

Article

Combining Linear Pixel Unmixing and STARFM for Spatiotemporal Fusion of Gaofen-1 Wide Field of View Imagery and MODIS Imagery

Jintian Cui ^{1,2}, Xin Zhang ^{1,*} and Muying Luo ³

¹ State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100101, China; cuijintian17@mails.ucas.ac.cn

² University of Chinese Academy of Sciences, Beijing 100101, China

³ School of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road, Wuhan 430079, China; luomuying@whu.edu.cn

* Correspondence: zhangxin@radi.ac.cn; Tel.: +86-10-6488-7892

Received: 5 May 2018; Accepted: 27 June 2018; Published: 3 July 2018



Abstract: Spatiotemporal fusion of remote sensing data is essential for generating high spatial and temporal resolution data by taking advantage of high spatial resolution and high temporal resolution imageries. At present, the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) is one of the most widely used spatiotemporal fusion technologies of remote sensing data. However, the quality of data acquired by STARFM depends on temporal information from homogeneous land cover patches at the MODIS (Moderate Resolution Imaging Spectroradiometer) imagery, and the estimation accuracy of STARFM degrades in highly fragmented and heterogeneous patches. To address this problem, we developed an innovative method to improve fusion accuracy, especially in areas of high heterogeneity, by combining linear pixel unmixing and STARFM. This method derived the input data of STARFM by downscaling the MODIS data with a linear spectral mixture model. Through this fusion method, the complement effect of the advantages of remote sensing information can be realized, and the multi-source remote sensing data can be realized for visual data mining. The developed fusion method was applied in Bosten Lake, the largest freshwater lake in China, and our analysis of results suggests that (1) after introducing the linear spectral mixture model, the fusion images illustrated improved spatial details to a certain extent and can be employed to identify small objects, as well as their texture distribution information; (2) for fragmented and highly heterogeneous areas, a stronger correlation between the predicted results and the real images was observed when compared to STARFM with small bias; and (3) the predicted red band and near infrared band can generate high-precision 16-m NDVI (Normalized Difference Vegetation Index) data with advantages in both spatial resolution and temporal resolution. The results are generally consistent with the Gaofen-1 wide field of view cameras (GF-1 WFV) NDVI in the same period and therefore can reflect the spatial distribution of NDVI in detail.

Keywords: high spatiotemporal resolution; data fusion; spectral mixture analysis; STARFM; GF-1 WFV; MODIS; NDVI

1. Introduction

Due to the development of remote sensing technology and the application of different Earth observation satellite sensors, remote sensing data obtained from the same area can present multi-platform, multi-sensor, multi-spectral, multi-resolution, and multi-temporal characteristics. Each type of remote sensing data has its advantages, but the information reflected or expressed is far from satisfactory for practical applications. Generally speaking, high spatial resolution sensors

typically provide a smaller image footprint, or spatial extent, thereby increasing the time it takes a satellite to revisit the same location on Earth [1]. Conversely, high temporal resolution sensors have more frequent revisit rates and produce wide area coverage with lower spatial resolution. Compared with single pieces of remote sensing data, the superiority of data with high spatiotemporal resolution is mainly reflected in their complementarity. The fusion of multi-source remote sensing data can result in more abundant and more accurate information than any single piece of remote sensing data. Studies have applied multi-sensor data fusion of medium- and high-resolution imagery for applications such as phenology analysis, management of wetlands, vegetation dynamics monitoring, land cover classification, and land surface temperature retrieval [2–11]. The first satellite of China's high-resolution Earth observation system, GaoFen-1 (GF-1), was successfully launched on 26 April 2013. The GF-1 satellite carries two high-resolution cameras with 2-m panchromatic and 8-m multi-spectral resolution, as well as four wide field imager cameras with 16-m multi-spectral resolution. GF-1 uses the key technologies of high spatial resolution and multi-spectral and wide swath, and it is widely used in the applications of dynamic monitoring and change detection, geological and mineral resource surveying, ecological environment investigation, geological hazard investigation, and so on [12–17]. However, due to a long revisit rate and cloud coverage, it is difficult to obtain continuous images for the same period in a large area. On the contrary, Moderate Resolution Imaging Spectroradiometer (MODIS) can provide images at varying spatial resolutions (from 250-m to 1000-m) with a temporal resolution of 1 day, therefore it can fully increase the accessibility of cloud-free imagery. MODIS data is widely used in the monitoring of land cover types on a large scale [18–20]. However, the spatial resolution of MODIS data is relatively low, and thus it cannot meet the requirements for the monitoring land cover and ecosystem changes in highly heterogeneous areas [21]. Therefore, the fusion of GF-1 data and the MODIS data can fully utilize the advantages of space, spectra, angles, and temporality, and obtain data with high temporal and spatial resolution at the same time. The availability and the application potential of data can then be improved greatly.

In recent years, scholars have proposed a series of data fusion models for obtaining high spatiotemporal resolution remote sensing data. One type is the linear spectral mixture model. With this model, Zhukov et al. [22] and Maselli et al. [23] considered the spatial variability of pixel reflectivity, selecting a subset involved in the calculation of the target pixel through a certain rule, such as the neighborhood centered on the target pixel or a certain distance from the target pixel. Furthermore, the spectral weight was proposed by Busetto et al. [24] to reduce the number of pixels with close spatial distances but large spectral differences. The accuracy of predicted NDVI (Normalized Difference Vegetation Index), however, decreased when the MODIS pixels corresponded to different seasonal crops. Cherchali et al. [25] presented a method for radiometrically unmixing coarse resolution signals through the inversion of linear mixture modeling of large regions to obtain high spatial and temporal resolution signals, but the precision was not high. However, the above methods are all fusion algorithms based directly on the linear spectral mixture model. The algorithm is simple and easy to use, but the class reflectance is the average reflectance of the whole area, and thus cannot reflect the spatial variability of the spectra of land cover types and the spatial variability of pixel reflectance. Therefore, this kind of fusion algorithm will result in serious image spots and is difficult to reflect the details of the features, especially in forest, farmland, and other large areas of similar land use types.

Some scholars improved the linear spectral mixture model and proposed the corresponding fusion model under the condition that the pixel reflectance is influenced by the environment. Gao et al. [26] proposed the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) to blend Landsat and MODIS data for predicting daily surface reflectance at Landsat spatial resolution and MODIS temporal resolution. This algorithm not only considers the distance in space between pixels and their neighborhood and spectral similarity but also considers the time difference. The STARFM algorithm has high prediction accuracy at a local scale and is widely used [27–30]. However, the STARFM still has some shortcomings: (1) if transient or abrupt change information of the Earth's surface is not recorded by the base Landsat images, the fusion image cannot capture the information [31]; (2) the bidirectional

reflectance distribution function (BRDF) effect; and (3) for broken and highly heterogeneous areas, the STARFM fusion result has lower accuracy. These shortcomings are due to the fact that when the STARFM predicts the centric pixel, there is a greater weight for the sub pixel of the MODIS pure pixel, and as such the influence of the MODIS mixed pixels on the reflectance estimation of the centric pixel is reduced. To this end, some researchers developed some improved algorithms based on the STARFM. For problem (1), Hilker et al. [32] proposed the Spatial Temporal Adaptive Algorithm for Mapping Reflectance Change (STAARCH) spatiotemporal fusion algorithm. This algorithm can identify the area of change and the specific time, and improve the fusion accuracy by selecting the data at the best time. For the BRDF effect mentioned in problem (2), Roy et al. [33] proposed a fusion method using a semi-physical model. This method fused the Landsat ETM+ data and MODIS BRDF/Albedo surface data to predict the ETM data on the corresponding date, and effectively solved the problem of the BRDF effect. However, as the method is based on the geometric optical model, the efficiency of preprocessing is low. In addition, it did not consider the neighbor spectral information of the centric pixels, the fusion result showed some patches, and the overall prediction result was poor. For heterogeneous landscapes, Zhu et al. [34] proposed the Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model (ESTARFM) based on the STARFM, which used the observed reflectance trend between two points in time and assumed that the different values of Landsat data at different times have a linear relationship. The increment coefficient and the weight were calculated using two Landsat images and three MODIS images at different times, and, finally, the predicted image was obtained accordingly. However, the algorithm needs to have at least two cloudless high spatial resolution images, thus increasing the difficulty of data acquisition. In addition, a number of fusion methods based on learning models, which mainly used compressed sensing or sparse representation [35], such as K-SVD algorithm [36] and spatiotemporal fusion model based on image learning [37], were proposed.

Compared with previous research, this study considers the following two aspects, and a new method was proposed. (1) For the fusion data sources, most of the research focused on the medium-resolution data and MODIS data, rarely using higher resolution data. Therefore, in this study, the Gaofen-1 wide field of view cameras (GF-1 WFV) data and MODIS data were used for fusion to obtain high spatiotemporal resolution data; and (2) taking into account the problem (3) of STARFM, the estimation accuracy of STARFM usually decreases when applied to heterogeneous areas. Due to the MODIS data with low spatial resolution, the phenomena of mixed pixels is particularly serious, especially in broken and highly heterogeneous areas. Moreover, the STARFM uses the resampled MODIS data, resulting in strong homogeneity among the resampled pixels within the same MODIS pixel and a large spectrum difference between neighboring MODIS pixels. Especially in the heterogeneous area, the difference is even greater. To address these problems, a combining linear pixel unmixing and STARFM fusion model was proposed in this paper. This method used the linear spectral mixture model to downscale the MODIS data and solved the mixed pixel problem at the sub-pixel level, and used endmembers and their fractions to perform the analysis based on spectral features, thus obtaining the surface reflectance of land cover types in each pixel of the MODIS data. Secondly, the downscaling data were used to replace the resampled low-resolution data in the STARFM, which were then fused with the GF-1 WFV image to produce the predicted image. Finally, the predicted result of the proposed method, the predicted result of the STARFM, and the real image at the same time were compared and analyzed, and this method was used to predict the heterogeneity and homogeneity of the Earth's surface and NDVI inversion in order to verify the effectiveness of the proposed algorithm.

2. Study Area and Datasets

2.1. Study Area

Bosten Lake is located in the arid area of northwest China between 41°56' and 42°14'N, and 86°40' and 87°26'E. It is the largest inland freshwater Lake in China, with an area of about 1100 km².

