Article

A Framework of Spatio-Temporal Fusion Algorithm Selection for Landsat NDVI Time Series Construction

Yangnan Guo [1], Cangjiao Wang [2,*], Shaogang Lei [2], Junzhe Yang [1] and Yibo Zhao [2]

[1]   China Energy Group Shendong Coal Group Technology Research Institute, Ordos 017209, China;
10031311@chnenergy.com.cn (Y.G.); 10018115@chnenergy.com.cn (J.Y.)

[2]   Engineering Research Center of Ministry of Education for Mine Ecological Restoration, China University of
Mining and Technology, Xuzhou 221116, China; lsgang@cumt.edu.cn (S.L.); zhaoyibo@cumt.edu.cn (Y.Z.)

*   Correspondence: wangcangjiao@cumt.edu.cn

check for
updates

Abstract: Spatio-temporal fusion algorithms dramatically enhance the application of the Landsat time series. However, each spatio-temporal fusion algorithm has its pros and cons of heterogeneous land cover performance, the minimal number of input image pairs, and its efficiency. This study aimed to answer: (1) how to determine the adaptability of the spatio-temporal fusion algorithm for predicting images in prediction date and (2) whether the Landsat normalized difference vegetation index (NDVI) time series would benefit from the interpolation with images fused from multiple spatio-temporal fusion algorithms. Thus, we supposed a linear relationship existed between the fusion accuracy and spatial and temporal variance. Taking the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) and the Enhanced STARFM (ESTARFM) as basic algorithms, a framework was designed to screen a spatio-temporal fusion algorithm for the Landsat NDVI time series construction. The screening rule was designed by fitting the linear relationship between the spatial and temporal variance and fusion algorithm accuracy, and then the fitted relationship was combined with the graded accuracy selecting rule (R^2) to select the fusion algorithm. The results indicated that the constructed Landsat NDVI time series by this paper proposed framework exhibited the highest overall accuracy (88.18%), and lowest omission (1.82%) and commission errors (10.00%) in land cover change detection compared with the moderate resolution imaging spectroradiometer (MODIS) NDVI time series and the NDVI time series constructed by a single STARFM or ESTARFM. Phenological stability analysis demonstrated that the Landsat NDVI time series established by multiple spatio-temporal algorithms could effectively avoid phenological fluctuations in the time series constructed by a single fusion algorithm. We believe that this framework can help improve the quality of the Landsat NDVI time series and fulfill the gap between near real-time environmental monitoring mandates and data-scarcity reality.

Keywords: spatiotemporal fusion algorithm; Landsat time series; screening scheme

1. Introduction

Landsat images recording the Earth's surface status since 1972 are irreplaceable in terrestrial ecosystem dynamics monitoring and biosphere processes modeling [1,2]. The Landsat normalized difference vegetation index (NDVI) acts on behalf of the land surface's greenness. Due to its high spatial resolution (30 m), Landsat NDVI time series has obvious superiority to the NDVI time series of the Moderate Resolution Imaging Spectroradiometer (MODIS), the SPOT Vegetation, and the Global Inventory Modeling and Mapping Studies (GIMMS) with a spatial resolution of 250 m, 1 km, and 8 km, respectively. The longer historical image archives of Landsat defeat the latest dual high-resolution (5-day temporal resolution and 10 m spatial resolution) Sentinel datasets in meeting the requirement of

long-term studies. As a result, the Landsat NDVI time series plays an indispensable role in classification accuracy improvement, phenological measurement, and long-term land cover change monitoring [3–6]. However, limited by the 16-day revisit time, frequent cloud contamination, and 22% data loss of the Enhanced Thematic Mapper Plus (ETM+) sensor since 2003, there is a dearth of dual high resolution (spatial and temporal resolution) Landsat NDVI time series [7–10]. These missing observation data have caused the Landsat time series to fail to catch land cover change events or to extract important phenological nodes [5,11]. Thus, missing image reconstruction is expected to solve the mentioned problems [12–14].

Spatio-temporal essentially focuses on downscaling spatially coarse images to spatially fine images. It aims to produce high temporal and high spatial resolution images for any missing data by taking full advantage of the complementary information of images with a frequently observed but coarse spatial resolution (referring to as coarse images) and the images with an infrequently observed but fine spatial resolution (referring to as fine images). Unlike interpretation methods that depend on the input number of homologous and available Landsat reference images, spatio-temporal fusion is flexible in input reference images and even performs well when just using one neighboring observed image [15,16]. Furthermore, the introduction of temporal information from coarse images further enhances the dynamic depiction of land cover. Currently, more than 50 spatio-temporal fusion algorithms have been proposed. Technically, these algorithms were generally categorized into five groups: the weight function-based, unmixing-based, learning-based, Bayesian-based, and hybrid fusion methods. The weight function-based methods, based on the weight function to integrate the spectral, temporal, and spatial contribution of reference images, are the most broadly used methods [8,17–19]. The unmixing-based methods unmix the coarse image into a fine-scale image based on linear spectral mixing theory [12,20,21]. Learning-based methods have been recently developed, but are promising, which predict fine images by the relationship between the coarse images and fine images learned by machine learning or deep learning technology [22–24]. The Bayesian-based methods use Bayes theory to estimate the fine image [25]. The hybrid fusion methods integrate two or more theories of the above-mentioned methods to realize image fusion [13].

Although spatio-temporal fusion methods have been booming, methods for directly producing Landsat NDVI time series based on all available Landsat images and coarse image time series are still in shortage [13,16,26]. Presently, most of the spatio-temporal fusion-based Landsat NDVI time series for land surface monitoring majorly utilize the strategy of filling the missing data by a single image fusion method [11,27]. However, different fusion algorithms have different pros and cons in image fusion. For example, the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) is feasible and efficient for image fusion as it only needs one pair of coarse and fine reference images. But it has limited performance in heterogonous land cover area [17]. While the Enhanced STARFM (ESTARFM) performs better in heterogonous land cover areas, it needs at least two pairs of coarse and fine reference images and is time-consuming [8]. Thus, the following decisions should be clear when fusing a single missing image for constructing the Landsat NDVI time series: (1) which fusion algorithm should be chosen for blending the missing image; and (2) whether the performance of the Landsat NDVI time series would be improved by employing fusion images generated from multiple fusion algorithms.

Therefore, the spatio-temporal variance, which quantitated the spatial and temporal change between the dates of a reference image and missing image, was introduced to support the spatio-temporal fusion algorithm selection for Landsat NDVI time series construction. We supposed a linear relationship existed between the fusion accuracy and spatial and temporal variance. Then, a framework was proposed for Landsat NDVI time series construction with different fusion algorithms by taking the STARFM and the ESTARFM as base algorithms due to their wide application, their performance in heterogeneous image fusion, and the minimum required number of input image pairs. Subsequently, the effectiveness of the screen rule in the proposed framework was verified. Moreover, the performance of the multi-algorithm-based Landsat NDVI time series in land cover change

monitoring and phenological stability was demonstrated by comparing to the single-algorithm-based Landsat NDVI time series. The remainder of this study was organized as follows: Section 2 described the theoretical basis and the workflow of Landsat NDVI time series construction; Section 3 presented the experimentation and effectiveness analysis; Section 4 discussed the advantages, uncertainties, and perspectives; and Section 5 drew conclusions.

2. Methodology

The spatio-temporal fusion algorithm aims to generate high spatial and temporal resolution images, which takes the fine images and coarse images in the base date and coarse images in the prediction date as inputs to predict fine images in the prediction date (Figure 1) [8,9,28]. However, each fusion algorithm has its pros and cons under different fusion situations. Hence, it is reasonable to assume that the performance of various spatio-temporal fusion algorithms would be different under different spatial and temporal conditions. Thus, the spatial and temporal variance was introduced to quantify the spatial and temporal difference between coarse images in the base and prediction date. The screening rule was constructed by first fitting the linear relationship between the spatial and temporal variance and fusion algorithm accuracy, and then the fitted relationship was combined with the graded accuracy selecting rule to select the fusion algorithm. Thus, the screened algorithm for a specific prediction date was complemented to predict the missing Landsat image. Then, those predicted Landsat images were utilized to calculate the NDVI for constructing the Landsat NDVI time series. Furthermore, the screening rule's performance was verified by the effectiveness of algorithm selection. The multiple-algorithm-based Landsat NDVI time series was assessed by the effectiveness of land cover change detection, and phenological extraction. It is worth noting that the STARFM and ESTARFM were selected as the basic algorithm to test the proposed framework due to their complementary characteristics of performance in homogenous or heterogeneous land cover, one pair or two pairs reference image, and efficiency.
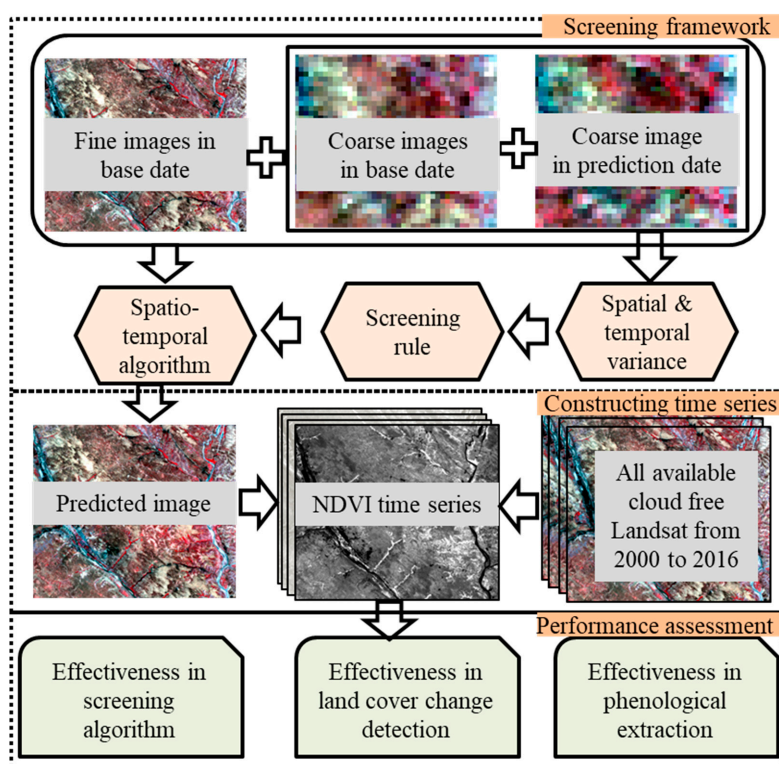

Figure 1. Framework for Landsat normalized difference vegetation index (NDVI) time series construction using the spatiotemporal fusion algorithm screening scheme.

