

Article

# A Spatio-Temporal Data Fusion Model for Generating NDVI Time Series in Heterogeneous Regions

Chunhua Liao <sup>1,\*</sup> , Jinfei Wang <sup>1</sup>, Ian Pritchard <sup>1</sup>, Jianguai Liu <sup>2</sup> and Jiali Shang <sup>2</sup>

<sup>1</sup> Department of Geography, Western University, London, ON N6A 5C2, Canada; jfwang@uwo.ca (J.W.); ianmpritchard@gmail.com (I.P.)

<sup>2</sup> Science and Technology Branch, Agriculture and Agri-Food Canada, 960 Carling Ave., Ottawa, ON K1A 0C6, Canada; Jianguai.Liu@agr.gc.ca (J.L.); Jiali.Shang@agr.gc.ca (J.S.)

\* Correspondence: xnliochunhua@gmail.com; Tel.: +1-519-661-2111 (ext. 80269)

Received: 2 September 2017; Accepted: 2 November 2017; Published: 4 November 2017

**Abstract:** Time series vegetation indices with high spatial resolution and high temporal frequency are important for crop growth monitoring and management. However, due to technical constraints and cloud contamination, it is difficult to obtain such datasets. In this study, a spatio-temporal vegetation index image fusion model (STVIFM) was developed to generate high spatial resolution Normalized Difference Vegetation Index (NDVI) time-series images with higher accuracy, since most of the existing methods have some limitations in accurately predicting NDVI in heterogeneous regions, or rely on very computationally intensive steps and land cover maps for heterogeneous regions. The STVIFM aims to predict the fine-resolution NDVI through understanding the contribution of each fine-resolution pixel to the total NDVI change, which was calculated from the coarse-resolution images acquired on two dates. On the one hand, it considers the difference in relationships between the fine- and coarse-resolution images on different dates and the difference in NDVI change rates at different growing stages. On the other hand, it neither needs to search similar pixels nor needs to use land cover maps. The Landsat-8 and MODIS data acquired over three test sites with different landscapes were used to test the spatial and temporal performance of the proposed model. Compared with the spatial and temporal adaptive reflectance fusion model (STARFM), enhanced spatial and temporal adaptive reflectance fusion model (ESTARFM) and the flexible spatiotemporal data fusion (FSDAF) method, the proposed STVIFM outperforms the STARFM and ESTARFM at three study sites and different stages when the land cover or NDVI changes were captured by the two pairs of fine- and coarse-resolution images, and it is more robust and less computationally intensive than the FSDAF.

**Keywords:** spatio-temporal; data fusion; Landsat; MODIS; NDVI; time-series

## 1. Introduction

The Normalized Difference Vegetation Index (NDVI) is a widely used vegetation index (VI) derived from optical remote-sensing data to evaluate the biophysical or biochemical information related to vegetation growth [1–5]. Large scale time series NDVI is generally used for assessment and monitoring of forest [6,7], grassland [8–10], ecological environment [11,12], wildlife habitat disturbance [13], and to estimate gross primary productivity [14], biomass [15,16], and evapotranspiration [17]. It can be calculated from the images acquired by sensors such as the Advanced Very High Resolution Radiometer (AVHRR), Moderate Resolution Imaging Spectroradiometer (MODIS), Medium Resolution Imaging Spectrometer (MERIS), Sea-Viewing Wide Field-of-View Sensor (SEAWIFS), or VEGETATION, with spatial resolution ranging from 250 m to a few kilometers [1]. In applications such as crop monitoring, time series images acquired by high spatial-resolution sensors such as Landsat-OLI (30 m)

and RapidEye (5 m) are required to provide spatial and temporal details. However, due to cost and technical limitations (trade-off between pixel spatial resolution and satellite temporal revisiting cycle) [18] and cloud cover problems, it is difficult to acquire images with both high spatial resolution and high temporal frequency. Thus, spatio-temporal data fusion techniques have been developed as a feasible and less expensive way to acquire remote sensing time series data for land surface dynamics monitoring [19–22].

In general, generating fine-resolution NDVI time series images through spatio-temporal data fusion can be conducted in two ways [23]: (i) conduct fusion of reflectance first and then calculate the NDVI (Blend-then-Index, BI); and (ii) calculate the NDVI first and then conduct fusion (Index-then-Blend, IB). The theoretical basis of the two ways are essentially the same, and some of the methods developed for reflectance images can also be used for NDVI images. However, the IB is preferred over BI due to less error propagation and fewer computational steps (blending one index band versus multiple reflectance bands) [23].

A few categories of spatio-temporal data fusion approaches have been originally proposed to blend reflectance images including data-assimilation based algorithms, unmixing based methods, dictionary-pair learning based methods, and weighted function based methods [22]. The data-assimilation based algorithms incorporate observation data with models and their uncertainties to minimize the residual errors [24]. The advantage of data-assimilation based algorithms is that a complete time series of fine-resolution images, rather than single synthetic image, can be synthesized in the implementation. A recursive Kalman filter algorithm (KF) was implemented to produce frequent fine-resolution NDVI time series using available fine-resolution images and a time series of coarse-resolution NDVI images in previous studies [25,26]. The uncertainty of these algorithms is correlated to the number of available fine-resolution observations, and these algorithms suffer large uncertainties when the available fine-resolution images are limited in heterogeneous areas.

The unmixing based methods include the Multisensor Multiresolution Technique (MMT) [27], the spatial temporal data fusion approach (STDFA) [28], and the spatial and temporal reflectance unmixing model (STRUM) [29]. These methods assume that remote sensing signal of coarse-resolution pixels is the weighted average of the mean reflectance of each class and a residual error. The weight of each class is its fractional cover within one coarse-resolution pixel, which can be calculated from a fine-resolution land-cover map. The mean reflectance of each land cover is estimated by solving a number of spectral unmixing model equations of coarse-resolution pixels, and the mean reflectance of each land cover is assigned to the fine-resolution pixels. The unmixing based concept can also be used for spatio-temporal fusion of NDVI images. Busetto et al. [1] proposed an approach for the estimation of sub-pixel NDVI time series by combining fine- and coarse-resolution NDVI images based on an unmixing approach. This approach aims to estimate the mean NDVI of each land cover within one coarser-resolution pixel from daily MODIS data by solving the linear weighted equations using manually selected coarse-resolution pixels, and disaggregate MODIS NDVI using weights calculated by Gaussian functions of MODIS spectral dissimilarity index and the Euclidean spatial distance between the target and each pixel. Rao et al. [30] proposed the NDVI Linear Mixing Growth Model (NDVI-LMGM) to produce high spatial resolution NDVI time-series data by using MODIS NDVI time series data and Landsat images. The NDVI-LMGM combines the NDVI linear mixing model with the NDVI linear growth model. It assumes that the short-term changes in NDVI can be interpreted as linear, and long-term NDVI changes can only be predicted reliably by several short-term predictions. The change rate of each land cover during the two dates on the TM/ETM+ pixel scale within the corresponding MODIS pixel can be estimated by solving the linear weighted equations. Another method, NDVI-Bayesian spatiotemporal fusion model (NDVI-BSFM) [31], employed the Bayesian pixel unmixing process through incorporating the prior multi-year average MODIS NDVI from the first day of the year to the last day of the year for each land cover type, and it can predict a single Landsat-like NDVI image as well as build a Landsat-like NDVI time-series dataset. However, these methods are computationally intensive and rely heavily on the quality of MODIS NDVI data. Furthermore,