The geographic extent of the study area is shown in Figure 1. The lake is 1048-m above sea level, with an average depth of 9-m, reaching 17-m in the deepest areas. There are more than 10 rivers flowing into Bosten Lake, but only the Kaidu River is perennial, for which the average annual inflow into the lake is $34.2 \times 10^8 \text{ m}^3$. This river basin lies in central Eurasia, with low precipitation, great evaporation, and long periods of sunshine. The annual mean precipitation is 68.2-mm, but the annual mean evaporation ranges from 1800 to 2000-mm, resulting in Bosten Lake having a continental desert climate. As a wetland ecosystem in the arid region of China, the Bosten Lake wetlands play an important role not only in maintaining regional ecological balance but also in maintaining biodiversity.

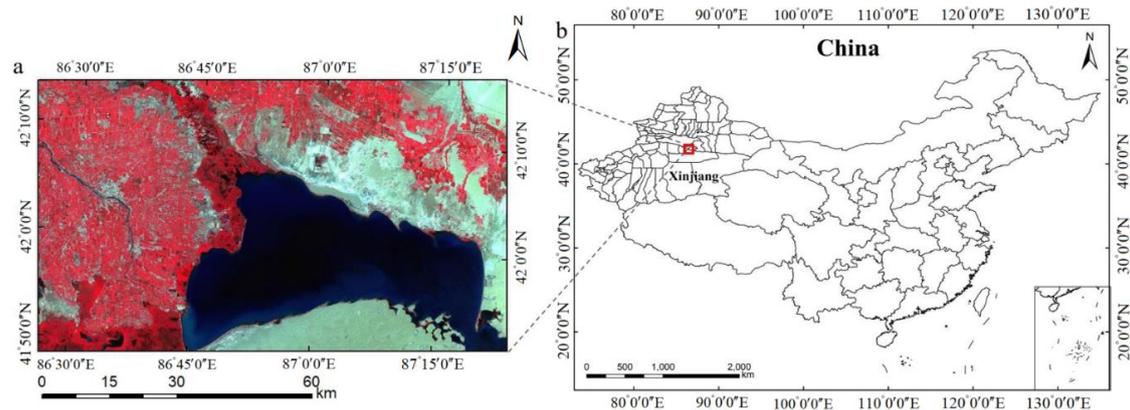


Figure 1. (a) Bosten Lake and (b) administrative boundary of China.

2.2. Datasets

2.2.1. GF-1 WFV Data

In this study, four GF-1 WFV scenes with low cloud cover in 2016 were collected. The dates of acquisition were 12 June 2016, 10 July 2016, 9 August 2016, and 21 August 2016. For the experiment, GF-1 WFV multispectral bands 1~4 were selected, for which the spatial resolution was 16-m. Preprocessing, including radiation calibration, atmospheric correction, and geometry correction, were performed. The radiation calibration parameters were obtained from the China Centre for Resources Satellite Data and Application, and the Second Simulation of the Satellite Signal in the Solar Spectrum model (6S model) [38] was used to perform the atmospheric correction in order to obtain the surface reflectance results. Based on the Landsat 8 panchromatic image with accurate spatial coordinate information, geometry correction was accomplished using the quadratic polynomial correction algorithm and adding the same name control point. The root mean square error was less than 0.5 pixels.

2.2.2. MODIS Data

500-m MODIS surface reflectance product MOD09GA (path:h24/v04) corresponding to GF-1 WFV data in time was selected. All MODIS images were re-projected into the UTM-WGS84 coordinate system using the MODIS Reprojection Tool (MRT) available from the United States Geological Survey (USGS) to ensure the same reference system with GF-1 WFV images, and then converted to the Geo-TIF data format. The bilinear interpolation method was used to resample the pixel size to a spatial resolution of 480-m, which was convenient for the subsequent spectral mixture analysis. Geometry correction was performed to MODIS data so it would correspond to the GF-1 WFV data in space. Finally, the GF-1 WFV data and the processed MODIS data were resized using ENVI software, and the data required for this study were obtained, as shown in Tables 1 and 2.

Table 1. Band information of GF-1 WFV and MOD09GA data.

GF-1 (WFV)					MODIS				
Band Number	Band	Wavelength/nm	Spatial Resolution/m	Temporal Resolution/d	Band Number	Band	Wavelength/nm	Spatial Resolution/m	Temporal Resolution/d
1	Blue	450–520	16	4	3	Blue	459–479	500	1
2	Green	520–590			4	Green	545–565		
3	Red	630–690			1	Red	620–670		
4	NIR ¹	770–890			2	NIR ¹	841–876		

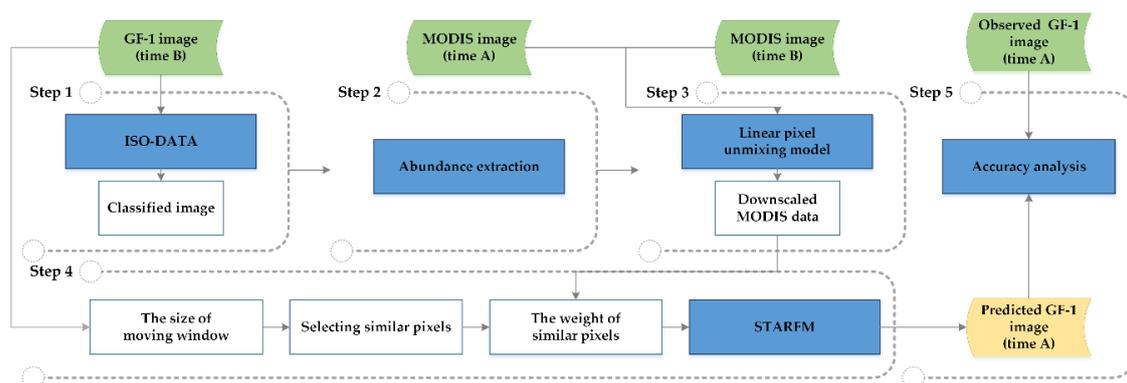
¹ is the near-infrared band.

Table 2. Main Purpose of the GF-1 WFV and MODIS data used in the experiment.

Data	Date	Purpose
GF-1 (WFV)	06-12	(1) Classification
	08-09	(2) Selection of similar pixels
	07-10 08-21	Accuracy assessment
MOD09GA	06-12	(1) Linear spectral mixture analysis
	07-10	
	08-09	(2) Calculation of the mean endmember spectral reflectance
	08-21	

3. Methodology

In this study, the combining linear pixel unmixing and STARFM fusion model with MODIS and GF-1 WFV data was used to estimate high spatiotemporal resolution imagery. As the estimation accuracy of STARFM usually declines when applied to heterogeneous areas, the linear spectral mixture model is used to downscale the MODIS data in order to improve the data fusion accuracy. The base data includes the high- and low-resolution images in the same period, and the low-resolution image of the predicted moment. The algorithm includes the following steps: (1) obtaining high-resolution (GF-1 WFV) classification image, (2) extracting endmembers and calculating the endmember fractions in order to estimate each class contribution to the signal of the lower spatial resolution image, (3) restoring the unmixed image pixel, (4) inputting the processed data into the STARFM to predict the high spatiotemporal resolution image, and (5) accuracy evaluation of predicted and real image. The flowchart is shown in Figure 2.

**Figure 2.** Flow chart of combining linear pixel unmixing and STARFM for high spatiotemporal fusion imagery.

3.1. MODIS Linear Spectral Mixture Analysis

Because of the complexity and diversity of nature, and the limitation of sensor spatial resolution, the mixed pixels are prevalent in remote sensing imagery. Therefore, the spectra of pixels in the image

do not comprise the spectrum of a single land cover type but a mixed spectrum as a combination of spectra for pure land cover types. MODIS data have low spatial resolution because of their large field of view, and mixed pixels are common. For broken and highly heterogeneous areas, the mixing phenomenon is more serious. When the STARFM predicts the centric pixel, the sub pixel of the MODIS pure pixel has a greater weight, and thus the influence of MODIS mixed pixels on the reflectance estimation of the centric pixel is reduced. However, it is very difficult to find pure MODIS pixels for the heterogeneous area, so the fusion result has lower accuracy. For this issue, the reflectance of land cover classes in each pixel can be used to replace the MODIS reflectance in the STARFM algorithm if it can be obtained based on the MODIS reflectance, and then it can be used to reduce the effect of MODIS mixed pixels on the accuracy of the prediction results and improve the quality of fusion data. Therefore, in this study, a linear spectral mixture model was used to downscale the low-resolution data and obtain the reflectance of land cover classes in each MODIS pixel. The mechanism of the linear spectral mixture model is shown in Figure 3. Thus, the downsampled data is utilized to replace the resampled low-resolution data in the STARFM in order to improve the estimation accuracy of the heterogeneous areas.

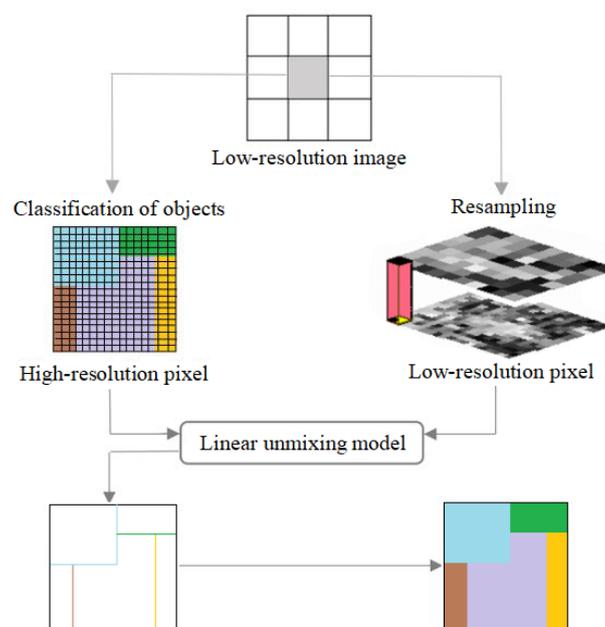


Figure 3. Mechanism of the linear spectral mixture model.

Endmember Selection and Fraction Calculation

At present, there are five types of spectral unmixing model: the linear model, the probabilistic statistical model, the random geometry model, the geometric-optical model, and the fuzzy model [39]. The linear model is the most widely used among them. The model describes a linear relationship among each land cover type, the surface fraction they cover, and their spectrum in each pixel of remote sensing imagery. The method continues to subdivide on a single-pixel basis, and the modeled spectra represents the linear summation of the spectrum of each land cover type. For each land cover type, the endmember is multiplied by the surface fraction they cover (called the weight coefficient) to describe the mixing situation. Endmember selection and fraction calculation are two important processes in spectral mixture analysis.