2.1. Base Spatiotemporal Fusion Algorithms

2.1.1. The Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM)

STARFM can yield satisfactory fusion results in the stage of non-linear land cover change as well as in homogenous regions [17], which assumed that the residual errors between the coarse image and fine image were immutable and caused by land cover and land use changes, the wavelength characteristics between two sensors, and the illumination conditions. Thus, the difference between the fine image and coarse image in the base and prediction date should be equivalent. Given that neighboring similar pixels would have similar reflectance, the STARFM was modified by combining the weight function, where the weight function is integrated spatial distance, spectral difference, and temporal difference [17]. The implementation of the STARFM can be summarized as first preparing one pair of fine and coarse images in the base date and one coarse image in the prediction date, then finding the similar neighbor pixels within a moving window, and finally predicting the fine image in the prediction date. STARFM can be defined as follows:

$$L\left(x_{w/2}, y_{w/2}, t_p\right) = \sum_{i=1}^{w} \sum_{j=1}^{w} \sum_{k}^{n} W_{ijk} \times \left(M\left(x_i, y_j, t_p\right) + L\left(x_i, y_j, t_k\right) - M\left(x_i, y_j, t_k\right)\right) \tag{1}$$

where $x_{w/2}, y_{w/2}$ is the center location of the moving window, where the moving window (w) was set as a 31-pixel width. More parameter settings can refer to [12,17]. L and M are the reflectance of the fine image and coarse image, respectively. n is the number of reference images, hence k is from 1 to n. t_p and t_k is the date of the fusion and reference images, respectively. W_{ijk} refers to the weight coefficient determined by the spatial distance, spectral similarity, and time difference, and is as follows:

$$W_{ijk} = \left(1/\left(S_{ijk} \times T_{ijk} \times D_{ijk}\right)\right) / \sum_{i=1}^{w} \sum_{j=1}^{w} \sum_{k}^{n} \left(1/\left(S_{ijk} \times T_{ijk} \times D_{ijk}\right)\right) \tag{2}$$

where S, T, and D indicate the spectral similarity, time difference, and spatial distance between the fine and coarse images.

2.1.2. The Enhanced STARFM (ESTARFM)

Unlike STARFM, ESTARFM is based on at least two pairs of coarse and fine reference images, which assume a linear land cover change between these two reference dates [8]. Thus, the linear conversion coefficient (V) between reference images should be calculated to represent the spectral change slope [8]. Like the STARFM, the weight function is utilized to integrate the spectral difference, time difference, and spatial distance of each neighboring similar pixel. Although ESTARFM has been enhanced, its fusion accuracy is affected by reference image quality and quantity, and its linear hypothesis [29,30]. ESTARFM can be defined as follows:

$$L\left(x_{w/2}, y_{w/2}, t_p\right) = L(x_{w/2}, y_{w/2}, t_k) + \sum_{i=1}^{N} W_i \times V_i \times \left(M\left(x_i, y_j, t_p\right) - M\left(x_i, y_j, t_k\right)\right) \tag{3}$$

where L and M are the reflectance of fine image and coarse image; t is the time; t_p and t_k are the dates of the fusion and reference images, respectively. The N is the number of similar pixels in the moving windows, where the size of the moving window was set as 31-pixel width in this paper. More parameter settings refer to [8]. W is the weight coefficient.

2.2. Spatial and Temporal Variance

The spatial and temporal variance partitions the variance into space and time [31,32]. These two variances represent the spatial and temporal differences of land cover caused by phenological changes or human activities. Mean spatial variance emphasizes the degree of surface spatial heterogeneity, while mean temporal variance reflects the temporal dynamics of objects. Therefore, the spatial

and temporal variance provides a new way of qualifying the difference in fusion algorithm images. The partition of the time-first and space-first formulation can respectively be defined as follows:

$$\overline{\delta_s^2} = \frac{(m \times n) - 1}{m(n-1)} \times \delta_g^2 - \frac{n(m-1)}{m(n-1)} \times \delta_t^2(u) \tag{4}$$

$$\overline{\delta_t^2} = \frac{(m \times n) - 1}{n(m-1)} \times \delta_g^2 - \frac{m(n-1)}{n(m-1)} \times \delta_s^2(u) \tag{5}$$

where δ_g^2 is the spatial and temporal variance across space and time, while m and n represent m scenes in time and n pixels in space, respectively ($m = 2$ for STARFM and $m = 3$ for ESTARFM). $\overline{\delta_s^2}$ and $\overline{\delta_t^2}$ is the spatial variance and temporal variance, respectively. $\delta_t^2(\mu)$ and $\delta_s^2(\mu)$ represent the temporal variance of spatial means and spatial variance of time means, respectively.

Here, we assumed that the fusion accuracy was linear change with the temporal and spatial variance. Thus, a linear relationship between fusion accuracy and temporal and spatial variances was established. Therefore, a quantitative comparison of different spatiotemporal fusion algorithms with the distinct date and reference images could be carried out. To simplify the model and make the screen rule robust, we used a graded accuracy instead of the direct comparison of the accuracy value of spatio-temporal fusion algorithm for algorithm selection. Moreover, the coefficient of determination (R^2) ranging from 0 to 1 was used as the evaluation index of fusion accuracy. Thus, fusion accuracy was divided into different grades including high accuracy ($1 \geq R^2 > 0.8$), general high accuracy ($0.8 \geq R^2 > 0.7$), middle accuracy ($0.7 \geq R^2 > 0.6$), and low accuracy ($0.6 \geq R^2$). During the screening process, the fusion algorithm with a higher fusion accuracy grade was selected as the suitable image fusion algorithm for the prediction date image. This rule even worked under some unexpected situations such as the prediction R^2 exceeded one. Moreover, if STARFM and ESTARFM have the same fusion accuracy grade, both algorithms are suitable for image prediction. The screening rule can be defined as follows:

$$R^2 = a\overline{\delta_s^2} + b\overline{\delta_t^2} + c \tag{6}$$

$$R_{i,lower}^2 < a\overline{\delta_s^2} + b\overline{\delta_t^2} + c \leq R_{i,upper}^2, \text{ where } \overline{\delta_s^2} \geq 0 \text{ and } \overline{\delta_t^2} \geq 0 \tag{7}$$

where R^2 is the coefficient of determination, and a, b, and c are the regression coefficients. $\overline{\delta_s^2}$ and $\overline{\delta_t^2}$ are the mean spatial variance and mean temporal variance, while subscript i is the i-grade of fusion accuracy, and subscripts *lower* and *upper* represent the lower and upper limitations of R^2, respectively.

2.3. Performance Assessment

The performance of the framework was assessed from three aspects including the effectiveness of the screening rule, the effectiveness of Landsat NDVI time series in land cover change detection, and the effectiveness of phenological extraction. First, the R^2 values between the NDVI from the observation image and fusion image from STARFM or ESTARFM were respectively calculated to assess the screening rule's effectiveness. The higher the R^2 graded level was, the less the prediction uncertainty of the spatio-temporal fusion algorithm would be. Second, the Breaks for Additive Seasonal and Trend (BFAST) monitor, which can be conducted with the R package (https://CRAN.R-project.org/package=bfast) was used to detect the land cover change, which includes seasonal change, gradual change, and abrupt change [33]. Technically, abrupt change, which generally shows the conversion of land cover type within the short term, was employed in this paper [34]. The default parameter setting for the BFAST monitor can refer to [33,35]. As a typical classification accuracy index, the overall accuracy, omission error, and commission error of change detection were referred to as the accuracy indicators. In this paper, overall accuracy was the proportion of correctly prediction samples to total samples in which the total samples include land cover change samples and unchanged samples. The error of omission referred to the ratio of samples omitted from the land cover changed samples to total samples.

Commission error was defined as the ratio of incorrectly unchanged samples to total samples. Third, the mean absolute difference (MAD), referred to as the evaluation index, was used to represent the effectiveness of phenological node extraction, where the phenological nodes included the start of the growing season (SOS) and the end of the growing season (EOS), which was extracted by the dynamic threshold method [36]. In the dynamic threshold method, a single growing season cycle was commonly suggested in China's semi-arid region. The SOS and EOS are technically defined as the date when the NDVI increase and decrease to the seasonal amplitude of the left and right edge in the Savitzky–Golay (S–G) filtered NDVI annual curve [37]. The seasonal amplitude was determined by the relative threshold of 0.5, as suggested by previous studies [36,38]. The BFAST monitor can simply be described as first fitting the historical stability period NDVI time series by the season-trend model, and then detecting the land cover change based on the moving sums (MOSUMs) of the residuals in the motoring period, where the season–trend model can be described as follows:

$$y_t = \alpha_1 + \alpha_2 t + \sum_{j=1}^{k} \gamma_j sin\left(\frac{2\pi jt}{f} + \delta_j\right) + \varepsilon_t \tag{8}$$

where y_t is the prediction value of NDVI in time t; and α_1 and α_2 represent the intercept and slope. γ_j and δ_j represent the amplitude and phases for the jth season cycle. ε_t is the error term at time t. If the MOSUMs significantly deviate from zero, the land cover conversion took place. The MOSUMs can be described as follows:

$$MO_t = \frac{1}{\hat{\sigma} \sqrt{n}} \sum_{s=t-h+1}^{t} \left(y_s - x_s^T \hat{\beta}\right) \tag{9}$$

where h is the bandwidth of the MOSUMs. More details on the parameters of the BFAST monitor can be found in [33,39].