the performance of these methods will be affected by the classification accuracies [32], especially when there are land cover changes within a year. These methods are not applicable when reliable Land cover/Land use (LC/LU) ancillary information is not available. For example, in some cropland area, the crop types change every year due to the annual rotation of the crops. Thus, the LC/LU data are generally produced at the end of the growing season.

The dictionary-pair learning based methods, such as the sparse representation-based spatio-temporal reflectance fusion model (SPSTFM) [33,34] and the error-bound-regularized semi-coupled dictionary learning (EBSCDL) [35], need one or two pairs of fine- and coarse-resolution images and one coarse-resolution image as input data. It builds relationships (dictionaries) between the features of fine-resolution images and their corresponding coarse-resolution images through dictionary-pair learning [36,37], and then applies the relationship (dictionary) to predict a fine-resolution image on the prediction date. The dictionary-pair learning based methods performed well in phenology change, but they still face challenges in heterogeneous regions with abrupt land cover type changes [22], and they are computationally complex [22,33].

The weighted function based methods include the spatial and temporal adaptive reflectance fusion model (STARFM) [19], and the enhanced spatial and temporal adaptive reflectance fusion model (ESTARFM) [18]. Due to simple input requirements (no ancillary land cover data or classification data required) and robustness, they are widely used in many applications [11,16,23,38,39]. The input data of the STARFM algorithm are one or two pairs of fine- and coarse-resolution images acquired at the same time and one coarse-resolution image acquired at the prediction time. This algorithm assumes that the reflectance of a given coarse-resolution pixel can be aggregated from fine-resolution homogeneous pixels. The STARFM attempts to search neighboring similar homogeneous coarse-resolution pixels within a moving window and endows the weights of these similar pixels calculated according to the spectral difference between coarse-resolution and fine-resolution data, the temporal difference between the input MODIS data, and the distance to the central pixel. The reflectance of the central fine-resolution pixel is estimated from the neighboring similar homogeneous pixels. However, the STARFM algorithm is not applicable to heterogeneous areas such as croplands [18]. Due to this limitation, an improved STARFM with help of unmixing-based method (USTARFM) was proposed [40]. However, it still subjects the problem of land cover changes. The ESTARFM was proposed by Zhu et al. [18] to enhance the ability of STARFM through the use of two pairs of fine-resolution and coarse-resolution images obtained on two dates. It is able to minimize the system biases between the sensors and is more suitable for heterogeneous areas by using two pairs of fine- and coarse-resolution images to detect land cover changes and it keeps more spatial details [18]. However, the ESTARFM method assumes that there is no temporal variation in change rate of the vegetation during a period, which is not reasonable if it is long a period between the input images. Furthermore, the ESTARFM neglects the variation of the relationship between the fine- and coarse-resolution images caused by different acquisition dates. The spatial and temporal nonlocal filter-based data fusion method (STNLFFM) [41] is a recently proposed method that can predict the fine-resolution reflectance accurately and robustly by introducing the idea of nonlocal filtering, for both heterogeneous landscapes and temporally dynamic areas. However, it is based on the assumption that the reflectance change rate is linear, which is not accurate over a long period. The flexible spatiotemporal data fusion (FSDAF) model [22] is a method using one pair of fine- and coarse-resolution images and one coarse-resolution image acquired on the prediction date. This method integrates the unmixing-based methods, spatial interpolation, and STARFM into one framework. It performs better in predicting abrupt land cover changes than other methods that use only one pair of fine- and coarse-resolution images. However, this method is computationally expensive and the prediction accuracy greatly depends on the extent of land cover changes between the two dates of the input images.

In this study, a spatio-temporal vegetation index image fusion model (STVIFM) is proposed to generate NDVI time series images with high spatial and high temporal resolution in heterogeneous regions such as croplands more accurately and robustly. Different from the methods mentioned above,

we aim to predict the NDVI change for each fine-resolution pixel by using a weighting system to disaggregate the total NDVI change within a moving window, which can be calculated from the coarse-resolution NDVI images acquired on two different dates. The proposed model employed: (1) a new weighting system; (2) a new method to obtain the relationship between the two resolution images; and (3) more reasonable assumptions on the NDVI change rate of non-evergreen vegetation, which considers the change rate variations at both spatial scale and temporal scale. This algorithm is tested by Landsat-OLI and MODIS data acquired in three study sites. The results generated by the STARFM, ESTARFM and FSDAF methods are validated by the real Landsat NDVI images or field measurements for three study sites and different growth stages of a cropland area.

## 2. Methodology

### 2.1. Theoretical Basis

Most of the spatio-temporal data fusion methods are based on the linear mixture model, which assumes that the reflectance of a coarse-resolution pixel (mixed pixel) can be modeled as the sum of the reflectance of different land cover endmembers (pure pixels) within the pixel, weighted by the corresponding fractional cover of each component [1,42]. This assumption can also be used for NDVI values, and it was demonstrated that this linear assumption for NDVI only led to minor inaccuracies [43]. The linear mixture model can be written as in Equation (1):

$$\text{NDVI} = \sum_{c=1}^k (f_c \times \text{NDVI}_c) + \varepsilon \quad (1)$$

where  $k$  is the number of classes.  $f_c$  represents the fractional cover of land cover class  $c$  in this pixel and  $\sum_{c=1}^k f_c = 1$ .  $\text{NDVI}_c$  is the NDVI of endmember of land cover class  $c$ .  $\varepsilon$  is the residual error term. This model can be applied to both fine- and coarse-resolution pixels.

The difference in NDVI between a single coarse-resolution pixel and a fine-resolution pixel results from the heterogeneity of the observed area and the systematic biases caused by the difference in sensor systems [18]. For heterogeneous areas, there may be great changes of fine-resolution pixels within one original coarse-resolution pixel, so it is inappropriate to build relationships between the individual fine-resolution pixels and coarse-resolution pixels with a linear regression method. The relationship between NDVI of a pure coarse-resolution pixel ( $\text{NDVI}'_c$ ) and NDVI of a pure fine-resolution pixel ( $\text{NDVI}_c$ ) for a given class  $c$  can be described with a linear model:

$$\text{NDVI}_c = a \times \text{NDVI}'_c + b \quad (c = 1, 2, \dots, k) \quad (2)$$

where  $a$  and  $b$  are the coefficients of the linear regression model between the coarse- and fine-resolution NDVI of pure pixels.