The mixed pixels are unmixed into different basic component units, called endmembers, which are important links to understand the image pixel spectra and reference spectra. In this study, the types of endmembers were obtained by the classification result of high-resolution images through the Iterative Self-Organizing Data Analysis Technology Algorithm (ISO-DATA) [40]. ISO-DATA is one of the

most commonly used classification algorithms for remote sensing images. The main principles of this algorithm are as follows. First, some samples are selected as the initial cluster centers randomly, and pixels are assigned based on the shortest-distance-to-the-center method. Then, it can be determined whether the initial clustering result meets the requirements. If not, clusters will be split or merged, and new cluster centers will be obtained. Through continuous splitting and merging, repeated iterations are performed until the clustering division is complete. Through the ISO-DATA, it can obtain high-resolution classification maps. Equation (1) is utilized to extract the fraction of each endmember in each mixed pixel. Assuming that the types and fraction do not change over time, Arc GIS is used to create a grid with a low-resolution scale on the classification map to obtain the fraction of each endmember in each grid.

$$f_c(i, c) = M/N \quad (1)$$

in which $f_c(i, c)$ is the fraction of land cover type c in low-resolution pixels at position i , M is the number of pixels of the land cover type c in low-resolution pixels, and N is the number of pixels of all land cover types in low-resolution pixels.

The fraction matrix and the reflectance of low spatial resolution pixels are input to perform spectral mixture analysis of MODIS data using the least-squares method in a certain size window (MODIS pixel). Then, the mean reflectance of each endmember in the window at time t can be obtained. The resulting value is put into the position of the corresponding pixel in the window according to the endmember type, and the data of spectral mixture analysis can be obtained. Finally, the resulting image of spectral mixture analysis is input to the STARFM rather than the original reflectance image of MODIS to estimate the high spatial resolution image at a given time. The linear mixture model is

$$R(i, t) = \sum_{c=0}^k f_c(i, c) \times \bar{r}(c, t) + \xi$$

$$\sum_{c=0}^k f_c(i, c) = 1$$

$$f_c(i, c) \geq 0 \quad (2)$$

in which $R(i, t)$ is the spectral reflectance of the low-resolution pixel of position i at time t , $\bar{r}(c, t)$ is the mean spectral reflectance of land cover type c at time t , k is the number of land cover types within the window, and ξ is an error term.

3.2. The Theoretical Basis of the STARFM

The STARFM is a kind of spatiotemporal data fusion model. This model acquires temporal information from low spatial resolution data, and spatial information on the distribution of ground objects from high spatial resolution data, using fusion technology to obtain high spatial and temporal resolution products.

Through obtaining the high-resolution data and the low-resolution data at the same time t_m , combined with the low-resolution at the predicted time t_0 , the STARFM calculates the spatial distribution difference between images to predict the high spatiotemporal data at t_0 . In the case of ignoring the registration and atmospheric correction error, the low-resolution image of homogeneous area of a single land cover type at t_m can be obtained by taking a weighted average of the high-resolution images the same time. Their reflectance would be biased considering the different spectra and temporal resolution and the variability of imaging light conditions. Assuming that the land cover type and system errors do not vary with time, the reflectance deviation is constant, and the high spatial resolution image at time t_0 can be obtained from the following Equation (3):

$$G(x_i, y_j, t_0) = M(x_i, y_j, t_0) + G(x_i, y_j, t_m) - M(x_i, y_j, t_m) \quad (3)$$

in which $G(x_i, y_j, t_0)$, $M(x_i, y_j, t_0)$ are the high- and low-resolution image reflectance of the predicted moment t_0 , and $G(x_i, y_j, t_m)$, $M(x_i, y_j, t_m)$ are the high- and low-resolution image reflectance of t_m .

Because of the complexity of the actual land types, the image is usually made up of heterogeneous pixels. Therefore, the derivation process of the above conditions cannot be satisfied. So, using the information of neighboring pixels of the target pixel, a weighted function of spectra, time, and space factors is constructed to calculate the reflectance value of the target pixel. The selection of similar pixels is an important step in the algorithm to construct the weighted function of the relationship between neighboring pixels and central pixels. The homogeneous neighboring pixels with similar spectral features are used as reference information to calculate the spectral information contribution of the similar pixels to the centric pixel in order to improve the accuracy of the prediction. To ensure that the neighboring pixels provide the correct auxiliary information, the similar pixels selected in this study are determined by the threshold value of each band of the high-resolution image at t_0 , meaning that the nearest pixel to the centric pixel in the moving window with spectral variability lower than the threshold will be a candidate pixel (Equation (4)). The threshold method has the characteristics of local applicability. Based on the surface features of each area, the predicted similar pixels of the area are calculated. Because the judgment standard varies with the position of the moving window, the selection bias of local similar pixels does not affect the accuracy of the overall image prediction.

$$|G(x_i, y_i, t_0) - G(x_{\omega/2}, y_{\omega/2}, t_0)| \leq \sqrt{\frac{1}{N} \sum_{i=1}^N (G(x_i, y_i, t_0) - \mu)^2 / m} \quad (4)$$

in which $G(x_i, y_i, t_0)$ denotes the spectral reflectance of the pixel of position i , $G(x_{\omega/2}, y_{\omega/2}, t_0)$ denotes the spectral reflectance of the centric pixel, N denotes the number of high-resolution pixels in the moving window, μ denotes the mean spectral reflectance in the window of the band, and m denotes the number of land cover types.

In a certain sized sliding window, similar pixels are used to calculate the value of the centric pixel. The weight value given to the similar pixel is measured by three indicators: spectral variability, temporal variability, and the relative distance between pixels and their neighborhood (Equation (5)). S_{ijm} represents the reflectance difference between the high-resolution image and the low-resolution image, which can indicate the spectral variability of images. The smaller the parameter is, the greater the spectral similarity between the given location and its neighborhood, and the higher the weight should be given in the calculation. T_{ijm} represents the reflectance difference between the image at t_0 and the image at t_m . The smaller the value is, the smaller the spectral change in this period, and the higher the weight should be in the calculation. D_{ijm} represents the geometric distance of the similar pixel and the center pixel, and the smaller the value is, the higher the weight is.

$$\begin{aligned} S_{ijm} &= |G(x_i, y_j, t_m) - M(x_i, y_j, t_m)| \\ T_{ijm} &= |M(x_i, y_j, t_m) - M(x_i, y_j, t_0)| \\ D_{ijm} &= 1 + \sqrt{(x_{\omega/2} - x_i)^2 + (y_{\omega/2} - y_j)^2} / A \end{aligned} \quad (5)$$

in which A is a constant value that denotes the importance of spatial distance relative to the spectral and temporal variability.

The reflectance of the center pixel is calculated using the reflectance of the similar pixels in the moving window. The convolution operation on the pixels is performed through the weight function in the window. Then, the weighted average method is used to realize the prediction of the reflectance

value of the centric pixel, and slide the window to traverse the full image. Equation (6) is utilized to obtain the predicted image. The calculation equation of the reflectance of the centric pixel is as follows:

$$W_{ijm} = \left(\frac{1}{S_{ijm} \times T_{ijm} \times D_{ijm}} \right) / \sum_{i=1}^{\omega} \sum_{j=1}^{\omega} \sum_{m=1}^n \left(\frac{1}{S_{ijm} \times T_{ijm} \times D_{ijm}} \right)$$

$$G(x_{\omega/2}, y_{\omega/2}, t_0) = \sum_{i=1}^{\omega} \sum_{j=1}^{\omega} \sum_{m=1}^n W_{ijm} \times (M(x_i, y_j, t_0) + G(x_i, y_j, t_m) - M(x_i, y_j, t_m)) \quad (6)$$

in which $(x_{\omega/2}, y_{\omega/2})$ is the centric pixel in the window and $G(x_{\omega/2}, y_{\omega/2}, t_0)$ is the predicted high-resolution image reflectance of the pixel $(x_{\omega/2}, y_{\omega/2})$ at t_0 ; $G(x_i, y_j, t_m)$ and $M(x_i, y_j, t_m)$ are the high- and low-resolution image reflectances, respectively, of the pixel (x_i, y_j) at t_m ; $M(x_i, y_j, t_0)$ is the low-resolution image reflectance of the pixel (x_i, y_j) at t_0 ; ω is the size of the moving window; and W_{ijm} is the weight value given to the similar pixel.

3.3. Data Quality Evaluation Method of Spatiotemporal Fusion

Using the base images obtains high spatiotemporal resolution image of the predicted moment t_0 , including the GF-1 WFV image of the moment t_m , and the MODIS image after linear spectral mixture analysis of the moment t_m and of the predicted moment t_0 . In terms of accuracy evaluation, this paper first qualitatively evaluates fusion results using visual interpretation and compares the detailed features of the predicted and the real images. Secondly, the surface reflectance of each pixel of the predicted image and the real image at the same time is extracted in order to use the scatter plot based on reflectance to quantitatively evaluate the prediction period and real GF-1 WFV images, and the correlation coefficient (R) and root mean square error ($RMSE_b$) between reflectance data at the moment t_m and t_0 are calculated as Equation (7):

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

$$RMSE_b = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad (7)$$

in which n is the number of pixels, i is the pixel position of the predicted and the real image, x_i and y_i denote the pixel values of the predicted and the real image in position i , and \bar{x} and \bar{y} denote the average of the pixels for the predicted and the real image.

In the accuracy evaluation, R shows the similarity between the predicted values and the real values, which is an important parameter to measure relevance. The greater the value, the higher the degree of relevance, and the better the fusion results. $RMSE_b$ is the root mean square error of band b between the predicted images and the real images. The smaller the value, the higher the quality of the fusion data.

4. Results

4.1. Spectral Mixture Analysis of MODIS Data

To perform spectral mixture analysis of MODIS data, the first step was to obtain the initial classification result of GF-1 WFV imagery using ISO-DATA. The number of initial classification classes was 10, the number of minimum iterations was 20, and the number of the smallest type of patches was set to 20. Because the initial clustering result was broken, the results were merged to six classes

according to the comparison between classes, and then the majority filter was applied to the small patches. Finally, through the seeded region growing algorithm, small patches were reclassified in the final classification map. The land cover types in the study area were divided into cultivated land, forest land, wetland, bare land, water, and residential land. Next, a grid of 480-m \times 480-m was created with Arc GIS 10.2. The fraction of each endmember in each grid was extracted, and then the fraction image of each land cover type was obtained. As the distribution of land cover types in the study area is relatively single and centralized, in this paper, a piece of typical area was selected in the study area to display the fraction image of each land cover type based on the entire area for clearer display and comparison (Figure 4). The selected area covered cultivated land, forest land, bare land, water, and residential land obtained by classification. The fraction values of the land cover types obtained by the fraction images were normalized to 0~1. The brighter the image, the higher the proportion of the endmember, and the higher the fraction. The reverse is true for darker images.