3. Experimentation and Effectiveness Analysis

The framework was implemented in the Shendong coal mining area (Figure 2). To construct the dual high-resolution Landsat NDVI time series for land cover monitoring, the NDVI extracted from all available Landsat images (Figure 3) was treated by maximum-value composites with a 16-day interval. Dates, without a synthetic image or synthetic images contaminated area (cloud, cloud shadow, and snow) higher than 50%, were screened as the prediction dates. These fine images in prediction dates were blended by the spatio-temporal fusion algorithm selected based on the screening rule. The screening rule in the framework was first determined by the linear fitted relationship between the spatial and temporal variance and spatio-temporal fusion algorithm accuracy (Table 2), and then the fitted relationship was combined with the graded accuracy selecting rule (R2) to select the fusion algorithm. Specifically, the fitting samples employed the spatio-temporal fusion results in 2005 (Figure 4). A 16-day interval Landsat NDVI time series was constructed by interpreting the fusion results instead of images in the prediction dates (Figure 5). After fusing the images in the prediction dates, the performance of the framework was conducted (Table 3, Figures 6–8).
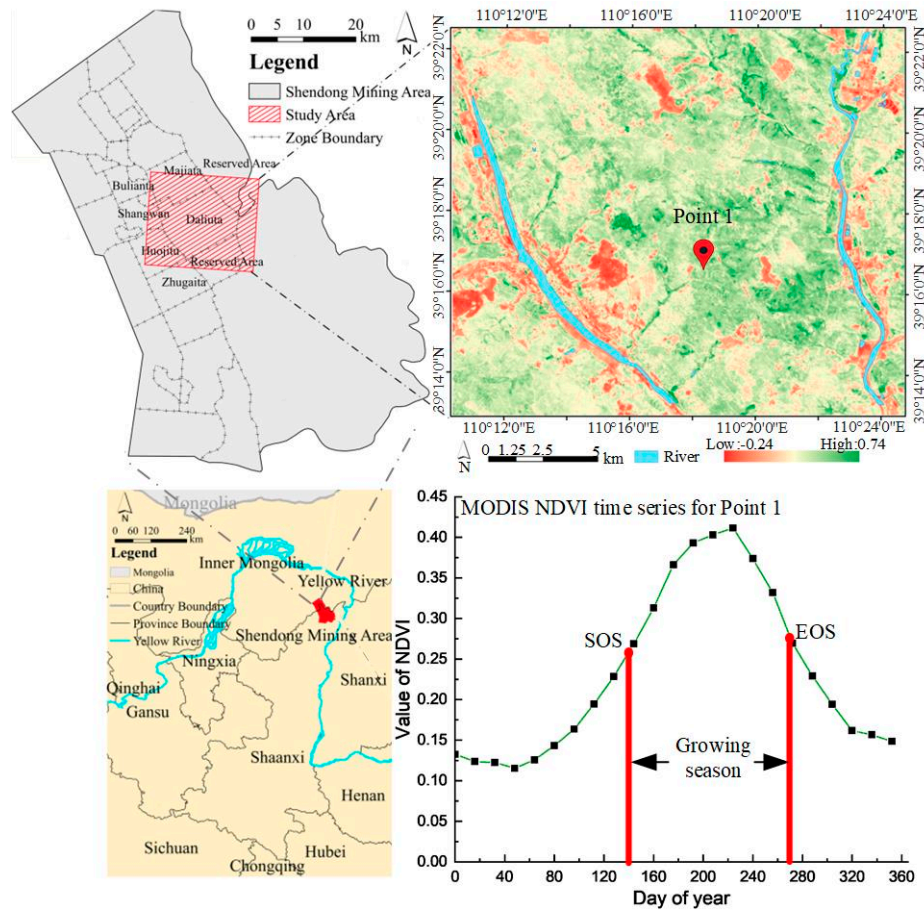
Figure 2. Location of the experimental area in the Shendong mining region. The normalized difference vegetation index (NDVI) was calculated by using a Landsat image (scene ID: LE07_L1TP_127033_20170730_20170825_01_T1_ANG). The phenology curve was extracted from the The Terra Moderate Resolution Imaging Spectroradiometer (MODIS) Vegetation Indices (MOD13Q1) NDVI time series at point 1, where the SOS and the EOS represent the start of growing season and the end of growing season, respectively.
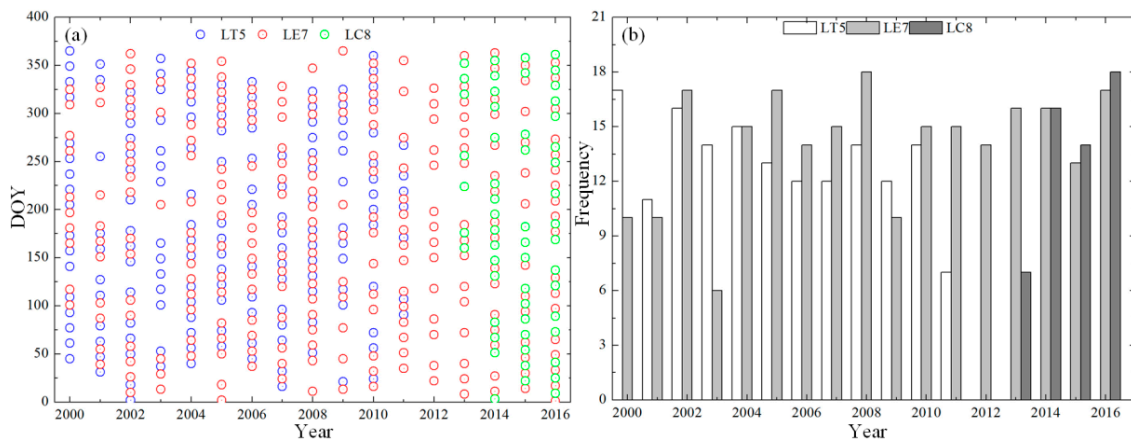

Figure 3. All available Landsat images with cloud coverage less than 70% from 2000 to 2016, where the LT5, LE7, and LC8 represent the Landsat 5 TM, the Landsat 7's ETM+ sensor, and the Landsat 8 OLI sensor, respectively. (**a**) Day of year (DOY) distribution of Landsat images, and (**b**) quantity statistics of Landsat images in every year.
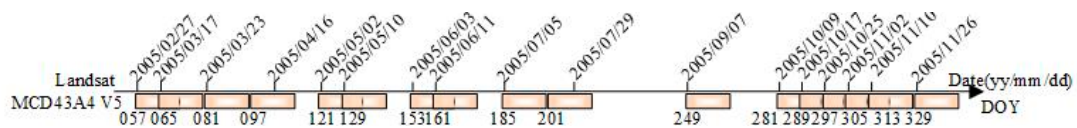
Figure 4. Landsat and MODIS in 2005 were used in the spatiotemporal fusion algorithm screening scheme establishment, where V5 represents version 5 of the MCD43A4 datasets. The date shows the observation date of the Landsat image, while the DOYs show the first day of year of the MCD43A4 16-day composite images.
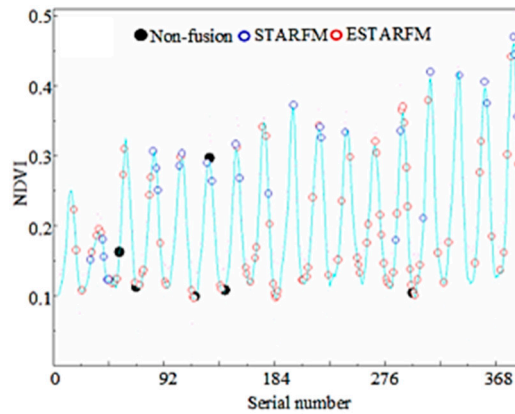

Figure 5. All predicted images from 2000 to 2016 where the spatio-temporal fusion algorithms were selected based on the screening rule. The non-fusion represents no suitable spatio-temporal algorithm.
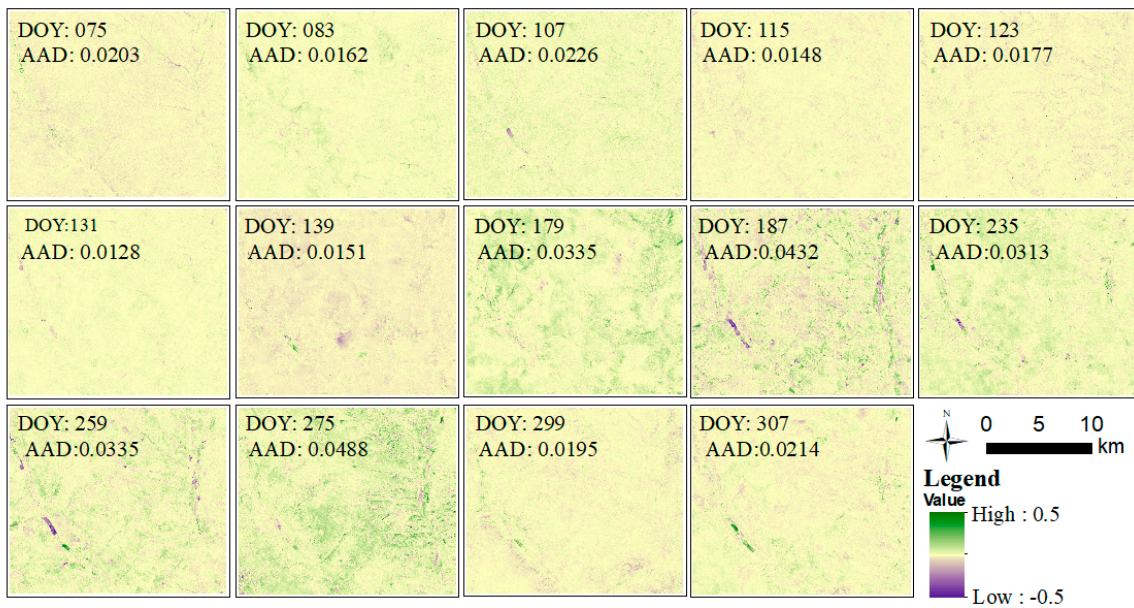