For fine-resolution pixels contained by a coarse resolution pixel, if we neglect the residual error, NDVI of the  $i$ th fine-resolution pixel can be calculated from Equations (1) and (2):

$$\text{NDVI}_i = \sum_{c=1}^k (f_{ci} \times (a \times \text{NDVI}'_c + b)) = a \times \left( \sum_{c=1}^k f_{ci} \times \text{NDVI}'_c \right) + b \quad (3)$$

The average fractional cover of class  $c$  of all the fine-resolution pixels within one coarse-resolution image is equal to the fractional cover of the coarse-resolution image. Therefore, the average NDVI of fine-resolution pixels can be obtained using Equation (4):

$$\overline{\text{NDVI}} = a \times \text{NDVI}' + b \quad (4)$$

where  $\text{NDVI}'$  is the NDVI of one coarse-resolution pixel. If the coarse-resolution image is resampled to the same spatial resolution as the fine-resolution image using the nearest neighbor method (the value

of each resampled pixel is the same within a coarse-resolution pixel), the NDVI value of each original coarse-resolution pixel is equal to the mean value of the resampled pixels within the original pixel. If there are three pairs of fine- and coarse-resolution images acquired on date  $m$  ( $t_m$ ), date  $n$  ( $t_n$ ), and the prediction date  $p$  ( $t_p$ ), which is between  $t_m$  and  $t_n$  ( $t_m < t_p < t_n$ ), the mean NDVI of the  $N$  fine-resolution image pixels on the three dates ( $\overline{\text{NDVI}}_m, \overline{\text{NDVI}}_n, \overline{\text{NDVI}}_p$ ) have linear relationships with the mean NDVI of the corresponding  $N$  resampled coarse-resolution image pixels ( $\overline{\text{NDVI}}'_m, \overline{\text{NDVI}}'_n, \overline{\text{NDVI}}'_p$ ) respectively. The relationships can be expressed as Equation (5):

$$\overline{\text{NDVI}}_k = a_k \times \overline{\text{NDVI}}'_k + b_k \quad (k = m, p, n). \quad (5)$$

where  $a_m, a_n$  and  $b_m, b_n$  can be obtained through the regression of the two pairs of images, respectively. As there is no available fine-resolution image at  $t_p$ , the correlation cannot be achieved through regression.  $a_p$  and  $b_p$  may be different from  $a_m, b_m$  or  $a_n, b_n$ ; thus, the total NDVI difference between the prediction date ( $t_p$ ) and the date before ( $t_m$ ) or after ( $t_n$ ) of the fine-resolution pixels within the moving window can be obtained by Equation (6):

$$\sum_{i=1}^N (\Delta \text{NDVI}_{pki}) = (\overline{\text{NDVI}}_p - \overline{\text{NDVI}}_k) \times N = (a_p \times \overline{\text{NDVI}}'_p - a_k \times \overline{\text{NDVI}}'_k + b_p - b_k) \times N \quad (k = m, n). \quad (6)$$

To obtain each fine-resolution pixel's NDVI change between  $t_m$  (or  $t_n$ ) and  $t_p$ , a disaggregation weighting system can be adopted to describe the contribution of each pixel to the total NDVI changes calculated from the coarse-resolution pixels within the moving window. Then, the NDVI for each pixel can be obtained from the image acquired at  $t_m$  or the image acquired at  $t_n$  by adding the predicted NDVI change of each fine-resolution pixel (Equation (7)):

$$\text{NDVI}_{pki} = \text{NDVI}_{ki} + W_{ki} \times \sum_{i=1}^N (\Delta \text{NDVI}_{pki}) \quad (k = m, n). \quad (7)$$

Theoretically, the NDVI at  $t_p$  can be predicted using the fine-resolution NDVI at  $t_m$  or  $t_n$ . In heterogeneous regions, local land cover changes may cause large inaccuracies if only one fine-resolution image is used as the base image. To reduce the inaccuracy caused by local land cover changes, the two estimations based on the two fine-resolution images can be combined according to the local similarity between the two coarse-resolution images to obtain a more robust result.

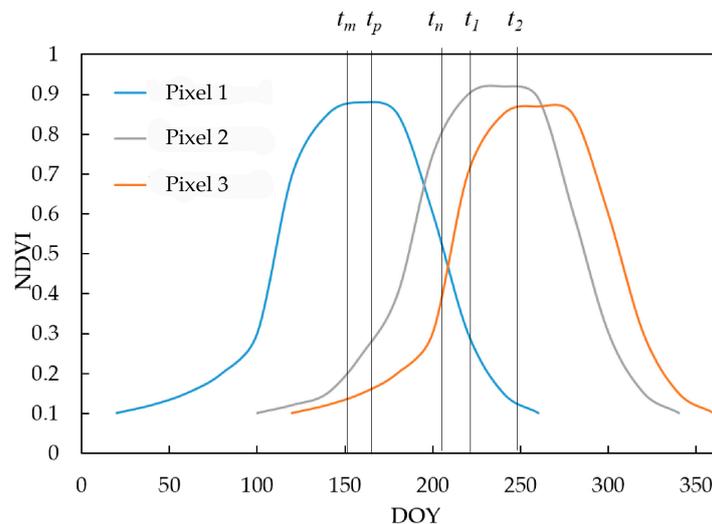
$$\text{NDVI}_{pi} = S_{lpm} \times \text{NDVI}_{pmi} + S_{lpn} \times \text{NDVI}_{pni}. \quad (8)$$

where  $S_{lpm}$  and  $S_{lpn}$  represent the local similarity weights,  $S_{lpm} + S_{lpn} = 1$ . The calculation of  $S_{lpm}$  and  $S_{lpn}$  will be given in Section 2.3.

## 2.2. Weighting System

To predict fine-resolution NDVI at  $t_p$ , the accurate calculation of weights, which aim to disaggregate the total NDVI change to each fine-resolution pixel within a moving window, is an important part of this algorithm. The traditional approach to calculate the NDVI change is to solve a system of linear mixture equations based on a prior classification map. However, this process is time consuming and the results ignored the difference of growth status in the same land cover type. Another approach is to calculate the change rate using the two fine-resolution images, which is adopted in the ESTARFM method [18]. The ESTARFM assumes that the change rate is stable during the period between  $t_m$  and  $t_n$ . This assumption is reasonable if the period between  $t_m$  and  $t_n$  is short enough (e.g., a few days), but, if the period is not short enough (e.g., 20 days), the change rate will vary during this period. In this study, we attempt to estimate the NDVI change based on the two acquired fine-resolution images by addressing the variations of NDVI change rate at spatial and temporal scale.