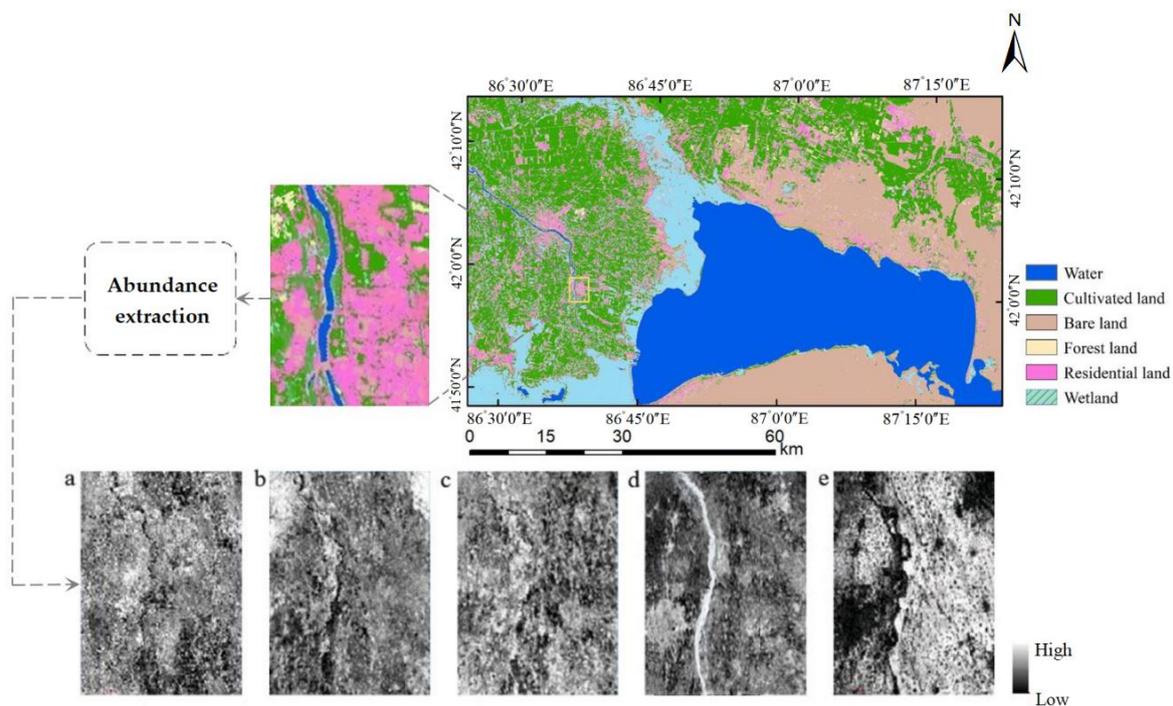


Figure 4. Fraction image of typical land cover types. (a) Cultivated land fraction, (b) forest land fraction, (c) bare land fraction, (d) water fraction, (e) residential land fraction.

The fraction matrix and MODIS image were input to perform spectral mixture analysis of MODIS data in the period of 12 June 2016 using the least-squares method in an appropriate size window. The sizes of the selected windows were 5 \times 5, 9 \times 9, 15 \times 15, 21 \times 21, and 31 \times 31 pixels (MODIS pixels). Based on the R, RMSE, and bias, accuracy assessment of the resulting imagery conducted by spectral mixture analysis with actual GF-1 WFV data was performed to obtain the best window size. Table 3 shows the R, RMSE, and bias of red, green, and NIR bands when using different window sizes. By analyzing the various accuracy assessment indexes, the result was the best when the window size was 15 \times 15 MODIS pixels. Therefore, this window size was chosen to perform spectral mixture analysis of MODIS data in the period of 10 July 2016, 9 August 2016, and 21 August 2016. Then, the reflectance of each land cover type in the window was obtained and put in the position of the corresponding pixel according to its type to obtain the downscaled MODIS data.

Table 3. Accuracy evaluation of downscaled MODIS data with different window sizes.

Window Size/MODIS Pixels	R			RMSE			Bias		
	Green	Red	NIR	Green	Red	NIR	Green	Red	NIR
5 × 5	0.7137	0.7960	0.8437	0.0303	0.0342	0.0587	0.0094	0.0062	0.0069
9 × 9	0.7649	0.8356	0.8769	0.0265	0.0283	0.0492	0.0094	0.0060	0.0066
15 × 15	0.8420	0.9048	0.9301	0.0228	0.0249	0.0448	0.0091	0.0059	0.0064
21 × 21	0.8776	0.8745	0.9170	0.0238	0.0241	0.0453	0.0092	0.0057	0.0064
31 × 31	0.8521	0.8823	0.9026	0.0253	0.0262	0.0486	0.0093	0.0058	0.0067

4.2. The Fusion of High Spatial and Temporal Resolution Imagery

Using the combining linear pixel unmixing and STARFM fusion model to deal with data, a high-resolution image of prediction time was obtained, and the predicted results were compared with the fusion image obtained by the STARFM algorithm to discuss the validity and reliability of the proposed algorithm for predicting the high spatiotemporal image. GF-1 WFV image and the unmixed MODIS image were used as reference data; the data from the GF-1 WFV image and MODIS image on 12 June, and the MODIS image on 10 July, were obtained as the first input of the model. The data from the GF-1 WFV image and MODIS image on 9 August, and the MODIS image on 21 August, were used as the second input of the model. This was to obtain fusion images with high spatiotemporal resolution on 10 July and 21 August.

When choosing the parameters in the combining linear pixel unmixing and STARFM fusion model for image fusion, the edge effect must be considered. When calculating the reflectance of the centric pixel, pixels which have similar spectra and are cloudless in the moving window, and can be involved in calculations. This is due to the fact that the actual distribution of ground objects is usually complex and ground objects may change the image feature greatly. Therefore, selecting the appropriate size of the moving window for similar pixel selection can greatly improve the accuracy of the prediction. Usually, the feature calculation result of the center pixel is not obvious if the window selected is too small. However, if the selected window is too large, the calculations will be greatly increased and the correlation between the edge pixels and the center pixel will be smaller, resulting in low accuracy of the prediction result. Considering the actual situation of the study area, the results were predicted at different window sizes, for example 9 × 9, 13 × 13, 33 × 33, 53 × 53, and 73 × 73 pixels (GF-1 WFV pixels). Table 4 shows the R, RMSE, and bias between the fusion data of red, green, and NIR bands and the real GF-1 WFV image on 10 July using different window sizes.

Table 4. Accuracy evaluation of prediction data by combining linear pixel unmixing and STARFM fusion model using different window sizes.

Window Size/GF-1 WFV Pixels	R			RMSE			Bias		
	Green	Red	NIR	Green	Red	NIR	Green	Red	NIR
9 × 9	0.8633	0.8764	0.9124	0.0333	0.0384	0.0462	−0.0020	0.0035	−0.0087
13 × 13	0.8694	0.8852	0.9187	0.0318	0.0376	0.0459	−0.0016	0.0032	−0.0079
33 × 33	0.8671	0.8890	0.9202	0.0317	0.0374	0.0424	−0.0012	0.0026	−0.0068
53 × 53	0.8620	0.8833	0.9195	0.0319	0.0370	0.0426	−0.0018	0.0030	−0.0072
73 × 73	0.8581	0.8801	0.9173	0.0321	0.0372	0.0430	−0.0023	0.0032	−0.0078

Through the analysis of accuracy with evaluation indexes using different window sizes, it can be found that the calculation results of each accuracy evaluation index were best when the size of the selected window was 33 × 33 pixels (GF-1 WFV pixels), so it was chosen for subsequent fusion processing. Figure 5 shows the scatter plots of the reflectance of red, green, and NIR bands of the predicted result using the combined linear pixel unmixing and STARFM fusion model and real GF-1 WFV imagery on 10 July and 21 August. The x-coordinate represents the real reflectance data, and the y-coordinate represents the predicted value of reflectance. From the correlation analysis, the scatter distribution is relatively concentrated. The majority of the scatter plots obtained by the combining

linear pixel unmixing and STARFM fusion model are distributed on both sides of $y = x$, but the scatter plots of 21 August are closer to the line $y = x$ than those of 10 July. Combining the scatter plots of each band to analyze the correlation coefficient between the fusion image and the real image, it can be found that the correlation between the fusion image and the real image was high. The correlation coefficients at the two phases of green and red bands were all over 0.8. Comparing the correlation of each band at different times, it can be found that the correlation between the predicted results and the real data increased from 10 July to 21 August. For the two prediction days, the correlation coefficient of the NIR band was higher than the respective coefficients of the green and red bands. Considering the cultivation land in the study area was large and the main crops were cotton, winter wheat, and corn (which are in a growth period between June and July), the reflectance features of land cover types have significant changes. In addition, due to the temporal difference between the predicted time 21 August and the initial time 9 August, the land cover change was relatively small, the correlation was greater on 21 August than on 10 July, and the predicted result was better.