Figure 6. Difference between NDVI images by using the NDVI calculated from Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) blended images minus the NDVI calculated from Enhanced STARFM (ESTRAFM) blended images.
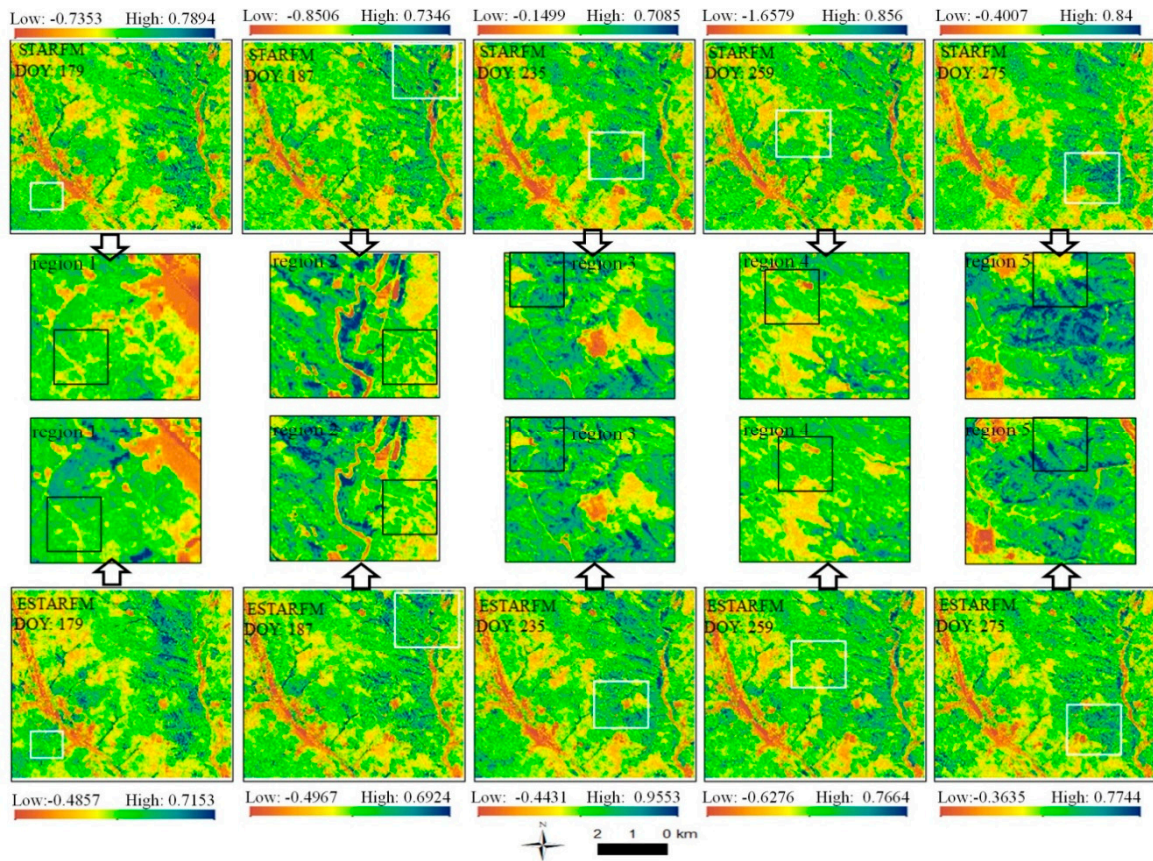
Figure 7. A comparison of the local enlarged region of the NDVI extracted from both STARFM and ESTARFM prediction images in the growing season.
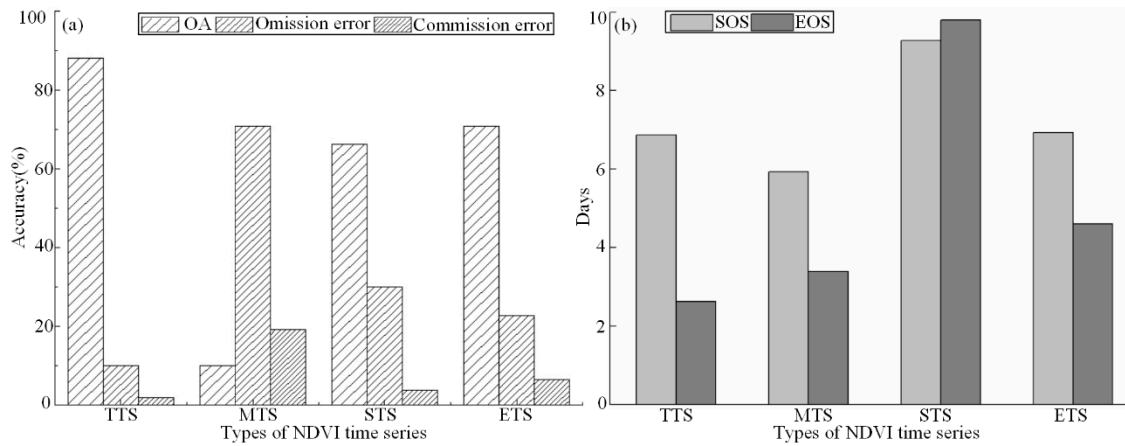

Figure 8. Landsat NDVI performance in land cover change detection and phenological stability analysis, where SOS and EOS represent the start of the growing season and the end of the growing season, and overall accuracy. (**a**,**b**) represent the statistics of land cover change detection and phenological stability, respectively.

3.1. Experimental Area and Datasets

3.1.1. Experimental Area

The experimental area was located on the border of Shanxi Province and Inner Mongolia [40]. It is situated in a typical ecologically fragile area with a high intensity of human disturbance due to exploitation of the Shendong coal mining area (the largest field and a typical underground mining

area in China). It is the transition zone between the Loess Plateau and the Mu Us Sandy Land [41]. This area has been gaining a lot of attention from ecologists, geographers, and surveyors in terms of analyzing ecosystem degradation [42], land cover change [34], and terrain deformation [43]. Moreover, compared with other ecosystems, the land cover in the coal mining area contains frequently abrupt and gradual change, and has an obvious season change, even in a small region [34,42]. Land cover change detection has always been the spatio-temporal fusion algorithm that designers care about [24,44]. Thus, this research selected a 689 × 590-pixel (30×30 m^2) area in the Shendong mining area as the experimental site (Figure 2), which contains multiple mining fields (including Majiata, Bulianta, Shangwan, Daliuta, Huojitu, and Zhugaita) and part reserved land. The experimental site range was from 39°13′57″N to 39°21′32″N, and 110°12′33″E to 110°22′52″E. This area has a plateau continental climate with arid, half-desert land. The perennial mean temperature is about 8.6 °C, the average yearly precipitation is about 380 mm, and the yearly evaporation ability is about 2113 mm. Compared with humid regions, vegetation cover in the study region is sparse and with an obvious single growing season. As a result of high-intensity mining and human activities, the landscape fragmentation of the land cover is high, and the majority of the land cover includes a mining pit, waste occupancy area, buildings, roads, water, forest, deserted forest land, mixed forests, shrubs, grassland and deserted grassland, desert, and farmland.

3.1.2. Datasets and Preprocessing

A total of 450 level-1 terrain-corrected Landsat images with cloud coverage less than 70% (from 2000 to 2016; Figure 3) including Thematic Mapper (TM), ETM+ and Operational Land Imager (OLI) products (WRS-2 path/row: 127/33) were download from the U.S. Geological Survey (USGS) of the Earth Resources Observation and Science (EROS) Center. The Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) [45] and Landsat 8 Surface Reflectance (L8SR) [46] were used to generated surface reflectance products of TM, ETM+, and OLI. The cloud, cloud shadow, and snow were masked from Landsat images using the Fmask algorithm [47]. Due to narrower spectral bands and improved radiometric resolution and signal-to-noise ratio, the surface reflectance product of OLI was significantly improved when compared with TM/ETM+ [48]. Therefore, taking ETM+ as the reference, relative normalization [49] was employed to reduce the surface reflectance inconsistencies between ETM+ and OLI. Furthermore, Landsat L1T level images reached a sub-pixel level, thus, Landsat did not need further geometric correction [50]. The MODIS MCD43A4 Version 5 Nadir Bidirectional Reflectance Distribution Function (BRDF)-Adjusted Reflectance (NBAR) dataset in tile h26/v05 was selected as the inputted coarse image for the spatio-temporal fusion algorithm [51] (https://ladsweb.modaps.eosdis.nasa.gov/). The MCD43A4 datasets were daily composite products according to a 16-day window, where each composite value is on behalf of the best BRDF possible [52]. Thus, the MODIS MCD43A4 datasets are broadly used as the coarse images for spatio-temporal fusion [44,52]. Moreover, due to its moderate spatial resolution (250 m) and high temporal resolution as well as its proven good performance in phenology, the terra moderate resolution imaging spectroradiometer (MODIS) vegetation indices (MOD13Q1) NDVI time series has been taken as a reference for phenology monitoring in many previous studies [53,54]. Thus, the phenological parameter extracted from the MOD13Q1 NDVI time series from 2000 to 2016 was employed as a base reference of the Landsat NDVI time series, where the phenological parameter was extracted based on the dynamic threshold method. To quantitate the relationship of the Earth's surface spatial and temporal changes with the fusion accuracy of the spatiotemporal fusion algorithm, the fusion results in 2005—with different time intervals by adjusting the reference images in the algorithm—were employed, and the historical fine images are represented in Figure 4. Furthermore, to compare the accuracy of near real-time change detection of land cover based on the NDVI time series constructed by different fusion strategies, five high-resolution images were selected as the reference for cross-validation samples ranging from 2012 to 2016 in the growing season (Table 1). All high-resolution images were processed with orthorectification, layer stacking, relative registration, fusion, and clipping.

Table 1. List of high-resolution images where ZY-03 and GF-1 represent the Chinese Ziyuan 3 and Gaofeng 1 satellites. Pan refers to the panchromatic band, MS refers to the multispectral band, and B, G, R, and NIR refer to blue, green, red, and near-infrared, respectively. "-" represents null.