According to the NDVI time series profile of pure vegetation pixels generated from MODIS NDVI time series data acquired in a growing season shown in previous studies [30,31] or the simulated NDVI time-series profile modeled by a double logistic function [44], the NDVI change rate increases with the increase of NDVI at the beginning of the growing stage, then reaches a maximum, and then decreases at the growing stage. The same trend is shown at the senescent stage. Theoretically, at the early growing stage, the growth rate, which is related to the NDVI change rate of healthy vegetation, is low due to the limited photosynthesis process caused by many factors such as temperature and chlorophyll content [45]. Then, the growth rate increases due to the optimal temperature, increasing chlorophyll content and other factors. At the later growing stage, the growth rate decreases due to the deficiency of nitrogen, water, and the change of temperature, etc. [45]. For heterogeneous regions, the NDVI change rates of different pixels may vary within a moving window. This variation may be caused by the difference of vegetation types or the difference in growth stages of the same vegetation type. Therefore, the spatial variation of fine-resolution NDVI change (spatial weight) varies with the prediction time  $t_p$ . For better understanding the variation of NDVI change rate at the temporal scale and spatial scale, we simulated the NDVI time series profiles for three pure crop pixels within a moving window according to the NDVI time series profile of vegetation pixel generated from MODIS NDVI time series data in previous studies [30,31] (Figure 1).



**Figure 1.** Diagram of simulated Normalized Difference Vegetation Index (NDVI) profiles for different crop pixels.

When image acquired at  $t_m$  is used as the base image, if  $\Delta t_{np}$  ( $t_n - t_p$ ) is short enough, the spatial variation of NDVI change from  $t_m$  to  $t_p$  can be determined by the spatial variation of the fine-resolution NDVI change from  $t_m$  to  $t_n$ . If  $\Delta t_{mp}$  ( $t_p - t_m$ ) is short enough (e.g., one day), the spatial variation of NDVI change from  $t_m$  to  $t_p$  cannot be accurately calculated using the spatial variation of the fine-resolution NDVI change from  $t_m$  to  $t_n$  due to the spatial and temporal variations of NDVI change rate. However, the spatial variation of fine-resolution NDVI change from  $t_m$  to  $t_p$  is closely related to the spatial variation of NDVI change rate at  $t_m$ . When  $t_p$  moves from  $t_m$  to  $t_n$ , the NDVI change between  $t_m$  and  $t_p$  becomes more and more related to the NDVI change between  $t_m$  and  $t_n$ . The idea is the same when image acquired at  $t_n$  is used as the base image. Thus, for any time  $t_p$  between  $t_m$  and  $t_n$ , we propose to calculate the final spatial weight for NDVI changes by combining the NDVI change rate calculated from the image acquired at  $t_m$  or  $t_n$ , and the temporal NDVI change between  $t_m$  and  $t_n$  using a temporal weight. Since two estimations can be calculated from the two base images, and there may be abrupt changes, a more robust final prediction can be achieved by combining the two predictions using a local similarity weight (Equation (8)). However, if there are peaks (growing stage to senescence

stage) or valleys (senescence stage to growing stage) for the vegetation NDVI profile between  $t_m$  and  $t_n$ , and  $t_p$  is neither close to  $t_m$  nor to  $t_n$ , large inaccuracy would be produced. The detailed idea of the weighting system is illustrated in Sections 2.2.1–2.2.3.

### 2.2.1. Weight Calculation Based on Temporal NDVI Change

As mentioned above, if  $t_p$  is very close to  $t_n$ , the variation of NDVI change from  $t_m$  to  $t_p$  can be determined by the variation of the fine-resolution NDVI change from  $t_m$  to  $t_n$  (Equation (9)).

$$D = \text{NDVI}_n - \text{NDVI}_m. \quad (9)$$

In heterogeneous regions, for example when increasing NDVI pixels may be mixed with decreasing NDVI pixels caused by harvesting, flooding or re-planting areas and unchanged NDVI pixels such as bare soil within the moving window, the weight calculation based on temporal NDVI change is complex. To keep the same sign of the weight calculated from temporal NDVI change, the three types of pixels should be processed separately. The land covers were classified into three categories according to the NDVI change from  $t_m$  to  $t_n$  (Equation (10)). Even though it is an unvegetated area such as bare soil, the NDVI may have minor temporal changes due to the slight variation of its spectral characteristics over time. Therefore,  $\pm 0.1$  are selected as the thresholds for the classification. The selection of the thresholds is also supported by the finding that the NDVI threshold for bare soil was 0.1 [46]. The weight of each pixel within the moving window was calculated separately according to their categories.

$$\begin{cases} D_i > 0.1 & \text{Category 1, growing vegetation} \\ D_i < -0.1 & \text{Category 2, senescent vegetation / disturbance} \\ -0.1 \leq D_i \leq 0.1 & \text{Category 3, unchanged area / short term changes} \end{cases}. \quad (10)$$

where  $D_i$  is the NDVI difference of the  $i$ th pixel in the moving window. If  $D_i$  is greater than 0.1 ( $D_i > 0.1$ ), the  $i$ th pixels is marked as growing vegetation. If  $D_i$  is less than  $-0.1$  ( $D_i < -0.1$ ), it is marked as a disturbance or senescent vegetation. If  $D_i$  is less than or equal to  $-0.1$  and greater than or equal to 0.1 ( $-0.1 \leq D_i \leq 0.1$ ), it is regarded as unchanged area or short-term changes that were not recorded in the two fine-resolution images. For Categories 1 and 2, the weight related to the temporal NDVI change between  $t_m$  and  $t_n$  for each pixel can be obtained by and Equation (11).

$$w_{tji} = D_i / \sum_{i=1}^{N_j} D_i \quad (j = 1, 2). \quad (11)$$

where  $w_{tji}$  is the weight calculated from temporal NDVI change for the  $i$ th pixel that belongs to the  $j$ th category.  $N_j$  is the number of the pixels that belong to category  $j$  within the moving window. However, for areas where there is no temporal NDVI change between  $t_m$  and  $t_p$  (Category 3), this calculation is not applicable.