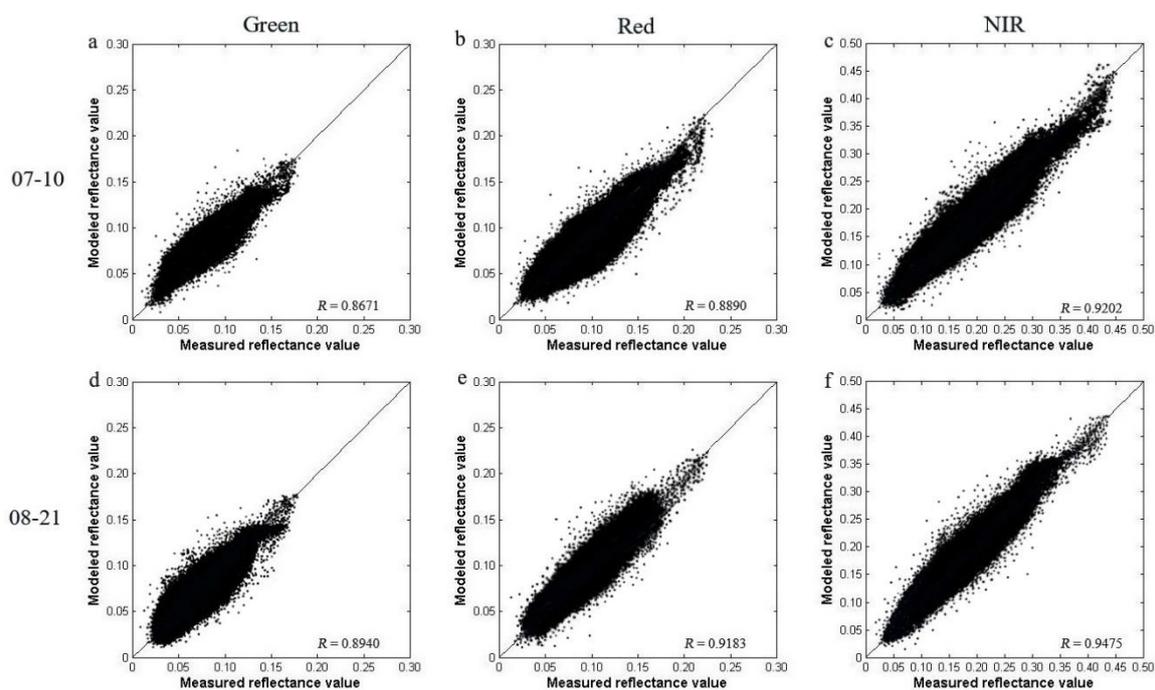


Figure 5. The scatter plots of the reflectance of the predicted values and the true values. (a–c) are the scatter plots of the green, red, and NIR bands on 07-10; (d–f) are the scatter plots of the green, red, and NIR bands on 08-21.

4.2.1. Qualitative Analysis of the Fusion Imagery

In order to discuss the validity and reliability of the combining linear pixel unmixing and STARFM fusion model for high spatiotemporal resolution image prediction, the predicted results were compared with the fusion images obtained by the STARFM algorithm, and the accuracy of the fusion results were evaluated by visual interpretation and correlation analysis. With the proposed method, high spatiotemporal resolution images were obtained for the prediction time and recorded as experiment 1 (Expt.1). The same parameters were set for the experimental data to perform image fusion using the original STARFM algorithm, recorded as experiment 2 (Expt.2).

The predicted green, red, and NIR bands using two fusion methods were composited in false color and the predicted images for 10 July and 21 August were obtained. Figure 6 shows the real MODIS image on 10 July 2016 and 21 August 2016, the real image of GF1 WFV at the same corresponding times, and the predicted images obtained by the Expt.1 and Expt.2. First, comparing and analyzing

the predicted images and the real images by visual interpretation, it can be found that the similarity between them was high in Expt.1 and Expt.2. The land cover types with large areas (water, residential land, and vast wetlands) were clearly identified, but there were still some differences, especially in parts of cultivated land on 10 July. The STARFM algorithm assumes that the reflectance difference between the high spatial resolution image and low spatial resolution image does not change over time. However, June and July were in the period of main crop harvest in the study area, when the land cover conditions undergo rapid change. As a result, the reflectance of land cover types changes, and therefore there is a certain difference in reflectance in parts of the cultivated land when comparing the fusion image and the real image. Compared with the predicted results of Expt.2, it can be seen that detailed spatial information was better represented when using the combining linear pixel unmixing and STARFM fusion model, especially for water boundaries and areas of cultivation. In addition, because Expt.1 input the MODIS unmixed image rather than the MODIS resampled image, the predicted image had a closer tone to the real image after composition in false color.

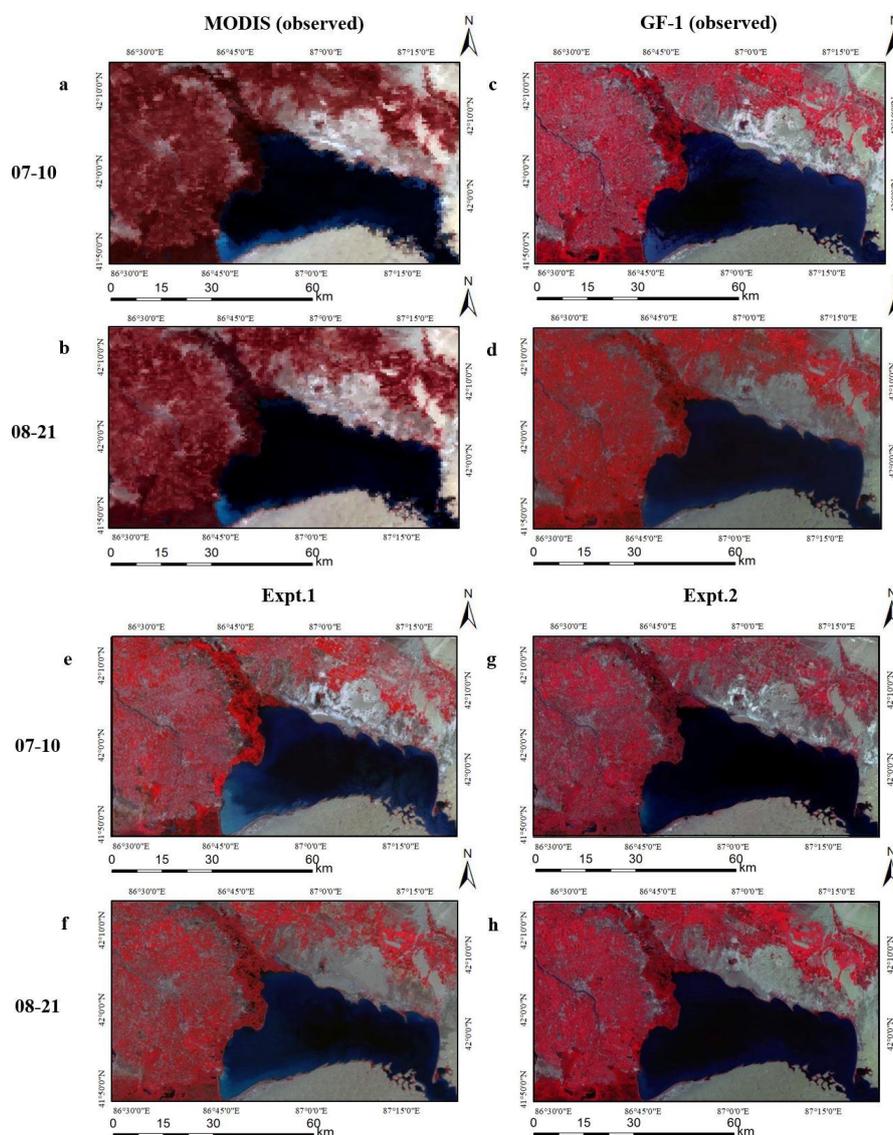


Figure 6. The real and predicted images (NIR–red–green). (a,c,e,g) show the MODIS true image, the GF-1 WFV true image, the predicted image of Expt.1, and the predicted image of Expt.2, respectively, on 10 July; (b,d,f,h) show the MODIS true image, the GF-1 WFV true image, the predicted image of Expt.1, and the predicted image of Expt.2, respectively, on 21 August.

A section of a typical area in the study area was selected to analyze the details of the predicted image on 10 July, as shown in Figure 7 with a black mark. The Expt.1 could identify small distribution information of ground objects and their texture, but the fusion image of the Expt.2 had an obvious patch phenomenon and could not predict accurately in heterogeneous areas. Since Expt.2 used MODIS resampled pixels as input to predict the image, sub pixels in the same MODIS pixel had strong homogeneity, while in the heterogeneous areas, the differences between the MODIS pixels were large. The model proposed in this paper can effectively solve the above problem. Because the mixed pixels in the image interfere with fusion accuracy, introducing the spectral mixture model into the spatiotemporal fusion technology can greatly improve the accuracy of the prediction result. For the prediction of changed areas (masked with a yellow box in Figure 7), prediction was not performed accurate in Expt.1 or Expt.2. Due to the hypothesis of the STARFM, when land cover types change, there is a larger difference between the fusion image and the real image, and the prediction accuracy of complex areas over long time intervals is lower.

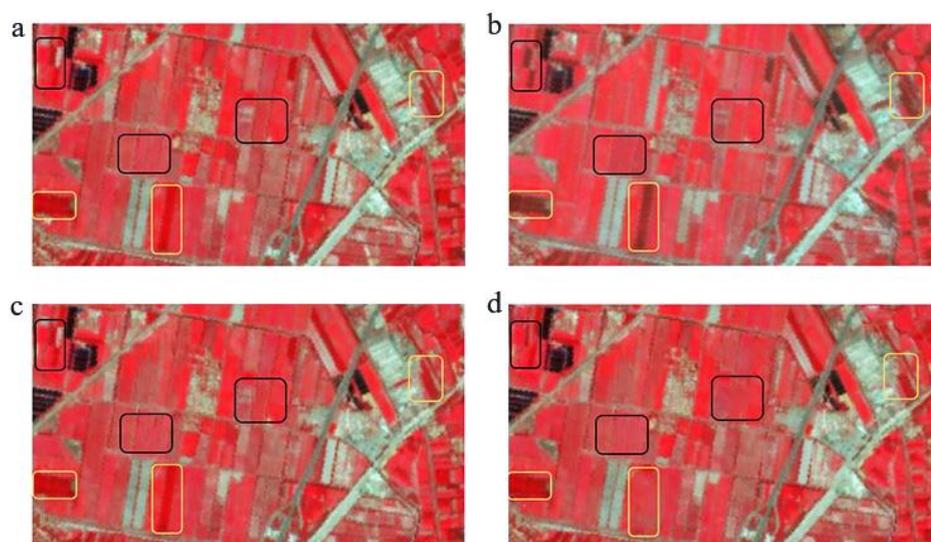


Figure 7. The local area comparison of the predicted images and the real images. (a) shows the real image on 12 June, (b) shows the real image on 10 July, (c) shows the predicted image of Expt.1 on 10 July, and (d) shows the predicted image of Expt.2 on 10 July.

4.2.2. Quantitative Analysis of the Fusion Imagery

In order to quantitatively analyze the predicted results of the two experiments, this paper used R and RMSE index to measure the similarity between the predicted image and the real image. The statistical results of R and RMSE in different bands in Expt.1 and Expt.2 are shown in Table 5. It can be seen that the correlation between the predicted value and the real value of Expt.1 is better than that of Expt.2. For the predicted time of 21 August, the correlation of each band improved to different degrees. In particular, the R of NIR band was the highest, reaching 0.9475. The RMSE of Expt.1 is also lower than that of Expt.2, which indicates that the predicted information of land cover is closer to the real images.