Date	Satellite	Resolution (Pan/MS)	Band	ID Scene
2012-9-30	ZY-03	2.1/5.8	B, G, R, NIR	ZY3_MUX_E110.4_N39.0_20120930_L1A0001611699
2013-9-6	SPOT 6	1.5/6	Pan, B, G, R, NIR	DIM_SPOT6_MS_201309060254349_SEN_14705884-0_01
2014-8-7	Pleiades	0.5/2	Pan, B, G, R, NIR	DS_PHR1A_201408070345035_SE1_PX_E110N39_0407_00804
2015-8-17	GF-1	-/16	B, G, R, NIR	GF1_WFV4_E110.4_N38.5_20150817_L1A0000985438
2016-9-13	GF-1	2/8	Pan, B, G, R, NIR	GF1_PMS1_E110.3_N39.4_20160913_L1A0001824149

3.2. Construction of Landsat Normalized Difference Vegetation Index (NDVI) Time Series

The MODIS prediction date and MODIS base date in 2005, as input images of the spatial and temporal variance, were used to express the image change with spatial and temporal variance. Additional experiments were carried out to increase the fitting samples by expanding the time interval between the fusion and reference images. Additionally, the samples were divided into growth and non-growing seasons due to the significant difference in fusion accuracy (Table 2). The fitting results indicated that a linear relationship existed among the fusion algorithm accuracy and spatiotemporal variance. It is worth noting that to improve the robustness of high accuracy, over-fitting accuracy prediction in the discrimination function was allowed in the present study. If two suitable algorithms exist, STARFM will be selected due to its advantage, whereas if no suitable algorithm exists, the image will not be predicted. No suitable algorithm situations appeared when the predicted R^2 was lower than 0.6. As a result, 126 scenes in the study area were fused, among which ESTARFM fused 29 scenes, STARFM fused 91 scenes, and six scenes were not fused (Figure 5). Moreover, we only predicted the images in the dates when lacking a synthetic image or synthetic images with a contamination area (cloud, cloud shadow, and snow) that was higher than 50%. As a result, the residual contamination area still existed in the interpreted Landsat NDVI time series. Thus, the S–G filter was employed to reduce the noise caused by residual contamination area (cloud, cloud shadow, and snow) in raw Landsat [55], which was implemented by the TIMESAT with a 5-pixel-width filter window (http://web.nateko.lu.se/TIMESAT/timesat.asp).

Table 2. Regression equation of spatiotemporal variance and R^2. S_GS and ES_GS, respectively, represent Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) and Enhanced STARFM (ESTARFM) in the growing season. S_NGS and ES_NGS represent STARFM and ESTARFM in the non-growing season, respectively.

Type	Regression Function	Adjusted R^2	F-Value	Samples
S_GS	$R^2 = 24.77\overline{\delta_s^2} - 39.67\overline{\delta_t^2} + 0.83$	0.82	3.64×10^{-6}	25
S_NGS	$R^2 = 607.57\overline{\delta_s^2} - 115.55\overline{\delta_t^2} + 0.50$	0.89	3.65×10^{-8}	25
ES_GS	$R^2 = 109.42\overline{\delta_s^2} - 13.40\overline{\delta_t^2} + 0.66$	0.85	9.57×10^{-13}	38
ES_NGS	$R^2 = 1110.81\overline{\delta_s^2} - 41.94\overline{\delta_t^2} + 0.25$	0.83	2.076×10^{-6}	29

3.3. Effectiveness Analysis

In this section, the proposed framework's performance in algorithm selection and the effectiveness of the Landsat NDVI time series to detect land cover change and phenological nodes were verified.

3.3.1. Effectiveness of Algorithm Screening

To test the effectiveness of the screening rule, a total of 14 Landsat scenes from 2008 were employed as validation data (Table 3). The spatial and temporal variances between the reference data for each scene were calculated, and then the screening scheme was applied to select a fusion algorithm. To select different regression functions for fusion images in different phenological phases, the average start and end of the growing-seasons extracted from MODIS13Q1 were used to distinguish the growing season from the non-growing season. The average SOS was the 154 day of the year, while the average EOS was the 278 day of the year. Table 3 shows that, except for two prediction images (DOY 075 and 107), the screening rule could effectively determine a suitable fusion algorithm for the special fusion image according to the R^2 accuracy grade calculated from the observational Landsat NDVI and predicted Landsat NDVI. The reason for choosing an unsuitable algorithm for image prediction in DOY 075 and 107 might have been caused by the potential cover of snow or ice in the reference images. These potential snow and ice cover areas caused a low variance in either the spatial or temporal dimension. However, it is worth noting that the scope of the screening rule was to select a higher R^2 grade algorithm from the basic algorithms, so its ability to select a higher value of R^2 under the same grade was poor. From the visualizing perspective (Figure 6), the difference between NDVIs of the prediction images from STARFM and ESTARFM was calculated by using the NDVI of STARFM blended images minus the NDVI of ESTRAFM blended images. The results indicated that the difference between the STARFM and ESTARFM was higher in the growing season than in the non-growing season. For example, the AAD in DOY 179, 187, 235, 259, and 275 were greater than 0.03, while the AAD in DOYs in the non-growing season was less than 0.03. Moreover, in the growing season, the NDVI difference showed that the NDVIs calculated from the STARFM predicted images was higher than the NDVIs calculated from the ESTARFM predicted images. A comparison of the enlarged local region of the NDVI extracted from both STARFM and ESTARFM predicted images in the growing season were conducted (Figure 7). The result indicated that the ESTARFM not only caught the spatial details (Figure 7 region 1, region 2), but predicted well in the gradual change (Figure 7 region 4, region 5). However, STARFM defeated the ESTARFM in land cover conversion caused by an abrupt change of land cover (Figure 7, region 3). It further verified the effectiveness of the screening rule by selected ESTARFM for predicting images in the growing season, if we assume that the land cover gradually changed in 2008.

Table 3. Screening results of the spatio-temporal fusion algorithm in 2008 based on the regression equation of spatiotemporal variance and R2. DOY represents the day of year.

DOY	R^2				Selected Algorithm
	STARFM	Accuracy Grade	ESTARFM	Accuracy Grade	
075	0.6321	middle accuracy	0.7326	general high accuracy	S
083	0.7781	general high accuracy	0.7091	general high accuracy	S
107	0.6950	middle accuracy	0.7220	general high accuracy	S
115	0.7640	general high accuracy	0.8292	high accuracy	E
123	0.8341	high accuracy	0.8717	high accuracy	E
131	0.8800	high accuracy	0.8902	high accuracy	E
139	0.8866	high accuracy	0.8733	high accuracy	E
179	0.9192	high accuracy	0.8618	high accuracy	E
187	0.8149	high accuracy	0.9105	high accuracy	E
235	0.9187	high accuracy	0.8863	high accuracy	E
259	0.9124	high accuracy	0.9416	high accuracy	E
275	0.8074	high accuracy	0.9054	high accuracy	E
299	0.8494	high accuracy	0.8068	high accuracy	E
307	0.7678	high accuracy	0.7533	general high accuracy	E

3.3.2. Land Cover Change Detection and Phenological Stability Analysis

The effectiveness of a method is often related to its application purpose. Thus, further steps for near real-time monitoring were carried out by comparing the target Landsat NDVI time series (TTS) generated from the framework that this paper proposed with the MODIS NDVI time series (MTS) and NDVI time series using single algorithm fusion (STS: the STRAFM based Landsat NDVI time series, and ETS: the ESTRAFM based Landsat NDVI time series). BFAST Monitor [33] was selected as the change detection algorithm, and 110 change samples ranging from 2012 to 2016 were selected as validation samples by visual interpretation (using high-resolution images; Table 1). The overall accuracy, omission error, and commission error were calculated as the accuracy indicators. Figure 6 shows the accuracy results of different NDVI time series. The results indicated that TTS obtained the highest overall accuracy (88.18%), but also the lowest omission error (1.82%) and commission error (10.00%), while in turn, MTS achieved the lowest overall accuracy (10.00%) but the highest omission (19.09%) and commission error (70.91%). When compared with STS and ETS, respectively, the overall accuracy of TTS increased by 21.82% and 17.27%, the omission error decreased by 1.82% and 4.54%, and the commission error decreased by 20% and 12.73%. The accuracy demonstrated that the construction of the Landsat NDVI time series using multiple fusion algorithms with the framework could efficiently improve the application accuracy. On the other hand, to test the ability of TTS in phenological change, a phenological stability analysis was implemented using MAD as the evaluation index, and the phenological parameters of SOS and EOS were extracted by the dynamic threshold method [56]. Under the hypothesis that phenological features exhibited long-term stability where the year 2000 was referred to as the reference year, a lower MAD indicated a more stable phenology. The phenological stability results (Figure 8b) demonstrated that the TTS's SOS stability and EOS stability were lower than the STS, but higher than ETS. STS possessed the highest stability in SOS and EOS with MAD, respectively equal to six days and four days, while ETS obtained the lowest stability with MAD, respectively equal to seven days and nine days. These results indicated that combining multiple fusion algorithms to construct a time series tends to gain more comprehensive results and effectively avoid phenological fluctuations caused by using a single fusion algorithm.

4. Discussion

In this section, the advantages and adaptability of the framework proposed in this paper are discussed. Although this framework worked well in the construction of Landsat NDVI time series, there are still many uncertainties and challenges from the datasets and screening rules. Thus, we discuss the framework in detail, so that we can be more familiar with it and its potentials and shortages.

