDAF require 105.37 min and 111.52 min. Due to multiple parallel processing of NDVI_FMHC, the computational efficiency is lower than that of NDVI-LMGM. However, the calculation efficiency of NDVI_FMHC is much higher than that of ESTARFM and FSDAF for two main reasons, including the use of advanced programming strategies, such as parallel computing, and no need to use ENVI and ArcMap software, reducing the function module calling time in the prediction process.

5.2. Optimal Parameters of the NDVI_FMHC

To obtain the best prediction effects and spend the least time, a large number of trial and error experiments were performed for the first four groups of test datasets, and the optimal parameters of the NDVI_FMHC were determined involving four main parameters: count of classes, distance calculation method, coarse pixel width, and Gaussian filter parameters. The optimal parameters of the NDVI_FMHC are based on MODIS and Landsat data while might need to be adjusted by future users to other satellite data spatiotemporal fusion.

The count of classes of hierarchical clustering based on histogram features is determined by users according to experience, and it is valuable to test the influence of count of classes on prediction accuracy. Supplemental Table S1 is the quantitative index evaluation table from 2 to 10 count of classes. In terms of trends, AAD, AE, and RMSE of cultivated land and flood areas decrease rapidly first with the increase in count of classes, and then tend to stabilise, while the changes in r and SSIM are opposite; AAD, AE, and RMSE of forests and urban areas increase slowly at first with the increase in count of classes, and then increase rapidly, while the changes in r and SSIM are opposite. In comparison, when the count of classes is four, the prediction effects of the NDVI_FMHC are relatively satisfactory.

When NDVI is classified by hierarchical clustering based on histogram features, the relationship between feature points in feature space is measured by pairwise distances. Therefore, the definition of distance measurement is of great significance to the study of the structure of feature space. In this study, we have analyzed the performance of six different distance measures including KL, L_1 , L_2 , L_∞ , \cap , and χ^2 [42]. By analyzing the quantitative index evaluation tables of six different distance measures (Table S2), for different distance measures, the difference in prediction results of the NDVI_FMHC is very small, indicating that the different distance measures have little effect on the prediction performance of the new spatiotemporal fusion framework. In addition, because L_1 and χ^2 have the smallest AAD, AE, and RMSE, and the largest r and SSIM, it is recommended that L_1 and χ^2 are the formulas of distance measurement for the NDVI_FMHC.

Coarse pixel width is a parameter to determine the size and change in coarse pixel width, which is also one of the important parameters of the NDVI_FMHC. Supplemental Table S3 is the quantitative index evaluation table under the width from 3 to 10 coarse pixels. In terms of trends, AAD, AE, and RMSE of the four test areas slowly decrease with the increase in the coarse pixel width, and then increase rapidly, while the changes in r and SSIM are opposite. The coarse pixel width may introduce a large error to the prediction result, leading to the decrease in the prediction effect, and the recommended coarse pixel width is 3 or 4.

Gaussian filtering is a kind of linear smoothing filter, and the smoothing degree depends on the standard deviation. The larger the standard deviation, the more dispersed the distribution, and the filter effect will be closer to the mean filter; the smaller the standard deviation, the more concentrated the distribution, and the filter effect will be weaker. Supplemental Table S4 is the quantitative index evaluation table under the change in standard deviation from 0.1 to 1. In terms of trends, AAD, AE, and RMSE of the four test areas first decrease with the standard deviation, then increase, and reach the minimum value at 0.5 or 0.6, while the change rule of r and SSIM is opposite, which means that the standard deviation is too small to eliminate the noise mixed with the prediction results, while the standard deviation is too large, some of the accurate values in the prediction results will also be eliminated as noise. Therefore, the recommended standard deviations are 0.5 and 0.6.

5.3. Limitations of the NDVI_FMHC

The NDVI_FMHC shows good robustness and prediction accuracy in predicting phenological changes and shape changes; however, the following aspects still need to be improved. First, in urban expansion, shape changes are generally large in number and small in size, and the low-resolution NDVI of the predicted date and a pair of high and low-resolution NDVI of the basic date did not record these changes, increasing the prediction difficult. Second, because MODIS and Landsat images were collected at an interval of approximately 0.5 h and the flood occurred very quickly, the range of shape changes recorded by two kinds of NDVI is not consistent, causing unsatisfactory fusion results. Third, for the multiple types of land cover converted into a single type due to the fire, although current fusion models can capture the overall shape change, the local prediction effect is not ideal because the prediction results still retain the texture information of the basic Landsat_NDVI. If we can propose a better new framework or add auxiliary data (other high-resolution images), the prediction accuracy of the above complex shape changes may be improved.