### 2.2.2. Weight Calculation Based on NDVI Change Rate

If  $\Delta t_{mp}$  ( $t_p - t_m$ ) is short enough (e.g., 1 day), the variation of fine-resolution NDVI change from  $t_m$  to  $t_p$  is closely related to the variation of NDVI change rate at  $t_m$ . Under this circumstance, we believe that the NDVI value of the fine-resolution image acquired at any time is related to the NDVI change rate at that time according to Figure 1 and the physiological characteristics of plants described in [45]. For heterogeneous regions, the spatial variation of NDVI values at  $t_m$  causes the spatial variation of the NDVI change rate, and accordingly the NDVI change between  $t_m$  and  $t_p$ . From the simulated NDVI profile shown in Figure 1 or the NDVI profile generated from remote sensing time series images [30,31], it can be assumed that for different types of vegetation or the same type of vegetation with different growth stages, the pixels with low or high NDVI values have lower NDVI change rate and the pixels with median NDVI values should have higher change rate at that time. The spatial variation of the

NDVI change rate can be interpreted using a change rate index (CRI) calculated by an exponential function (Equation (12)) based on one acquired fine-resolution NDVI image.

$$CRI_k = e^{-\frac{(NDVI_k - d)^2}{\sigma^2}} \quad (k = m, n). \quad (12)$$

where  $\sigma^2$  determines the variance of the transformed values, and  $d$  represents the NDVI value where the change rate is maximum. Since the theoretical NDVI values for vegetation pixels range from 0 to 1, the median value 0.5 was selected as the value of  $d$  in this study.  $\sigma^2$  was set to 0.1 to obtain the change rate index with a dynamic range from 0 to 1.

The weight calculated from the NDVI change rate for the  $i$ th pixel that belongs to category  $j$  at  $t_m$  ( $w_{mji}$ ) or  $t_n$  ( $w_{nji}$ ) can be calculated by Equation (13).

$$w_{kji} = \frac{CRI_{ki}}{\sum_{i=1}^{N_j} CRI_{ki}} \quad (k = m, n; j = 1, 2, 3). \quad (13)$$

### 2.2.3. Final Weight Calculation

For heterogeneous vegetated areas, the more similar the land cover on the base date ( $t_m$  or  $t_n$ ) is to the land cover on the prediction date, the more accurate is the predicted image. The time interval can be an indicator for the similarity of land cover but there are exceptions. For example, in Figure 1, the NDVI change is larger from  $t_n$  to  $t_1$  than from  $t_1$  to  $t_2$  for pixel 2 even though  $t_1 - t_n$  is less than  $t_2 - t_1$ . To avoid the manual input of dates for the acquired images, and the circumstance that the NDVI change is larger in a shorter period, the correlation coefficient ( $r_w$ ) (Equation (14)) between two coarse-resolution images for the whole region is selected to calculate the temporal weight ( $T$ ) (Equation (15)). Compared with other statistical indices, the correlation coefficient is more suitable for indicating land cover similarity in heterogeneous regions as it reflects the similarity in the trend of differences associated with each pixel.

$$r_{wpk} = \frac{\sum_{i=1}^N (\overline{NDVI'_{ki}} - \overline{NDVI'_k}) (\overline{NDVI'_{pi}} - \overline{NDVI'_p})}{\sqrt{\sum_{i=1}^N (\overline{NDVI'_{ki}} - \overline{NDVI'_k})^2} \times \sqrt{\sum_{i=1}^N (\overline{NDVI'_{pi}} - \overline{NDVI'_p})^2}} \quad (k = m, n) \quad (14)$$

$$T_{pk} = \frac{r_{wpk}^2}{r_{wpm}^2 + r_{wpn}^2} \quad (k = m, n) \quad (15)$$

where  $r_{wpk}$  is the correlation coefficient of the whole image at  $t_p$  and  $t_k$  ( $k = m, n$ ).  $N$  is the number of pixels in the whole region.  $T_{pk}$  represents the temporal weight between image  $t_p$  and  $t_k$  ( $k = m, n$ ).

For any prediction time between  $t_m$  and  $t_n$ , the final weight for vegetated areas (Category 1,2) can be calculated using the temporal weighted average of the weight calculated from the spatial NDVI variation and the temporal NDVI change (Equation (16)).

$$\begin{cases} W_{mji} = T_{pm} \times w_{mji} + T_{pn} \times w_{tji} \\ W_{nji} = T_{pn} \times w_{nji} + T_{pm} \times w_{tji} \end{cases} \quad (j = 1, 2) \quad (16)$$

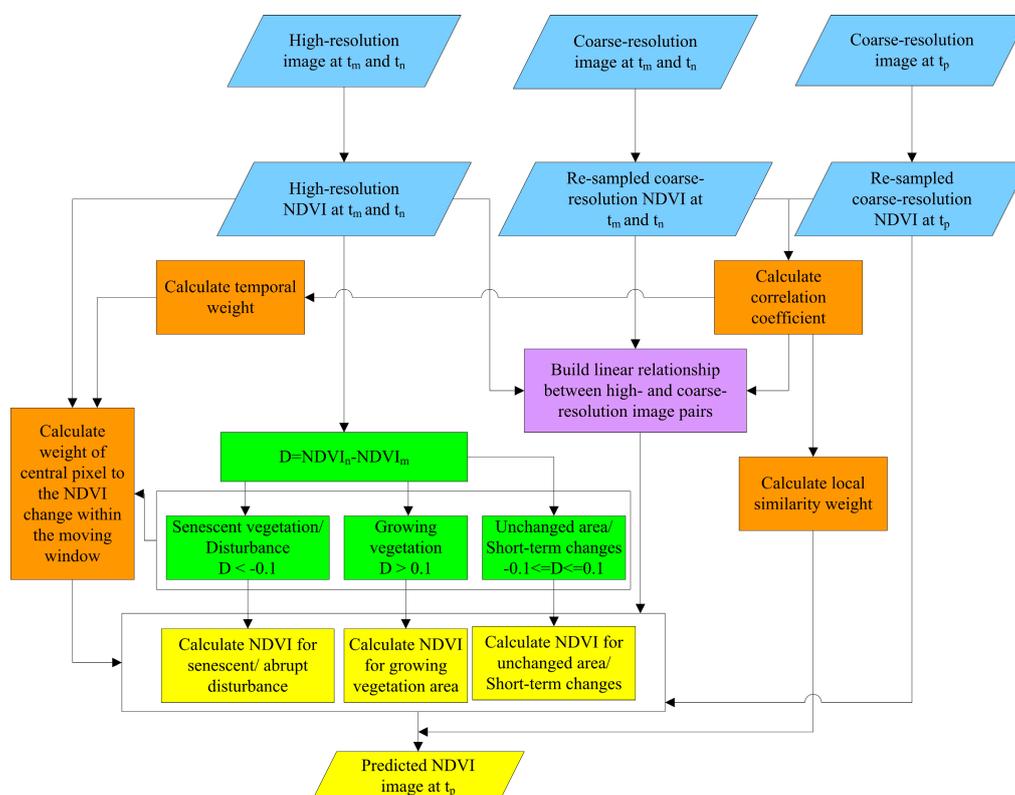
where  $W_{mji}$  or  $W_{nji}$  is the final weight for the  $i$ th pixel calculated based on the fine-resolution image acquired at  $t_m$  or  $t_n$  for the  $j$ th category. For areas where there is no temporal NDVI change (Category 3) between  $t_m$  and  $t_p$ , the final weight is the same as the weight calculated based on the NDVI change rate (Equation (17)).

$$\begin{cases} W_{mji} = w_{mji} \\ W_{nji} = w_{nji} \end{cases} \quad (j = 3) \quad (17)$$

### 2.3. Implementation of the STVIFM

The STVIFM requires two pairs of fine- and coarse-resolution images acquired on the same date and one coarse-resolution image on the prediction date. All of the input images need to be preprocessed (re-projection, geometric correction, and NDVI calculation). A moving window is adopted for implementing the STVIFM. The step of the moving window is one fine-resolution pixel and each step calculates the NDVI of the center pixel in the moving window.