4.1. Advantages and Adaptability

As one of the most effective approaches to predict the missing high-resolution images for time series construction, spatio-temporal fusion algorithms have been extensively developed for different purposes [30,57]. However, a balance between the required minimal number of input image pairs, performance of heterogeneous land cover, and calculation complexity is always needed when using a spatio-temporal fusion algorithm to generate a Landsat NDVI time series [44,58]. Emelyanova et al. (2013) proposed that different algorithms had their advantages and disadvantages by comparing the accuracies of the advanced blending algorithms (STARFM and ESTARFM) and the simple benchmarking algorithms (including linear interpolation model (LIM) and global empirical image fusion model (GEIFM)) [50] by taking the spatial and temporal variance as assessment metrics. Thus, this paper explored the possibility of employing multiple spatio-temporal fusion algorithms to construct a dual high-resolution Landsat NDVI time series. Hence, the prediction images would be predicted by an adaptable algorithm. It is worth noting that the basic spatio-temporal fusion algorithms in the framework should be complementary. For example, STARFM and ESTARFM were selected in this paper as basic algorithms because the STARFM was efficient and performed well in a homogeneous

land cover area with one reference image pair, and the ESTARFM was relatively time-consuming, needed at least two coarse and fine reference image pairs, but performed well in a heterogeneous land cover area [7,24]. The performance of this framework would be worse if the selected base algorithm not only had a good performance in a heterogeneous area and different kind of land cover change area, but also operated efficiently and just needed one pair of reference images, which include algorithms such as the Spatial and Temporal Reflectance Unmixing Model (STRUFM) in [59], the improved STRUFM (ISTRUFM) in [60], and the Flexible Spatiotemporal DAta Fusion (FSDAF) method in [12]. To the best of the authors' knowledge, no previous studies have used multiple fusion algorithms to construct a long-term time series, and most of the research work to date has only employed a single algorithm [7,61]. Möller et al. (2017) generated a 30-m and 16-day interval time series utilizing STARFM to monitor soil erosion [4]. Moreover, other STARFM-generated time series were described in the study by Tian et al. (2013), which analyzed the land cover trends in the Loess Plateau [5] as well as in the studies by Schmidt et al. (2015) and Bhandari et al. (2012), which respectively detected the forest disturbance and recovery in Queensland, Australia [6,53]. From the application perspective, the constructed Landsat NDVI time series based on the framework proposed in this paper could be broadly applied for land cover change monitoring, phenological extraction, and land use and land cover classification. The near real-time monitoring method or its improved version is perfect for catching the land surface change including abrupt change, gradual change, and seasonal change, if the land cover change detection is the only application scope [33,35,39]. Therefore, a screening framework for multiple fusion algorithm selection can enrich the theory for Landsat NDVI time series construction, and result in the inclusion of more details in the Landsat NDVI time series compared to that in a time series constructed based on a single algorithm. On the other hand, temporal and spatial changes among images were emphasized in the proposed screening rule using the spatial and temporal variance, which provides an opportunity to quantitatively express the correlation between fusion performance and the Earth's surface change.

4.2. Uncertainty and Challenges

Although the framework performed well in the algorithm selected and the constructed Landsat NDVI time series was proven to be valid in land cover detection and phenological extraction, there are still many uncertainties that might affect the results. First, this paper fused the reflectance of images first, and then calculated the NDVI. This process has been reported as inferior to the direct fusion NDVI process. As a result, extra errors would be introduced in the Landsat NDVI time series [62,63]. Additionally, this paper only blended these Landsat images with the fine pixel covered area less than 50%. Those images with contaminated areas by cloud, shallow, ice, and observed gaps were interpreted by the S–G filter. The interpretation process might bring extra errors into the Landsat NDVI time series construction [64]. These interpolated areas showed an obvious difference from the true value (areas in the red rectangle of Figure 9). Moreover, the screened algorithm was confirmed to meet the R^2 graded accuracy, but it was unsatisfactory for selecting the spatio-temporal fusion algorithm with the highest accuracy value (Table 4). Furthermore, as this paper simply supposed a linear relationship of the fusion graded accuracy and spatial and temporal variance, the overfitting or non-linear relationship between the fusion accuracy and the spatial and temporal variance might also introduce errors in algorithm selection [15,39]. Ideally, it was expected to improve the stability of the screening rule when employing the comprehensive accuracy indexes or other accuracy assessment indexes [14,62]. Other accuracy assessment indexes include the average absolute difference (AAD), the root mean square error (RMSE), the structural similarity (SSIM) [65], and the universal image quality index (UIQI) [66].
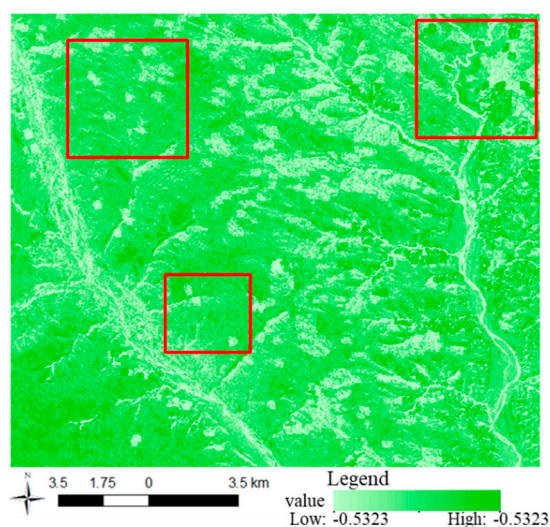
Figure 9. The effect of cloud, cloud shadow, or ice to one of NDVI in 2000, which was interpolated by the Savitzky–Golay (S–G) filter.

Table 4. Accuracy comparison of STARFM and ESTARFM, where AAD, RMSE, SSIM, and UIQI represent the average absolute difference (AAD), the root mean square error (RMSE), the structural similarity (SSIM), and the universal image quality index (UIQI).

DOY	AAD		RMSE		SSIM		UIQI	
	STARFM	ESTARFM	STARFM	ESTARFM	STARFM	ESTARFM	STARFM	ESTARFM
75	0.0209	0.0164	0.0282	0.0227	0.4409	0.5438	0.9565	0.9764
83	0.0163	0.0212	0.0223	0.0285	0.6111	0.5703	0.9791	0.9594
107	0.0305	0.0221	0.0379	0.0285	0.4381	0.4751	0.9382	0.9635
115	0.0183	0.0145	0.0243	0.0195	0.5841	0.7126	0.9759	0.9846
123	0.0150	0.0179	0.0202	0.0248	0.6557	0.7056	0.9869	0.9815
131	0.0226	0.0248	0.0315	0.0336	0.7748	0.7713	0.9772	0.9742
139	0.0228	0.0264	0.0314	0.0361	0.7206	0.7089	0.9788	0.9724
179	0.0273	0.0283	0.0364	0.0390	0.8277	0.7928	0.9820	0.9752
187	0.0437	0.0261	0.0636	0.0356	0.6709	0.8415	0.9543	0.9830
235	0.0659	0.0818	0.0830	0.1002	0.7495	0.7243	0.9592	0.9389
259	0.0685	0.0512	0.0881	0.0629	0.7296	0.8356	0.9433	0.9665
275	0.0393	0.0310	0.0538	0.0407	0.6190	0.8035	0.9716	0.9796
299	0.0300	0.0291	0.0388	0.0389	0.6164	0.6136	0.9671	0.9661
307	0.0236	0.0255	0.0312	0.0361	0.5105	0.5427	0.9694	0.9563

Furthermore, even if the construction of Landsat NDVI time series using the framework has been proven to be effective for land cover change detection, some uncertainties should be considered, and further studies should be conducted. Additionally, more samples from different years need to be generated. This study considered the fusion result of the year with the most cloudless images (2005) as the fitting samples (Figure 10), which may introduce errors in the quantitative model. Therefore, further analysis of the fitting model using all fusion images of all available and cloud-free images as samples needs to be carried out. Additionally, due to its computing efficiency [8,67], the STARFM was artificially more frequently selected than ESTARFM in the Landsat NDVI time series when the same R^2 grade of STARFM and ESTARFM appeared in the algorithm screening process. Furthermore, the fusion images around the phenological nodes (SOS and EOS) should receive more attention. The comparison of fusion algorithms demonstrated that the effectiveness of each fusion algorithm has significant differences between the growth and non-growing seasons [50]. Thus, the fitting function for each fusion algorithm was divided into the growing season and the non-growing season. As a result, the determination of the growth stage should be performed before image fusion, and the lack of phenology reference for incomplete years increases the uncertainty in the fusion algorithm

selection. Furthermore, the framework proposed in this paper was applied in a step by step time series construction using multiple fusion algorithms, while continuous automated screening mechanisms and time series construction need to be implemented in further research. Finally, compared to all pixels, using a sliding window to quantify spatiotemporal differences among reference images may obtain a local optimization fusion result since STARFM and ESTARFM have their respective advantages in different land types. Last but not least, the framework is only a kind of proof-of-concept. These algorithms, which not only consider the spatial and temporal difference, but integrate the multiple processes of Landsat NDVI time series construction, would be desirable [13,16,26].
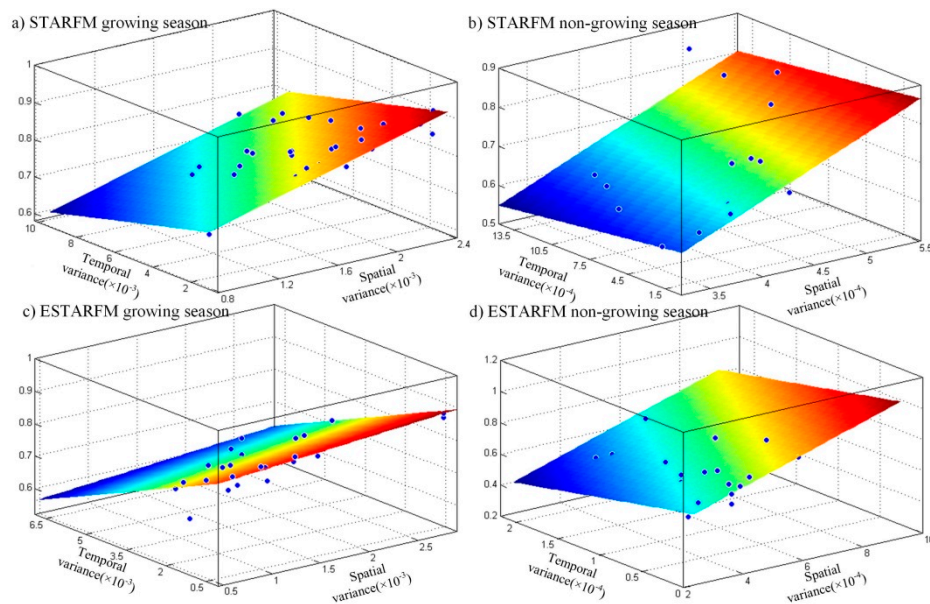

Figure 10. Fitting surface and its samples. (**a**,**c**) respectively represent the STARFM and ESTARFM fitting result in the growing season, while (**b**,**d**) indicate the non-growing season fitting results of STARFM and ESTARFM, respectively.