6. Conclusions

This study proposes an effective high spatiotemporal NDVI fusion model based on histogram feature clustering to obtain NDVI data with high spatiotemporal resolution. Compared with the three fusion models of ESTARFM, NDVI-LMGM, and FSDAF, we also have found that except for fire areas, the NDVI_FMHC has the highest prediction accuracy, the best spatial detail retention, and the strongest ability to capture shape changes. The new spatiotemporal fusion framework proposed by the NDVI_FMHC can better predict phenological changes and shape changes because classification map 2 is obtained by combining HS_{t0} and HS_{tp1} , which can distinguish shape changes that occurred from those that did not occur during the prediction period. Meanwhile, the NDVI_FMHC also uses four strategies to reduce the error, including the construction of the overdetermined linear mixed model, Multiscale prediction, residual distribution, and Gaussian filtering.

Although the NDVI_FMHC has good robustness and prediction accuracy, it still needs to improve. For example, this model cannot predict small changes, instantaneous changes caused by floods, and the shape changes from the multiple types of land cover converted into a single type. In addition, this study does not test the fusion performance of the NDVI_FMHC on other vegetation indexes. We have designed a friendly and concise user interface for the NDVI_FMHC, which will be shared (https://github.com/xingxuejun1989/NDVI_FMHC) for interested users.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2072-4292/12/22/3774/s1>.

Author Contributions: X.X. and C.Y. jointly conceived and designed the experiments; X.X. developed the processing program and wrote the paper; X.X. and Y.J. jointly performed the experiments and analyzed the data; H.J., J.L. and G.L. reviewed and edited the paper as supervisor. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of China (Grant No. 41730752 and 41971277) and Key Program of Guizhou Provincial Science and Technology Foundation ([2018]2200).