Table 5. Accuracy assessment of Expt.1 and Expt.2.

Date	Band	Expt.1		Expt.2	
		R	RMSE	R	RMSE
07–10	Green	0.8671	0.0317	0.8337	0.0411
	Red	0.8890	0.0374	0.8663	0.0382
	NIR	0.9202	0.0424	0.8932	0.0436
08–21	Green	0.8940	0.0338	0.8675	0.0387
	Red	0.9183	0.0227	0.8841	0.0359
	NIR	0.9475	0.0390	0.9153	0.0401

4.3. The Comparison of Estimation Accuracy of Reflectance in Homogeneous/Heterogeneous Areas

Taking homogeneous/heterogeneous areas as examples, the estimation accuracy of the two land cover types on 10 July was analyzed. The original image was cropped, and four areas of different land cover types were obtained (Figure 8). By using the method of Expt.1 and Expt.2, we performed prediction for images from 10 July of areas a, b, c, d, and used the mean absolute difference (MAD) to reflect the actual situation of error between the predicted values and the real values (Table 6), and to quantitatively evaluate the prediction accuracy of the combining linear pixel unmixing and STARFM fusion model in homogeneous/heterogeneous areas.

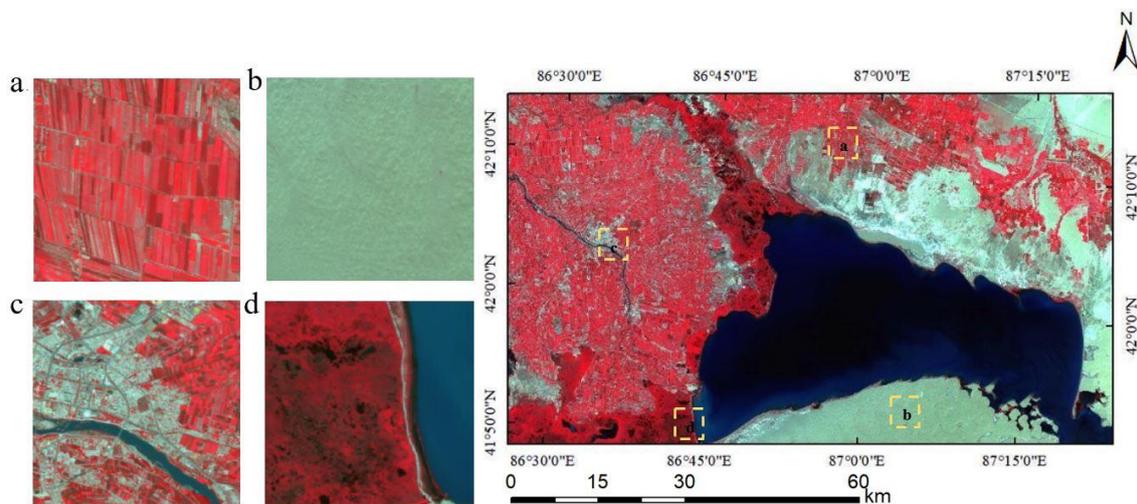


Figure 8. Selection of homogeneous/heterogeneous areas. Homogeneous areas: (a) mainly distributes cultivated land and (b) mainly distributes bare land. Heterogeneous areas: (c) mainly distributes residential land, cultivated land, and water; and (d) mainly distributes wetland and water.

Table 6. The estimation accuracy of reflectance in homogeneous/heterogeneous areas.

Experiment	Band	a	b	c	d
Expt.1	Green	0.0092	0.0086	0.0114	0.0097
	Red	0.0062	0.0047	0.0121	0.0083
	NIR	0.0049	0.0053	0.0092	0.0075
Expt.2	Green	0.0099	0.0091	0.0168	0.0132
	Red	0.0075	0.0060	0.0147	0.0130
	NIR	0.0058	0.0059	0.0122	0.0099

Comparing the prediction results in homogeneous/heterogeneous areas of Expt.1 and Expt.2, it can be found that the MAD decreases to some extent, and the fusion result is closer to the true

reflectance than the result of the Expt.2. In the homogeneous areas, the fusion images show little differences between Expt.1 and Expt.2. For area a, the MAD of green, red, and NIR bands of Expt.1 are reduced by 0.0007, 0.0013, and 0.0009, respectively, as compared to Expt.2. For area b, the MAD in the green band increases by 0.0005, and in the red and NIR bands, decreases by 0.0013 and 0.0009, respectively. However, in heterogeneous areas where the local spatial autocorrelation is low, the extent of the decrease of the average absolute differences in Expt.1 is much larger than Expt.2. For area c, the MAD in the green, red, and NIR bands of Expt.1 shows reductions of 0.0054, 0.0026, and 0.0030, respectively, compared to those of Expt.2, and for area d, reductions of 0.0035, 0.0047, and 0.0024. Therefore, in heterogeneous areas, using unmixed reflectance as the input data, the STARFM algorithm can effectively improve the estimation accuracy of the fusion image and increase the probability of searching for pure pixels of the STARFM algorithm to reduce the intra-class errors. In addition, the MAD between the prediction results and the real images in areas c and d is all larger than that of areas a and b, which shows that MODIS pure pixels are still an important factor for controlling the prediction accuracy.

4.4. NDVI Estimation and Comparative Analysis

NDVI is an important index for reflecting plant growth status and distribution, which is based on the strong absorption characteristics of the plant foliage in the red band and the strong reflection characteristics in the NIR band to combine different bands of remote sensing satellites [41]. To evaluate the potential of the proposed fusion algorithm in remote sensing applications, this study utilized the NDVI to further explore the validity and reliability of this fusion method. The NDVI images of 10 July and 21 August were calculated based on the predicted images of the red band and the NIR band. The comparison between the predicted NDVI images at two times and the real MODIS NDVI images at the same time is shown in Figure 9, and the differences between the predicted NDVI value and that of the real GF-1 WFV in the same period are shown in Figure 10. As seen in Figure 9, the predicted NDVI of Expt.1 is more consistent with MODIS NDVI in the same period than that of Expt.2 and can reflect the spatial detailed distribution difference of NDVI at GF-1 WFV resolution (16-m). By visualizing Figure 10 and the histogram of difference map (Figure 11), it can be seen that the prediction of NDVI is still different from that of GF-1 WFV NDVI in the same period. However, the range distribution of Expt.1 is more concentrated, and the difference with the actual GF-1 WFV NDVI is even smaller.

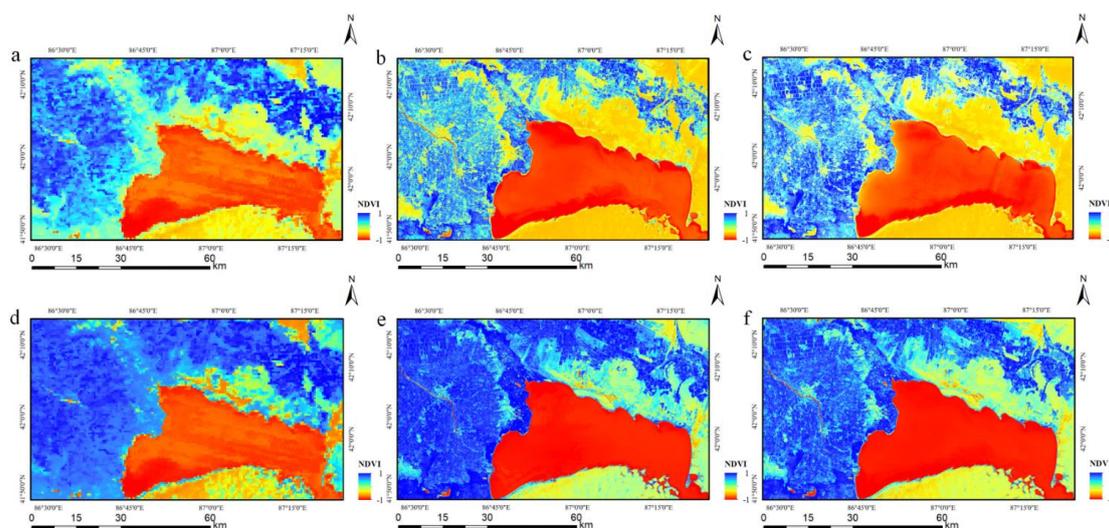


Figure 9. The observed NDVI and predicted NDVI. (a–c) show the MODIS NDVI, the predicted NDVI of Expt.1, and the predicted NDVI of Expt.2, on 10 July and (d–f) show the MODIS NDVI, the predicted NDVI of Expt.1, and the predicted NDVI of Expt.2, on 21 August.

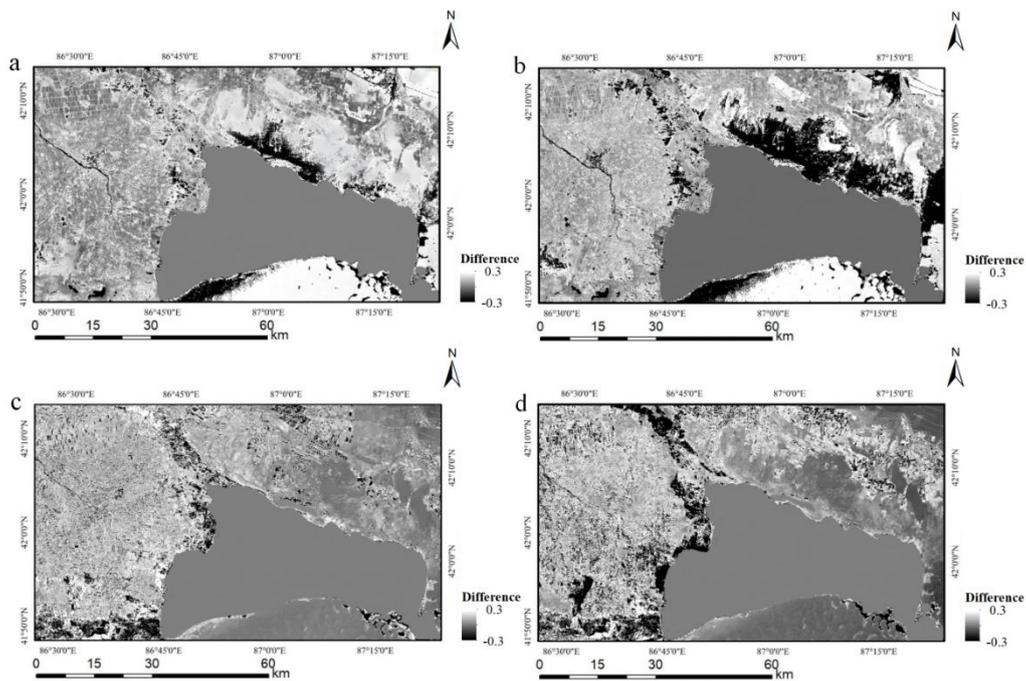


Figure 10. The difference map of the predicted NDVI and real GF-1 WFV NDVI at the same period. (a,b) were obtained by Expt.1 and Expt.2 on 10 July, respectively, and (c,d) were obtained by Expt.1 and Expt.2 on 21 August, respectively.