4.3. Perspectives

The framework proposed in this paper can be regarded as a proof-of-concept, and its implementation was proven to be successful. The greatest prospect of this framework was to construct long-term and updated timely Landsat time series for international institutions or programs such as for UNITAR's Operational Satellite Applications Program (UNOSAT) in disaster rapid mapping, the United Nations Collaborative Program on Reducing Emissions from Deforestation and Forest Degradation (UN-REDD) in deforestation detection, and the United Nations Environment Program (UNEP) in environment hotspots, to increase the ability of near real-time environmental protection and monitoring. Furthermore, this framework is expected to be used in a cloud-based platform such as Google Earth Engine (GEE), which not only contains petabyte-scale Landsat archives, but can also provide multiple MODIS products [68]. Using the Internet-accessible application programming interface (API) of GEE, this framework might realize full automation as preprocessing of dual high-resolution Landsat time series construction. Indeed, users could directly utilize Landsat time series, fused images, and algorithm screening carried out in a cloud computing platform and would only need to determine the required data attributes (such as position, spatial resolution, time resolution, feature index, etc.) for near real-time monitoring.

5. Conclusions

Each model has its advantages and disadvantages. Making full use of complementary information among different models is an effective way to improve application accuracy. This study assumed that

selecting an appropriate algorithm from multiple complementary spatio-temporal fusion algorithms according to the spatial and temporal variance for each prediction image in a time series construction would improve the Landsat NDVI time series performance. In addition, we supposed that a linear relationship existed between the fusion accuracy and spatial and temporal variance. Thus, a framework for Landsat NDVI time series construction with multiple spatiotemporal fusion algorithms was established. This framework was applied in a typical fragile and high heterogeneity ecosystem to construct a Landsat NDVI time series with a 30 m spatial resolution with a 16-day interval. The application of near real-time monitoring proved the effectiveness of this framework. This paper mainly contains two contributions. First, a fusion algorithm screening framework for constructing a Landsat time series was proposed, resulting in more detailed information in the time series than that included when using a single algorithm. Second, the Earth's surface change was divided into space and time according to recorded images, and the algorithm suitability was quantitatively expressed, which can make the screening scheme more adaptable and robust.

Author Contributions: Conceptualization, Yangnan Guo; Data curation, Yangnan Guo and Cangjiao Wang; Formal analysis, Yangnan Guo; Methodology, Cangjiao Wang; Software, Shaogang Lei; Supervision, Shaogang Lei; Validation, Shaogang Lei; Visualization, Yangnan Guo and Cangjiao Wang; Writing—original draft, Yangnan Guo and Cangjiao Wang; Writing—review & editing, Shaogang Lei, Junzhe Yang, and Yibo Zhao. All authors have read and agreed to the published version of the manuscript.

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

Abbreviations

STARFM	Spatial and Temporal Adaptive Reflectance Fusion Model
ESTARFM	Enhanced STARFM
NDVI	Normalized Difference Vegetation Index
MODIS	Moderate Resolution Imaging Spectroradiometer
GIMMS	Global Inventory Modeling And Mapping Studies
ETM+	Enhanced Thematic Mapper Plus
R^2	Coefficient Of Determination
BFAST	Breaks For Additive Seasonal and Trend
MAD	Mean Absolute Difference
SOS	Start Of Growing Season
EOS	End of Growing Season
S–G	Savitzky–Golay Filter
MOSUMs	Moving Sums
TM	Thematic Mapper
OLI	Operational Land Imager
LEDAPS	Landsat Ecosystem Disturbance Adaptive Processing System
USGS EROS	U.S. Geological Survey Of The Earth Resources Observation And Science Center
L8SR	Landsat 8 Surface Reflectance
BRDF-NBAR	Nadir Bidirectional Reflectance Distribution Function-Adjusted Reflectance
DOY	Day Of Year
TTS	Target Landsat NDVI Time Series
MTS	MODIS NDVI Time Series
STS	STRAFM Based Landsat NDVI Time Series
ETS	ESTRAFM Based Landsat NDVI Time Series
LIM	Linear Interpolation Model

STRUFM	Spatial And Temporal Reflectance Unmixing Model
ISTRUFM	Improved STRUFM
FSDAF	Flexible Spatiotemporal DAta Fusion
SSIM	Structural Similarity
RMSE	Root Mean Square Error
UIQI	Universal Image Quality Index
UNOSAT	UNITAR's Operational Satellite Applications Program
UN-REDD	United Nations Collaborative Program On Reducing Emissions From Deforestation And Forest Degradation
GEE	Google Earth Engine
API	Application Programming Interface

References

1. Gómez, C.; White, J.C.; Wulder, M.A. Optical remotely sensed time series data for land cover classification: A review. *ISPRS J. Photogramm. Remote Sens.* **2016**, *116*, 55–72. [CrossRef]

2. Chen, J.; Zhu, X.; Vogelmann, J.E.; Gao, F.; Jin, S. A simple and effective method for filling gaps in Landsat ETM+ SLC-off images. *Remote Sens. Environ.* **2011**, *115*, 1053–1064. [CrossRef]

3. Moller, M.; Gerstmann, H.; Thurkow, D.; Gao, F. Coupling of phenological information and synthetically generated time-series for crop types as indicator for vegetation coverage information. In Proceedings of the Analysis of Multitemporal Remote Sensing Images, Annecy, France, 22–24 July 2015; pp. 1–4.

4. Möller, M.; Gerstmann, H.; Gao, F.; Dahms, T.C.; Förster, M. Coupling of phenological information and simulated vegetation index time series: Limitations and potentials for the assessment and monitoring of soil erosion risk. *CATENA* **2017**, *150*, 192–205. [CrossRef]

5. Tian, F.; Wang, Y.; 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]

6. Schmidt, M.; Lucas, R.; Bunting, P.; Verbesselt, J.; Armston, J. Multi-resolution time series imagery for forest disturbance and regrowth monitoring in Queensland, Australia. *Remote Sens. Environ.* **2015**, *158*, 156–168. [CrossRef]

7. Chen, B.; Huang, B.; Xu, B. Comparison of Spatiotemporal Fusion Models: A Review. *Remote Sens.* **2015**, *7*, 1798–1835. [CrossRef]

8. Zhu, X.; Chen, J.; Gao, F.; 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]

9. Gao, F.; Hilker, T.; Zhu, X.; Anderson, M.; Masek, J.; Wang, P.; Yang, Y. Fusing Landsat and MODIS Data for Vegetation Monitoring. *IEEE Geosci. Remote Sens. Mag.* **2015**, *3*, 47–60. [CrossRef]

10. Huang, B.; Zhang, H. Spatio-temporal reflectance fusion via unmixing: Accounting for both phenological and land-cover changes. *Int. J. Remote Sens.* **2014**, *35*, 6213–6233. [CrossRef]

11. Jia, D.; Gao, P.; Cheng, C.; Ye, S. Multiple-feature-driven co-training method for crop mapping based on remote sensing time series imagery. *Int. J. Remote Sens.* **2020**, *41*, 8096–8120. [CrossRef]

12. Zhu, X.; Helmer, E.H.; Gao, F.; Liu, D.; Chen, J.; Lefsky, M.A. A flexible spatiotemporal method for fusing satellite images with different resolutions. *Remote Sens. Environ.* **2016**, *172*, 165–177. [CrossRef]

13. Rao, Y.; Zhu, X.; Chen, J.; Wang, 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.* **2015**, *7*, 7865–7891. [CrossRef]

14. Zhu, Z.; Woodcock, C.E.; Holden, C.; Yang, Z. Generating synthetic Landsat images based on all available Landsat data: Predicting Landsat surface reflectance at any given time. *Remote Sens. Environ.* **2015**, *162*, 67–83. [CrossRef]

15. Samanta, A.; Costa, M.H.; Nunes, E.L.; Vieira, S.A.; Xu, L.; Myneni, R.B. Comment on "Drought-Induced Reduction in Global Terrestrial Net Primary Production from 2000 Through 2009". *Science* **2011**, *329*, 940–943. [CrossRef]

16. Luo, Y.; Guan, K.; Peng, J. STAIR: A generic and fully-automated method to fuse multiple sources of optical satellite data to generate a high-resolution, daily and cloud-/gap-free surface reflectance product. *Remote Sens. Environ.* **2018**, *214*, 87–99. [CrossRef]

17. 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.

18. 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]

19. Wu, B.; Huang, B.; Cao, K.; Zhuo, G. Improving spatiotemporal reflectance fusion using image inpainting and steering kernel regression techniques. *Int. J. Remote Sens.* **2017**, *38*, 706–727. [CrossRef]

20. Zhang, W.; Li, A.; Jin, H.; Bian, J.; Zhang, Z.; Lei, G.; Qin, Z.; Huang, C. 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]

21. Wu, M.; Niu, Z.; Wang, C.; Wu, C.; Wang, L. Use of MODIS and Landsat time series data to generate high-resolution temporal synthetic Landsat data using a spatial and temporal reflectance fusion model. *J. Appl. Remote. Sens.* **2012**, *6*, 063507.

22. Song, H.; Huang, B. Spatiotemporal satellite image fusion through one-pair image learning. *IEEE Geosci. Remote. Sens. Lett.* **2012**, *51*, 1883–1896. [CrossRef]

23. Song, H.; Liu, Q.; Wang, G.; Hang, R.; Huang, B. Spatiotemporal satellite image fusion using deep convolutional neural networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 821–829. [CrossRef]

24. Jia, D.; Song, C.; Cheng, C.; Shen, S.; Ning, L.; Hui, C. A Novel Deep Learning-Based Spatiotemporal Fusion Method for Combining Satellite Images with Different Resolutions Using a Two-Stream Convolutional Neural Network. *Remote Sens.* **2020**, *12*, 698. [CrossRef]

25. Xue, J.; Leung, Y.; Fung, T. A Bayesian data fusion approach to spatio-temporal fusion of remotely sensed images. *Remote Sens.* **2017**, *9*, 1310. [CrossRef]

26. Moreno-Martínez, Á.; Izquierdo-Verdiguier, E.; Maneta, M.P.; Camps-Valls, G.; Robinson, N.; Muñoz-Marí, J.; Sedano, F.; Clinton, N.; Running, S.W. Multispectral high resolution sensor fusion for smoothing and gap-filling in the cloud. *Remote Sens. Environ.* **2020**, *247*, 111901. [CrossRef]

27. Walker, J.; De Beurs, K.; Wynne, R.; 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. Hobyb, A.I.E.; Radgui, A.; Tamtaoui, A.; Er-Raji, A.; Hadani, D.E.; Merdas, M.; Smiej, F.M. Evaluation of spatiotemporal fusion methods for high resolution daily NDVI prediction. In Proceedings of the 5th International Conference on Multimedia Computing and Systems (ICMCS), Marrakech, Morocco, 29 September–1 October 2016; pp. 121–126.