The flowchart for the STVIFM algorithm is shown in Figure 2. There are four main steps for the STVIFM implementation. The first step is to detect NDVI changes according to the two input fine-resolution NDVI images and classify the land cover into three categories. The second step is to calculate the correlation coefficient between the coarse-resolution images for the whole region and within the moving window, and then the weights. The third step is to determine the coefficients  $a$  and  $b$  through a linear regression between the two fine-resolution and coarse-resolution image pairs. The last step is to calculate the final weight and the NDVI value of the center pixel on the prediction date for its category. As the weight calculations have been introduced in Section 2.2, this section mainly illustrates the last two steps.



**Figure 2.** Flowchart of the spatio-temporal vegetation index image fusion model (STVIFM) algorithm. The steps are shaded by different colors (**blue**: satellite data preprocessing; **green**: NDVI change detection; **orange**: weight calculation; **purple**: coefficients determination; **yellow**: NDVI prediction).

#### 2.3.1. Determination of Coefficients between the Fine- and Coarse-Resolution Images

Due to differences in remote sensor systems, systematic biases exist between different sensor imagery. In addition, directional effects and weather conditions can also lead to biases between different sensor images on different dates. In this study, the coefficients between the fine-resolution and coarse-resolution images on different dates were acquired through regression analyses. For the two pairs of fine-resolution and coarse-resolution images, a moving window was used to calculate the mean NDVI of the fine- and coarse-resolution images within this window, then the linear regression

with the mean NDVI values was conducted to obtain the coefficients  $a_m$ ,  $b_m$  and  $a_n$ ,  $b_n$ . The window size should be the odd number which is near to the integer multiple of the ratio between the coarse resolution and fine resolution to accommodate the original coarse-resolution pixel. The step of the moving window is the same as the window size (rather than one-pixel step) so as to avoid self-correlation with the mean values of the NDVI. For example, if the window size is 9 by 9, the step will be 9. It is difficult to determine  $a_p$  and  $b_p$  due to the unavailability of the fine-resolution image at  $t_p$ . The coefficients  $a_p$  and  $b_p$  were calculated by the temporal weighted average of  $a_m$ ,  $a_n$  and  $b_m$ ,  $b_n$  respectively (Equations (18) and (19)), with the assumption that the more similar the two coarse-resolution images, the greater the weight:

$$a_p = T_{pm} \times a_m + T_{pn} \times a_n. \quad (18)$$

$$b_p = T_{pm} \times b_m + T_{pn} \times b_n. \quad (19)$$

### 2.3.2. Local Similarity Weight Calculation and NDVI Prediction for the Central Pixel

There may be local land cover changes caused by harvesting or flooding for heterogeneous regions such as croplands with crops growing in different seasons. The correlation coefficient of the local area within the moving window ( $r_l$ ) is used to calculate the local similarity weight ( $S_l$ ), which is mentioned in Section 2.1 for local heterogeneous area, whereas the mean absolute difference within the moving window ( $MAD_l$ ) (Equation (20)) is used to calculate local similarity weight for local homogeneous area (Equation (21)).

$$MAD_{l_{pk}} = \frac{1}{N_l} \sum_{i=1}^{N_l} abs(NDVI'_{ki} - NDVI'_{pi}) \quad (k = m, n) \quad (20)$$

$$\begin{cases} S_{l_{pk}} = \frac{r_{l_{pk}}^2}{r_{l_{pm}}^2 + r_{l_{pn}}^2} \quad (k = m, n) \text{ heterogeneous area} \\ S_{l_{pm}} = \frac{MAD_{l_{pn}}}{MAD_{l_{pm}} + MAD_{l_{pn}}} \text{ homogeneous area} \\ S_{l_{pn}} = \frac{MAD_{l_{pm}}}{MAD_{l_{pm}} + MAD_{l_{pn}}} \text{ homogeneous area} \end{cases} \quad (21)$$

where  $MAD_{l_{pk}}$  is the local mean absolute difference between coarse-resolution image at  $t_k$  ( $k = m, n$ ) and  $t_p$ .  $N_l$  is the number of pixels within the local moving window.  $S_{l_{pk}}$  is the local similarity weight between coarse-resolution image at  $t_k$  ( $k = m, n$ ) and  $t_p$ . If  $MAD_{l_{pm}} = MAD_{l_{pn}} = 0$ ,  $S_{l_{pm}} = S_{l_{pn}} = 0.5$ .

To determine the heterogeneous area and homogeneous area, the standard deviation (SD) was calculated for all three coarse-resolution images (Equation (22)). If the SD for the three images satisfies Equation (23), the area within the moving window is determined as homogeneous, otherwise heterogeneous.

$$SD_k = \sqrt{\frac{\sum_{i=1}^{N_l} (NDVI'_{ki} - \overline{NDVI'_k})^2}{N_l - 1}} \quad (k = m, n, p) \quad (22)$$

$$SD_m < 0.002 \times NDVI'_{maxm} \text{ and } SD_n < 0.002 \times NDVI'_{maxn} \text{ and } SD_p < 0.002 \times NDVI'_{maxp} \quad (23)$$

where  $N_l$  is the number of pixels in the local moving window, and  $NDVI'_{maxk}$  means the maximum NDVI of the whole coarse image acquired on date  $k$  ( $k = m, n, p$ ).  $NDVI'_{ki}$  and  $\overline{NDVI'_k}$  represent NDVI of the  $i$ th pixel and mean NDVI within the moving window on the coarse-resolution image acquired on date  $k$  ( $k = m, n, p$ ) respectively.

In the moving window, the weight was calculated within each category (Equations (16) and (17)), and the final NDVI for the central pixel on the prediction date can be predicted by Equations (7) and (8) using the pixels which have the same category as the central pixel.