Acknowledgments: The authors would like to thank to Daisy He and anonymous reviewers for their insightful review of and valuable comments on the manuscript, which helped to improve the quality of the paper. Thanks Xiaolin Zhu for providing the source code of FSDAF and ESTARFM.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kayastha, N.; Thomas, V.; Galbraith, J.; Banskota, A. Monitoring Wetland Change Using Inter-Annual Landsat Time-Series Data. *Wetlands* **2012**, *32*, 1149–1162. [[CrossRef](#)]
2. Baassou, B.; He, M.; Mei, S.; Zhang, Y. Unsupervised hyperspectral image classification algorithm by integrating spatial-spectral information. In Proceedings of the International Conference on Audio, Language & Image Processing, Shanghai, China, 16 July 2012.
3. Guo, B.; Zhou, Y.; Wang, S.; Tao, H. The relationship between normalized difference vegetation index (NDVI) and climate factors in the semiarid region: A case study in Yalu Tsangpo River basin of Qinghai-Tibet Plateau. *J. Mt. Sci. Engl.* **2014**, *11*, 926–940. [[CrossRef](#)]
4. Wang, J.; Zhao, Y.; Li, C.; Yu, L.; Liu, D.; Gong, P. Mapping global land cover in 2001 and 2010 with spatial-temporal consistency at 250 m resolution. *ISPRS J. Photogramm. Remote Sens.* **2015**, *103*, 38–47. [[CrossRef](#)]
5. Fan, X.; Liu, Y. A global study of NDVI difference among moderate-resolution satellite sensors. *ISPRS J. Photogramm. Remote Sens.* **2016**, *121*, 177–191. [[CrossRef](#)]
6. De la Casa, A.; Ovando, G.; Bressanini, L.; Martinez, J.; Diaz, G.; Miranda, C. Soybean crop coverage estimation from NDVI images with different spatial resolution to evaluate yield variability in a plot. *ISPRS J. Photogramm.* **2018**, *146*, 531–547. [[CrossRef](#)]
7. Laibao, L.; Yang, W.; Zheng, W.; Delong, L.; Yatong, Z.; Dahe, Q.; Shuangcheng, L. Elevation-dependent decline in vegetation greening rate driven by increasing dryness based on three satellite NDVI datasets on the Tibetan Plateau. *Ecol. Indic.* **2019**, *107*, 105569.
8. Chu, H.; Venevsky, S.; Wu, C.; Wang, M. NDVI-based vegetation dynamics and its response to climate changes at Amur-Heilongjiang River Basin from 1982 to 2015. *Sci. Total Environ.* **2019**, *650*, 2051–2062. [[CrossRef](#)]
9. Zhang, Y.; Ling, F.; Foody, G.M.; Ge, Y.; Boyd, D.S.; Li, X.; Du, Y.; Atkinson, P.M. Mapping annual forest cover by fusing PALSAR/PALSAR-2 and MODIS NDVI during 2007–2016. *Remote Sens. Environ.* **2019**, *224*, 74–91. [[CrossRef](#)]
10. Gao, F.; Masek, J.; Schwaller, M.; Hall, F. On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2207–2218.
11. Zhu, X.; Jin, C.; Feng, G.; Chen, X.; Masek, J.G. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sens. Environ.* **2010**, *114*, 2610–2623. [[CrossRef](#)]
12. Zhu, X.; Cai, F.; Tian, J.; Williams, T. Spatiotemporal Fusion of Multisource Remote Sensing Data: Literature Survey, Taxonomy, Principles, Applications, and Future Directions. *Remote Sens. Basel* **2018**, *10*, 527.
13. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [[CrossRef](#)]
14. Fensholt, R.; Rasmussen, K.; Nielsen, T.T.; Mbow, C. Evaluation of earth observation based long term vegetation trends—Intercomparing NDVI time series trend analysis consistency of Sahel from AVHRR GIMMS, Terra MODIS and SPOT VGT data. *Remote Sens. Environ.* **2009**, *113*, 1886–1898. [[CrossRef](#)]
15. Abdollah, J.; Tim, M.V.; Thomas, V.N.; Irina, E.; John, C.; Kasper, J. Blending landsat and MODIS data to generate multispectral indices: A comparison of “index-then-blend” and “Blend-Then-Index” approaches. *Remote Sens. Basel* **2014**, *6*, 9213–9238.
16. Chen, X.; Liu, M.; Zhu, X.; Chen, J.; Zhong, Y.; Cao, X. “Blend-then-Index” or “Index-then-Blend”: A Theoretical Analysis for Generating High-resolution NDVI Time Series by STARFM. *Photogramm. Eng. Remote Sens.* **2018**, *84*, 65–73. [[CrossRef](#)]
17. Liu, X.; Deng, C.; Chanussot, J.; Hong, D.; Zhao, B. StfNet: A Two-Stream Convolutional Neural Network for Spatiotemporal Image Fusion. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 6552–6564. [[CrossRef](#)]