29. Knauer, K.; Gessner, U.; Fensholt, R.; Kuenzer, C. An ESTARFM Fusion Framework for the Generation of Large-Scale Time Series in Cloud-Prone and Heterogeneous Landscapes. *Remote Sens.* **2016**, *8*, 425. [CrossRef]

30. Li, Q.; Ding, F.; Wu, W.; Chen, J. Improvement of ESTARFM and its application to fusion of Landsat-8 and MODIS Land Surface Temperature images. In Proceedings of the Fourth International Workshop on Earth Observation and Remote Sensing Applications, Guangzhou, China, 4–6 July 2016; pp. 33–37.

31. Sun, F.; Roderick, M.L.; Farquhar, G.D.; Lim, W.H.; Zhang, Y.; Bennett, N.; Roxburgh, S.H. Partitioning the variance between space and time. *Geophys. Res. Lett.* **2010**, *37*, 107. [CrossRef]

32. Li, B.; Chen, Y.; Shi, X.; Chen, Z.; Li, W. Temperature and precipitation changes in different environments in the;arid region of northwest China. *Theor. Appl. Climatol.* **2013**, *112*, 589–596. [CrossRef]

33. Verbesselt, J.; Zeileis, A.; Herold, M. Near real-time disturbance detection using satellite image time series. *Remote Sens. Environ.* **2012**, *123*, 98–108. [CrossRef]

34. Wang, C.; Lei, S.; Elmore, A.J.; Jia, D.; Mu, S. Integrating Temporal Evolution with Cellular Automata for Simulating Land Cover Change. *Remote Sens.* **2019**, *11*, 301. [CrossRef]

35. Verbesselt, J.; Hyndman, R.; Newnham, G.; Culvenor, D. Detecting trend and seasonal changes in satellite image time series. *Remote Sens. Environ.* **2010**, *114*, 106–115. [CrossRef]

36. He, Z.; Du, J.; Zhao, W.; Yang, J.; Chen, L.; Zhu, X.; Chang, X.; Liu, H. Assessing temperature sensitivity of subalpine shrub phenology in semi-arid mountain regions of China. *Agric. For. Meteorol.* **2015**, *213*, 42–52. [CrossRef]

37. Chen, J.; Jönsson, P.; Tamura, M.; Gu, Z.; Matsushita, B.; Eklundh, L. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. *Remote Sens. Environ.* **2004**, *91*, 332–344. [CrossRef]

38. Zhou, J.; Cai, W.; Qin, Y.; Lai, L.; Guan, T.; Zhang, X.; Jiang, L.; Du, H.; Yang, D.; Cong, Z. Alpine vegetation phenology dynamic over 16 years and its covariation with climate in a semi-arid region of China. *Sci. Total Environ.* **2016**, *572*, 119–128. [CrossRef]

39. Ghaderpour, E.; Vujadinovic, T. The Potential of the Least-Squares Spectral and Cross-Wavelet Analyses for Near-Real-Time Disturbance Detection within Unequally Spaced Satellite Image Time Series. *Remote Sens.* **2020**, *12*, 2446. [CrossRef]

40. Chen, G.; Wang, Y. Land Use Changes in Shendong Coal Mining Area. In Proceedings of the International Conference on Multimedia Technology, Ningbo, China, 29–31 October 2010; pp. 1–4.

41. Lei, S.; Ren, L.; Bian, Z. Time–space characterization of vegetation in a semiarid mining area using empirical orthogonal function decomposition of MODIS NDVI time series. *Environ. Earth Sci.* **2016**, *75*, 516. [CrossRef]

42. Liu, Y.; Lei, S.; Cheng, L.; Cheng, W.; Xiong, J.; Bian, Z. Leaf photosynthesis of three typical plant species affected by the subsidence cracks of coal mining: A case study in the semiarid region of Western China. *Photosynthetica* **2019**, *57*, 75–85. [CrossRef]

43. Ma, C.; Cheng, X.; Yang, Y.; Zhang, X.; Guo, Z.; Zou, Y. Investigation on mining subsidence based on multi-temporal InSAR and time-series analysis of the small baseline subset—Case study of working faces 22201-1/2 in Bu'ertai mine, Shendong coalfield, China. *Remote Sens.* **2016**, *8*, 951. [CrossRef]

44. Huang, B.; Song, H. Spatiotemporal Reflectance Fusion via Sparse Representation. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3707–3716. [CrossRef]

45. Masek, J.G.; Vermote, E.F.; Saleous, N.E.; Wolfe, R.; Hall, F.G.; Huemmrich, K.F.; Gao, F.; Kutler, J.; Lim, T.K. A Landsat surface reflectance dataset for North America, 1990–2000. *IEEE Geosci. Remote Sens. Lett.* **2006**, *3*, 68–72. [CrossRef]

46. Vermote, E.; Justice, C.; Claverie, M.; Franch, B. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sens. Environ.* **2016**, *185*, 46–56. [CrossRef]

47. Zhu, Z.; Woodcock, C.E. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sens. Environ.* **2012**, *118*, 83–94. [CrossRef]

48. Li, P.; Jiang, L.; Feng, Z. Cross-Comparison of Vegetation Indices Derived from Landsat-7 Enhanced Thematic Mapper Plus (ETM+) and Landsat-8 Operational Land Imager (OLI) Sensors. *Remote Sens.* **2013**, *6*, 310–329. [CrossRef]

49. Canty, M.J.; Nielsen, A.A. Automatic radiometric normalization of multitemporal satellite imagery with the iteratively re-weighted MAD transformation. *Remote Sens. Environ.* **2008**, *112*, 1025–1036. [CrossRef]

50. Emelyanova, I.V.; Mcvicar, T.R.; Niel, T.G.V.; Li, L.T.; Dijk, A.I.J.M.V. 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]

51. Wang, Z.; Schaaf, C.B.; Sun, Q.; Kim, J.H.; Erb, A.M.; Gao, F.; Román, M.O.; Yang, Y.; Petroy, S.; Taylor, J.R. Monitoring land surface albedo and vegetation dynamics using high spatial and temporal resolution synthetic time series from Landsat and the MODIS BRDF/NBAR/albedo product. *Int. J. Appl. Earth. Obs. Geoinf.* **2017**, *59*, 104–117. [CrossRef]

52. Chen, B.; Huang, B.; Xu, B. Constucting a unified framework for multi-source remotely sensed data fusion. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 2574–2577.

53. Bhandari, S.; Phinn, S.; Gill, T. Preparing Landsat Image Time Series (LITS) for Monitoring Changes in Vegetation Phenology in Queensland, Australia. *Remote Sens.* **2012**, *4*, 1856–1886. [CrossRef]

54. Hmimina, G.; Dufrêne, E.; Pontailler, J.-Y.; Delpierre, N.; Aubinet, M.; Caquet, B.; De Grandcourt, A.; Burban, B.; Flechard, C.; Granier, A. Evaluation of the potential of MODIS satellite data to predict vegetation phenology in different biomes: An investigation using ground-based NDVI measurements. *Remote Sens. Environ.* **2013**, *132*, 145–158. [CrossRef]

55. Eklundh, L.; Jönsson, P. TIMESAT: A software package for time-series processing and assessment of vegetation dynamics. In *Remote Sensing Time Series*; Springer: Cham, Switzerland, 2015; pp. 141–158.

56. Jonsson, P.; Eklundh, L. Seasonality extraction by function fitting to time-series of satellite sensor data. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 1824–1832. [CrossRef]

57. Amorós-López, J.; Gómez-Chova, L.; Alonso, L.; Guanter, L.; Zurita-Milla, R.; Moreno, J.; Camps-Valls, G. Multitemporal fusion of Landsat/TM and ENVISAT/MERIS for crop monitoring. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *23*, 132–141. [CrossRef]

58. Liao, L.; Song, J.; Wang, J.; Xiao, Z.; Wang, J. Bayesian Method for Building Frequent Landsat-Like NDVI Datasets by Integrating MODIS and Landsat NDVI. *Remote Sens.* **2016**, *8*, 452. [CrossRef]

59. Gevaert, C.M.; García-Haro, F.J. A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion. *Remote Sens. Environ.* **2015**, *156*, 34–44. [CrossRef]

60. 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.* **2018**, *10*, 1388. [CrossRef]

61. Qu, Y.; Zhang, Y.; Wang, J. A dynamic Bayesian network data fusion algorithm for estimating leaf area index using time-series data from in situ measurement to remote sensing observations. *Int. J. Remote Sens.* **2012**, *33*, 1106–1125. [CrossRef]

62. Chen, Y.; Cao, R.; Chen, J.; Zhu, X.; Zhou, J.; Wang, G.; Shen, M.; Chen, X.; Yang, W. A New Cross-Fusion Method to Automatically Determine the Optimal Input Image Pairs for NDVI Spatiotemporal Data Fusion. *IEEE Geosci. Remote. Sens. Lett.* **2020**. [CrossRef]

63. 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]

64. Zhao, M.; Running, S.W. Drought-Induced Reduction in Global Terrestrial Net Primary Production from 2000 Through 2009. *Science* **2010**, *329*, 940–943. [CrossRef]

65. 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]

66. Wang, Z.; Bovik, A.C. A universal image quality index. *IEEE Signal Process. Lett.* **2002**, *9*, 81–84. [CrossRef]

67. Liu, X.; Deng, C.; Chanussot, J.; Hong, D.; Zhao, B. Stfnet: A two-stream convolutional neural network for spatiotemporal image fusion. *IEEE Geosci. Remote. Sens. Lett.* **2019**, *57*, 6552–6564. [CrossRef]

68. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]