## 3. Results of Data Fusion

### 3.1. Test Sites and Data

Three test sites of distinct geographic locations and climate zones were chosen to test the STVIFM algorithm. The first site (42°53'N 81°35'W, 12 km × 12 km) is located in the Mixedwood Plains Ecozone

in southwestern Ontario, Canada, near the city of London. This area is constantly contaminated by cloud cover during the growing season. For example, from April to October, about 57% of Landsat-8 images contain cloud cover greater than 35%. The dominant crops are winter wheat, corn, and soybean. The winter wheat is generally seeded in October in the previous year and harvested in late July, whereas the corn and soybean are generally seeded in May and harvested in September or October. The second site (37°01'N 99°07'W, 24 km × 24 km) is located near Dodge City in Kansas, United States. This study site contains a large area of grassland as well as crops such as winter wheat. This area receives around 532 mm of rainfall with annual average temperature of 13.3 °C, and altitude between 700 and 800 m above sea level. The third site (44°13'N 87°53'E, 45 km × 45 km) is situated in the south of the Junggar Basin in Xinjiang, China, bordered by the Gurbantunggut Desert in the north. The dominant crop types in this area are cotton, corn, and winter wheat. The cotton and corn are planted in April and harvested in September, whereas wheat is generally seeded in October in the previous year and harvested in May. This area receives scarce rainfall and has long and cold winters with short and hot summer with sharp contrast between daytime and night temperature [47].

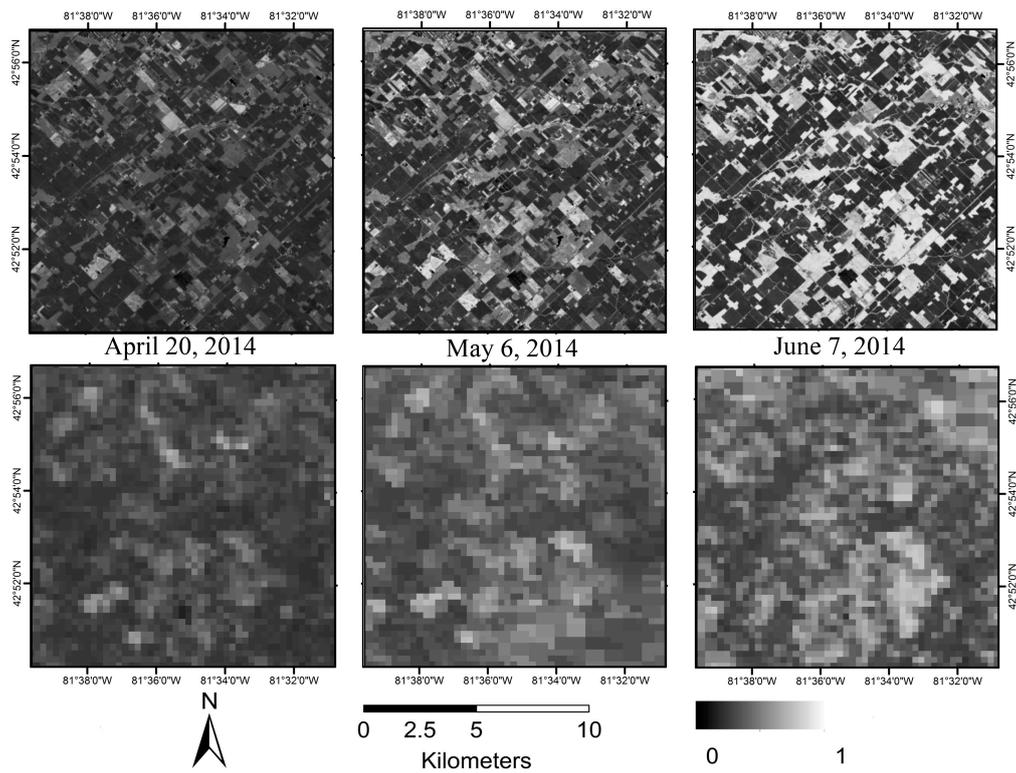
Landsat-8 OLI data and Moderate Resolution Imaging Spectroradiometer (MODIS) surface reflectance products (MOD09GQ and MOD09Q1) were obtained. The Landsat-8 images, with 9 spectral bands, 16-day temporal frequency, and 30 m spatial resolution, were downloaded from the United States Geological Survey (USGS) (<http://landsat.usgs.gov/landsat8.php>). MOD09GQ is daily reflectance product and MOD09Q1 is a level-3 eight-day composite product of MOD09GQ, which provides surface reflectance of Band 1 and Band 2 at 250 m resolution. Each MOD09Q1 pixel represents the best observation during an eight-day period [48]. The eight-day MODIS reflectance products (MOD09Q1) were downloaded from the National Aeronautics and Space Administration (NASA) Reverb portal (<http://reverb.echo.nasa.gov/reverb/>) for Ontario site due to the frequent cloud contamination of the daily reflectance products. The daily MODIS reflectance products MOD09GQ were downloaded for Kansas and Xinjiang sites. The MODIS data were re-projected and mosaicked using the MODIS re-projection tool (MRT). They were then resampled to 30 m resolution using the nearest neighbor method and geo-rectified to the corresponding Landsat images. To avoid the influence of clouds, the MODIS and Landsat-8 images were then clipped to the areas where there was no cloud presence. Finally, the NDVI was calculated.

The dates of cloud-free Landsat-8 OLI and the corresponding MODIS images acquired near the dates of the Landsat acquisitions are shown in Table 1. The two pairs of MODIS and Landsat-8 NDVI images acquired before and after the prediction date, and one MODIS NDVI image acquired on the prediction date were used to predict the synthetic Landsat-like NDVI image. The Landsat-8 images acquired on the prediction date at each study site were used to validate the synthetic Landsat-like NDVI images.