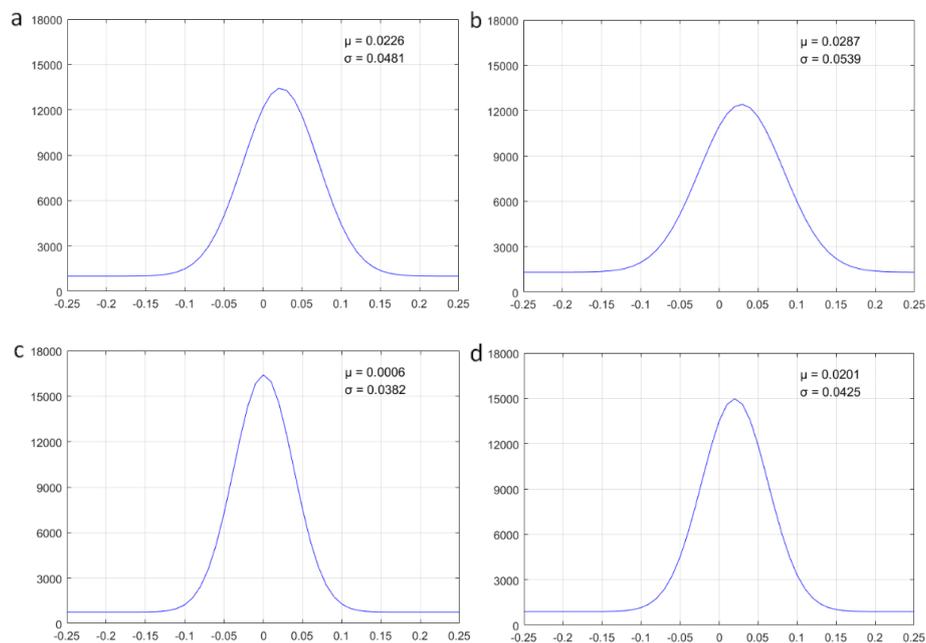


Figure 11. Difference image histogram of the difference map. (a,b) were obtained by Expt.1 and Expt.2 on 10 July, respectively, and (c,d) were obtained by Expt.1 and Expt.2 on 21 August, respectively.

Table 7 shows the R and RMSE of the predicted NDVI and real GF-1 WFV NDVI, and the mean value, standard deviation, and the percentage of the statistics of pixel absolute value less than 0.1 and 0.2 of the difference images. Through the comparison of R and RMSE, it can be seen that the proposed method achieved a higher accuracy than STARFM algorithm in the prediction of NDVI,

and the linear fitting degree with real NDVI value is better. In addition, it can be found that the NDVI prediction result on 21 August is better than that on 10 July. The possible reason is that the accuracy of the predicted red band and NIR band on 21 August is higher, which then indirectly affects the prediction accuracy of NDVI. By analyzing the percentage statistics of pixels, the proportion of pixel absolute value less than 0.1 is generally over 90% in Expt.1, and the proportion less than 0.2 is higher than Expt.2 to some extent. Although the predicted NDVI still has some differences from GF-1 WFV NDVI in the same period, it can be considered that the proposed method can effectively reconstruct high-precision NDVI with high spatiotemporal resolution.

Table 7. NDVI accuracy evaluation.

Index	07-10		08-21	
	Expt.1	Expt.2	Expt.1	Expt.2
R	0.9120	0.8729	0.9437	0.9184
RMSE	0.0323	0.0458	0.0264	0.0307
μ ¹	0.0226	0.0287	0.0006	0.0201
σ ²	0.0481	0.0539	0.0382	0.0425
$p^3 < 0.1$	89.74%	86.24%	94.71%	90.26%
$p^3 < 0.2$	95.48%	90.73%	98.61%	93.80%

¹ is the mean value, ² is the standard deviation, and ³ is the pixel absolute value.

5. Discussion

For many applications, the spatiotemporal fusion technology of remote sensing data must meet growing needs in terms of spatial and time dimensions. The purpose of fusion is to obtain high-resolution data when information is unavailable to make up for the lack of data, and to obtain continuous and valid data that can be used in quantitative retrieval. To date, most spatiotemporal fusion models have focused mainly on the fusion of Landsat and MODIS data, and few studies have researched and compared other satellites in depth, although the STARFM is theoretically also effective for other satellites. Therefore, in this paper, on the basis of previous studies, the spatiotemporal fusion model was applied to the high-resolution satellite GF-1 and the high temporal resolution satellite MODIS in order to expand the range of applications of the spatiotemporal fusion model. Although the STARFM can obtain relatively good fusion results, the phenomenon of mixed pixels is common because of the limitations of the algorithm assumption and the large field of MODIS view. As a result, some patchy effects appear, and prediction in highly heterogeneous areas is not sensitive. In order to solve these problems, the combining linear pixel unmixing and STARFM fusion model was proposed in this paper. The MODIS data were downsampled using a linear spectral mixture model and then replaced the resampled MODIS data in order to perform fusion. Although the proposed method has achieved good results in general, the fusion results may still be affected by the following factors:

- (1) The differences in multi-source remote sensing data. Due to different measurement principles and observation environments, different sensors have different spectral settings and spectral response functions. In addition, errors such as radiation calibration, atmospheric correction, and geometric correction are unavoidable in data processing. These are nonlinear factors that cause systematic errors within sensor data.
- (2) The change in land cover types. The date of the prediction imagery and the reference imagery of the STARFM algorithm should be within a season. If the time interval is too long, land cover changes become more likely, and the corresponding relationship between the band reflectance data before and after fusion is not maintained well, resulting in a decrease in the precision of prediction results. Therefore, the selection of reference images should be taken into consideration in fusion.

- (3) The variability of reflectance of pixels. The linear spectral mixture model is based on the assumption that the same land cover types have the same spectral features, and that there is a linear summation of the spectra. In fact, the spectral reflectance of each land cover type is combined in nonlinear forms and has variability. In addition, due to randomness and variability, natural, random residuals are unavoidable. All these factors may affect the accuracy of the model and the results of spectral mixture analysis.

6. Conclusions

Considering that the estimation accuracy of the STARFM algorithm is low in areas of high heterogeneity, this paper proposed a new fusion model. Based on the STARFM, a linear spectral mixture model was introduced to unmix the input MODIS data so that it could combine the advantages of the spectral mixture model and the STARFM, and Gf-1 WFV data and MODIS data were used to test the proposed method. Some following conclusions were drawn:

- (1) This proposed method can generate high-precision predicted images with high spatiotemporal resolution. The experimental results showed that the improved method can increase the probability of searching for pure pixels in moving windows. The scatter plots of each band showed that the correlation coefficient between the predicted image and the real image was better than STARFM, especially for NIR band.

- (2) The combining linear pixel unmixing and STARFM fusion model has resulted in better spatial details, especially in water boundaries and cultivated areas, and the predicted image had a closer tone to the real image after composition in false color. However, a big difference was found in the areas with changed land cover types.

- (3) The proposed method was successfully applied in highly heterogenous areas, and it was able to identify small objects and their texture distribution information. However, for the fusion image of the STARFM algorithm, an obvious patchy phenomenon was found in heterogeneous areas and the fusion accuracy was relatively low.

- (4) With the resultant fusion imagery, 16-m high-precision NDVI data can be successfully generated by the predicted red band and the NIR band, which contained not only spatial difference information of high spatial resolution data but also the time advantage of high temporal resolution data. The application further verified the accuracy and adaptability of the fusion data obtained from the proposed algorithm in retrieving land surface parameters.

Due to the many complex problems of the spatiotemporal fusion method, future research will study practical applications and fully consider the complex problems that arise, including in extreme cases. It will also be necessary to investigate influencing factors in order to set parameters reasonably, as well as effective preprocessing in order to improve the accuracy of fusion and broaden the scope of application. Moreover, we usually have much prior knowledge of the study area. Discovering how to make full use of this prior knowledge and known ground information for integration into the final prediction result will be an important breakthrough in future research.

Author Contributions: C.J. and Z.X. are the directors of the corresponding contents. C.J. and Z.X. designed the algorithm and performed the experiments. L.M. completed the data processing and manuscript polishing.

Funding: The study is funded by the National Key R&D Programs of China (No. 2017YFB0504201), the National Natural Science Foundation of China (No. 61473286).

Acknowledgments: The authors would like to thank the anonymous reviewers for their constructive comments and suggestions.