18. Rao, Y.; Zhu, X.; Chen, L.; Wang, J.; Chen, J. An Improved Method for Producing High Spatial-Resolution NDVI Time Series Datasets with Multi-Temporal MODIS NDVI Data and Landsat TM/ETM+ Images. *Remote Sens. Basel* **2015**, *7*, 7865–7891. [[CrossRef](#)]
19. Belgiu, M.; Stein, A. Spatiotemporal Image Fusion in Remote Sensing. *Remote Sens. Basel* **2019**, *11*, 818. [[CrossRef](#)]
20. Hilker, T.; Wulder, M.A.; Coops, N.C.; Linke, J.; Mcdermid, G.; Masek, J.G.; Gao, F.; White, J.C. A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on Landsat and MODIS. *Remote Sens. Environ.* **2009**, *113*, 1613–1627. [[CrossRef](#)]
21. Wu, P.; Shen, H.; Zhang, L.; Göttsche, F.M. 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.* **2015**, *156*, 169–181. [[CrossRef](#)]
22. Ma, J.; Zhang, W.; Marinoni, A.; Gao, L.; Zhang, B. An Improved Spatial and Temporal Reflectance Unmixing Model to Synthesize Time Series of Landsat-Like Images. *Remote Sens. Basel* **2018**, *10*, 1388. [[CrossRef](#)]
23. Bo, H.; Zhang, H. Spatio-temporal reflectance fusion via unmixing: Accounting for both phenological and land-cover changes. *Int. J. Remote Sens.* **2014**, *35*, 6213–6233.
24. Zhu, X.; Helmer, E.H.; Gao, F.; Liu, D.; Chen, J. A flexible spatiotemporal method for fusing satellite images with different resolutions. *Remote Sens. Environ.* **2016**, *172*, 165–177. [[CrossRef](#)]
25. Huang, B.; Song, H. Spatiotemporal Reflectance Fusion via Sparse Representation. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3707–3716. [[CrossRef](#)]
26. Chen, B.; Huang, B.; Xu, B. A hierarchical spatiotemporal adaptive fusion model using one image pair. *Int. J. Digit. Earth* **2017**, *10*, 639–655. [[CrossRef](#)]
27. Zhong, D.; Zhou, F. Improvement of Clustering Methods for Modelling Abrupt Land Surface Changes in Satellite Image Fusions. *Remote Sens. Basel* **2019**, *11*, 1759. [[CrossRef](#)]
28. Zhao, Y.; Huang, B.; Song, H. A robust adaptive spatial and temporal image fusion model for complex land surface changes. *Remote Sens. Environ.* **2018**, *208*, 42–62. [[CrossRef](#)]
29. Maselli, F.; Rembold, F. Integration of LAC and GAC NDVI data to improve vegetation monitoring in semi-arid environments. *Int. J. Remote Sens.* **2002**, *23*, 2475–2488. [[CrossRef](#)]
30. Liao, L.; Song, J.; Wang, J.; Xiao, Z.; Jian, W. Bayesian Method for Building Frequent Landsat-Like NDVI Datasets by Integrating MODIS and Landsat NDVI. *Remote Sens. Basel* **2016**, *8*, 452. [[CrossRef](#)]
31. Liu, M.; Yang, W.; Zhu, X.; Chen, J.; Chen, X.; Yang, L.; Helmer, E.H. An Improved Flexible Spatiotemporal DATA Fusion (IFSDAF) method for producing high spatiotemporal resolution normalized difference vegetation index time series. *Remote Sens. Environ.* **2019**, *227*, 74–89. [[CrossRef](#)]
32. Chen, X.; Li, W.; Chen, J.; Rao, Y.; Yamaguchi, Y. A Combination of TsHARP and Thin Plate Spline Interpolation for Spatial Sharpening of Thermal Imagery. *Remote Sens.* **2014**, *6*, 2845–2863. [[CrossRef](#)]
33. Xie, D.; Zhang, J.; Zhu, X.; Pan, Y.; Liu, H.; Yuan, Z.; Yun, Y. 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* **2016**, *16*, 207. [[CrossRef](#)] [[PubMed](#)]
34. Weng, Q.; Fu, P.; Gao, F. Generating daily land surface temperature at Landsat resolution by fusing Landsat and MODIS data. *Remote Sens. Environ.* **2014**, *145*, 55–67. [[CrossRef](#)]
35. Zhang, H.K.; Zhang, M.; Huang, B.; Cao, K.; Yu, L. A generalization of spatial and temporal fusion methods for remotely sensed surface parameters. *Int. J. Remote Sens.* **2015**, *36*, 4411–4445. [[CrossRef](#)]
36. Sun, Y.; Zhang, H.; Shi, W. A spatio-temporal fusion method for remote sensing data Using a linear injection model and local neighbourhood information. *Int. J. Remote Sens.* **2019**, *40*, 2965–2985. [[CrossRef](#)]
37. Emelyanova, I.V.; Mcvicar, T.R.; Van Niel, T.G.; Li, L.T.; Van Dijk, A.I.J.M. 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.* **2013**, *133*, 193–209. [[CrossRef](#)]
38. Cammalleri, C.; Anderson, M.C.; Gao, F.; Hain, C.R.; Kustas, W.P. A data fusion approach for mapping daily evapotranspiration at field scale. *Water Resour. Res.* **2013**, *49*, 4672–4686. [[CrossRef](#)]
39. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)]
40. Yao, W.; Loffeld, O.; Datcu, M. Application and Evaluation of a Hierarchical Patch Clustering Method for Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 2279–2289. [[CrossRef](#)]

41. Dubey, S.K.; Vijay, S. A Review of Image Segmentation using Clustering Methods. *Int. J. Appl. Eng. Res.* **2018**, *13*, 2484–2489.
42. Buch, A.G.; Kraft, D. Local Point Pair Feature Histogram for Accurate 3D Matching. In Proceedings of the 29th British Machine Vision Conference, BMVC, Newcastle, UK, 11 September 2018; pp. 1–12.

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).