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

References

1. Coops, N.C.; Johnson, M.; Wulder, M.A.; White, J.C. Assessment of QuickBird high spatial resolution imagery to detect red attack damage due to mountain pine beetle infestation. *Remote Sens. Environ.* **2006**, *103*, 67–80. [[CrossRef](#)]
2. Tian, F.; Wang, Y.J.; Fensholt, R.; Wang, K.; Zhang, L.; Huang, Y. Mapping and Evaluation of NDVI Trends from Synthetic Time Series Obtained by Blending Landsat and MODIS Data around a Coalfield on the Loess Plateau. *Remote Sens.* **2013**, *5*, 4255–4279. [[CrossRef](#)]
3. Jarihani, A.A.; McVicar, T.R.; Van Niel, T.G.; Emelyanova, I.V.; Callow, J.N.; Johansen, K. Blending Landsat and MODIS Data to Generate Multispectral Indices: A Comparison of “Index-then-Blend” and “Blend-then-Index” Approaches. *Remote Sens.* **2014**, *6*, 9213–9238. [[CrossRef](#)]
4. Tewes, A.; Thonfeld, F.; Schmidt, M.; Oomen, R.J.; Zhu, X.L.; Dubovyk, O.; Menz, G.; Schellberg, J. Using RapidEye and MODIS Data Fusion to Monitor Vegetation Dynamics in Semi-Arid Rangelands in South Africa. *Remote Sens.* **2015**, *7*, 6510–6534. [[CrossRef](#)]
5. Rao, Y.H.; Zhu, X.L.; Chen, J.; Wang, J.M. An Improved Method for Producing High Spatial-Resolution NDVI Time Series Datasets with Multi-Temporal MODIS NDVI Data and Landsat TM/ETM plus Images. *Remote Sens.* **2015**, *7*, 7865–7891. [[CrossRef](#)]
6. Schmidt, M.; Udelhoven, T.; Gill, T.; Roder, A. Long term data fusion for a dense time series analysis with MODIS and Landsat imagery in an Australian Savanna. *J. Appl. Remote Sens.* **2012**, *6*, 063512.
7. Zhang, W.; Li, A.N.; Jin, H.A.; Bian, J.H.; Zhang, Z.J.; Lei, G.B.; Qin, Z.H.; Huang, C.Q. An Enhanced Spatial and Temporal Data Fusion Model for Fusing Landsat and MODIS Surface Reflectance to Generate High Temporal Landsat-Like Data. *Remote Sens.* **2013**, *5*, 5346–5368. [[CrossRef](#)]
8. Bai, Y.; Wong, M.S.; Shi, W.Z.; Wu, L.X.; Qin, K. Advancing of Land Surface Temperature Retrieval Using Extreme Learning Machine and Spatio-Temporal Adaptive Data Fusion Algorithm. *Remote Sens.* **2015**, *7*, 4424–4441. [[CrossRef](#)]
9. Singh, D. Generation and evaluation of gross primary productivity using Landsat data through blending with MODIS data. *Int. J. Appl. Earth Obs. Geoinf.* **2011**, *13*, 59–69. [[CrossRef](#)]
10. Weng, Q.H.; 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](#)]
11. Chen, B.; Ge, Q.; Fu, D.; Yu, G.; Sun, X.; Wang, S.; Wang, H. A data-model fusion approach for upscaling gross ecosystem productivity to the landscape scale based on remote sensing and flux footprint modelling. *Biogeosciences* **2010**, *7*, 2943–2958. [[CrossRef](#)]
12. Liu, K.; Su, H.B.; Li, X.K.; Wang, W.M.; Yang, L.J.; Liang, H. Quantifying Spatial-Temporal Pattern of Urban Heat Island in Beijing: An Improved Assessment Using Land Surface Temperature (LST) Time Series Observations From LANDSAT, MODIS, and Chinese New Satellite GaoFen-1. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 2028–2042. [[CrossRef](#)]
13. Chao, Y.F.; Zhang, H.T.; Dan, G.; Hao, W. Wetland mapping and wetland temporal dynamic analysis in the Nanjishan wetland using Gaofen One data. *Ann. GIS* **2016**, *22*, 259–271.
14. Wu, M.Q.; Huang, W.J.; Niu, Z.; Wang, C.Y. Combining HJ CCD, GF-1 WFV and MODIS Data to Generate Daily High Spatial Resolution Synthetic Data for Environmental Process Monitoring. *Int. J. Environ. Res. Public Health* **2015**, *12*, 9920–9937. [[CrossRef](#)] [[PubMed](#)]
15. Fu, B.L.; Wang, Y.Q.; Campbell, A.; Li, Y.; Zhang, B.; Yin, S.B.; Xing, Z.F.; Jin, X.M. Comparison of object-based and pixel-based Random Forest algorithm for wetland vegetation mapping using high spatial resolution GF-1 and SAR data. *Ecol. Indic.* **2017**, *73*, 105–117. [[CrossRef](#)]
16. Li, J.; Chen, X.L.; Tian, L.Q.; Huang, J.; Feng, L. Improved capabilities of the Chinese high-resolution remote sensing satellite GF-1 for monitoring suspended particulate matter (SPM) in inland waters: Radiometric and spatial considerations. *ISPRS J. Photogramm. Remote Sens.* **2015**, *106*, 145–156. [[CrossRef](#)]
17. Song, Q.; Hu, Q.; Zhou, Q.B.; Hovis, C.; Xiang, M.T.; Tang, H.J.; Wu, W.B. In-Season Crop Mapping with GF-1/WFV Data by Combining Object-Based Image Analysis and Random Forest. *Remote Sens.* **2017**, *9*, 1184. [[CrossRef](#)]
18. Marchane, A.; Jarlan, L.; Hanich, L.; Boudhar, A.; Gascoin, S.; Tavernier, A.; Filali, N.; Le Page, M.; Hagolle, O.; Berjamy, B. Assessment of daily MODIS snow cover products to monitor snow cover dynamics over the Moroccan Atlas mountain range. *Remote Sens. Environ.* **2015**, *160*, 72–86. [[CrossRef](#)]

19. Chen, Y.L.; Song, X.D.; Wang, S.S.; Huang, J.F.; Mansaray, L.R. Impacts of spatial heterogeneity on crop area mapping in Canada using MODIS data. *ISPRS J. Photogramm. Remote Sens.* **2016**, *119*, 451–461. [[CrossRef](#)]
20. Portillo-Quintero, C.A.; Sanchez, A.M.; Valbuena, C.A.; Gonzalez, Y.Y.; Larreal, J.T. Forest cover and deforestation patterns in the Northern Andes (Lake Maracaibo Basin): A synoptic assessment using MODIS and Landsat imagery. *Appl. Geogr.* **2012**, *35*, 152–163. [[CrossRef](#)]
21. Wu, M.Q.; Wu, C.Y.; Huang, W.J.; Niu, Z.; Wang, C.Y.; Li, W.; Hao, P.Y. An improved high spatial and temporal data fusion approach for combining Landsat and MODIS data to generate daily synthetic Landsat imagery. *Inf. Fusion* **2016**, *31*, 14–25. [[CrossRef](#)]
22. Zhukov, B.; Oertel, D.; Lanzl, F.; Reinhackel, G. Unmixing-based multisensor multiresolution image fusion. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 1212–1226. [[CrossRef](#)]
23. Maselli, F. Definition of spatially variable spectral endmembers by locally calibrated multivariate regression analyses. *Remote Sens. Environ.* **2001**, *75*, 29–38. [[CrossRef](#)]
24. Busetto, L.; Meroni, M.; Colombo, R. Combining medium and coarse spatial resolution satellite data to improve the estimation of sub-pixel NDVI time series. *Remote Sens. Environ.* **2008**, *112*, 118–131. [[CrossRef](#)]
25. Cherchali, S.; Amram, O.; Flouzat, G. Retrieval of temporal profiles of reflectances from simulated and real NOAA-AVHRR data over heterogeneous landscapes. *Int. J. Remote Sens.* **2000**, *21*, 753–775. [[CrossRef](#)]
26. 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.
27. Walker, J.J.; de Beurs, K.M.; Wynne, R.H.; Gao, F. Evaluation of Landsat and MODIS data fusion products for analysis of dryland forest phenology. *Remote Sens. Environ.* **2012**, *117*, 381–393. [[CrossRef](#)]
28. Swain, R.; Sahoo, B. Mapping of heavy metal pollution in river water at daily time-scale using spatio-temporal fusion of MODIS-aqua and Landsat satellite imageries. *J. Environ. Manag.* **2017**, *192*, 1–14. [[CrossRef](#)] [[PubMed](#)]
29. Walker, J.J.; de Beurs, K.M.; Wynne, R.H. Dryland vegetation phenology across an elevation gradient in Arizona, USA, investigated with fused MODIS and Landsat data. *Remote Sens. Environ.* **2014**, *144*, 85–97. [[CrossRef](#)]
30. Olsoy, P.J.; Mitchell, J.; Glenn, N.F.; Flores, A.N. Assessing a Multi-Platform Data Fusion Technique in Capturing Spatiotemporal Dynamics of Heterogeneous Dryland Ecosystems in Topographically Complex Terrain. *Remote Sens.* **2017**, *9*, 981. [[CrossRef](#)]
31. Hilker, T.; Wulder, M.A.; Coops, N.C.; Seitz, N.; White, J.C.; Gao, F.; Masek, J.G.; Stenhouse, G. Generation of dense time series synthetic Landsat data through data blending with MODIS using a spatial and temporal adaptive reflectance fusion model. *Remote Sens. Environ.* **2009**, *113*, 1988–1999. [[CrossRef](#)]
32. 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](#)]
33. Roy, D.P.; Ju, J.; Lewis, P.; Schaaf, C.; Gao, F.; Hansen, M.; Lindquist, E. Multi-temporal MODIS-Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data. *Remote Sens. Environ.* **2008**, *112*, 3112–3130. [[CrossRef](#)]
34. Zhu, X.L.; Chen, J.; Gao, F.; Chen, X.H.; Masek, J.G. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sens. Environ.* **2010**, *114*, 2610–2623. [[CrossRef](#)]
35. Yang, J.C.; Wright, J.; Huang, T.; Ma, Y. Image super-resolution via sparse representation. *IEEE Trans. Image Process.* **2010**, *19*, 2861–2873. [[CrossRef](#)] [[PubMed](#)]
36. Zeyde, R.; Elad, M.; Protter, M. On Single Image Scale-Up Using Sparse Representations. In Proceedings of the 7th International Conference on Curves and Surfaces, Avignon, France, 24–30 June 2012; pp. 711–730.
37. Song, H.H.; Huang, B. Spatiotemporal Satellite Image Fusion through One-Pair Image Learning. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 1883–1896. [[CrossRef](#)]
38. Vermote, E.F.; Tanre, D.; Deuze, J.L.; Herman, M.; Morcrette, J.J. Second Simulation of the Satellite Signal in the Solar Spectrum, 6S: An overview. *IEEE Trans. Geosci. Remote Sens.* **1997**, *35*, 675–686. [[CrossRef](#)]
39. Ichoku, C.; Karnieli, A. A review of mixture modeling techniques for sub-pixel and cover estimation. *Remote Sens. Rev.* **1996**, *13*, 161–186. [[CrossRef](#)]

40. Dunn, J.C. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *J. Cybern.* **1973**, *3*, 32–57. [[CrossRef](#)]
41. Lumbierres, M.; Mendez, P.F.; Bustamante, J.; Soriguer, R.; Santamaria, L. Modeling Biomass Production in Seasonal Wetlands Using MODIS NDVI Land Surface Phenology. *Remote Sens.* **2017**, *9*, 392. [[CrossRef](#)]



© 2018 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/>).