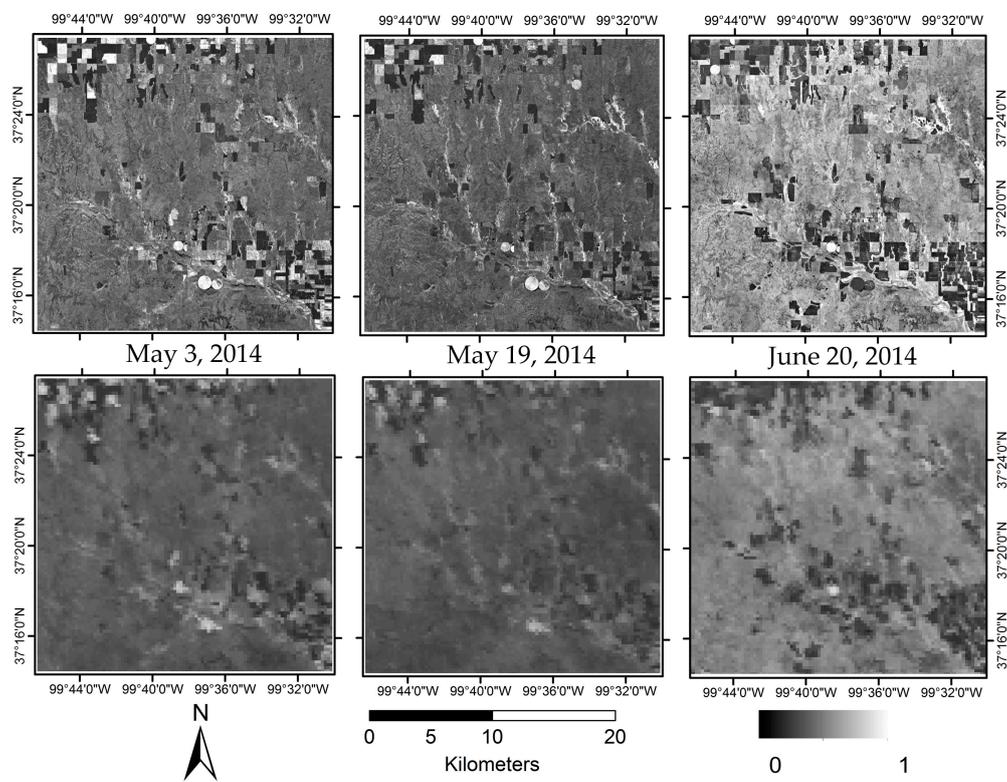
**Table 1.** Dates of MODIS and Landsat-8 OLI images.

	Composite Period (Date) of Obtained MODIS Images	Date of Obtained Landsat OLI Images
Ontario, Canada	15 April 2014–22 April 2014 (DOY: 105–112)	20 April 2014 (DOY: 110)
	1 May 2014–8 May 2014 (DOY: 121–128)	6 May 2014 (DOY: 126) (validation)
	2 June 2014–9 June 2014 (DOY: 153–160)	7 June 2014 (DOY: 158)
Kansas, U.S.	3 May 2014 (DOY: 123)	3 May 2014 (DOY: 123)
	19 May 2014 (DOY: 139)	19 May 2014 (DOY: 139) (validation)
	20 June 2014 (DOY: 171)	20 June 2014 (DOY: 171)
Xinjiang, China	27 May 2014 (DOY: 147)	27 May 2014 (DOY: 147)
	12 June 2014 (DOY: 163)	12 June 2014 (DOY: 163) (validation)
	28 June 2014 (DOY: 179)	28 June 2014 (DOY: 179)

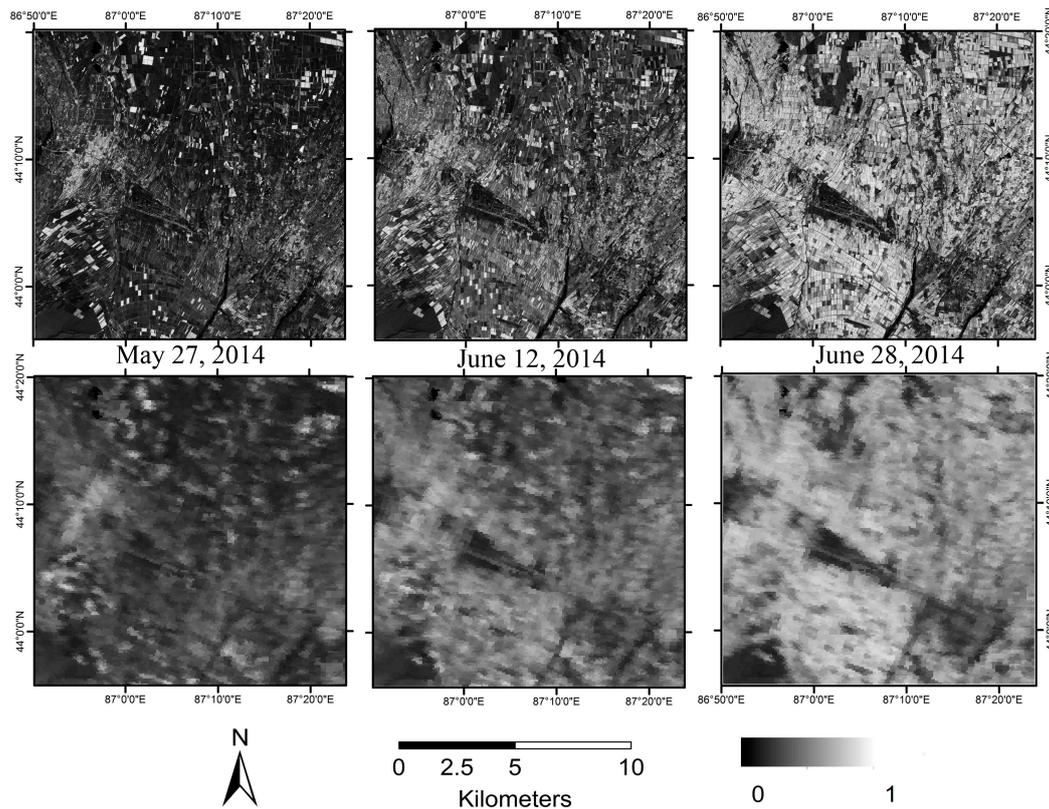
Figures 3–5 show the NDVI images obtained on three dates over three study sites. The sub-images (30 m resolution, 400 × 400 pixels) at the upper rows are Landsat-8 NDVI images (fine-resolution, 30 m) and lower rows are MODIS NDVI images (coarse-resolution, 250 m resampled to 30 m).



**Figure 3.** Landsat (upper row); and MODIS (lower row) NDVI images from three dates in 2014 over a cropland area in Ontario, Canada.

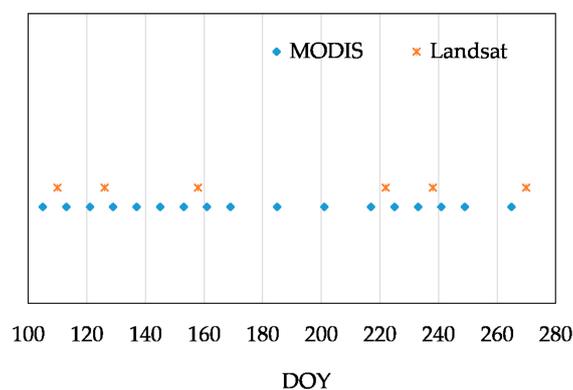


**Figure 4.** Landsat (upper row); and MODIS (lower row) NDVI images from three dates in 2014 over a mixed crop and grassland area in Kansas, U.S.

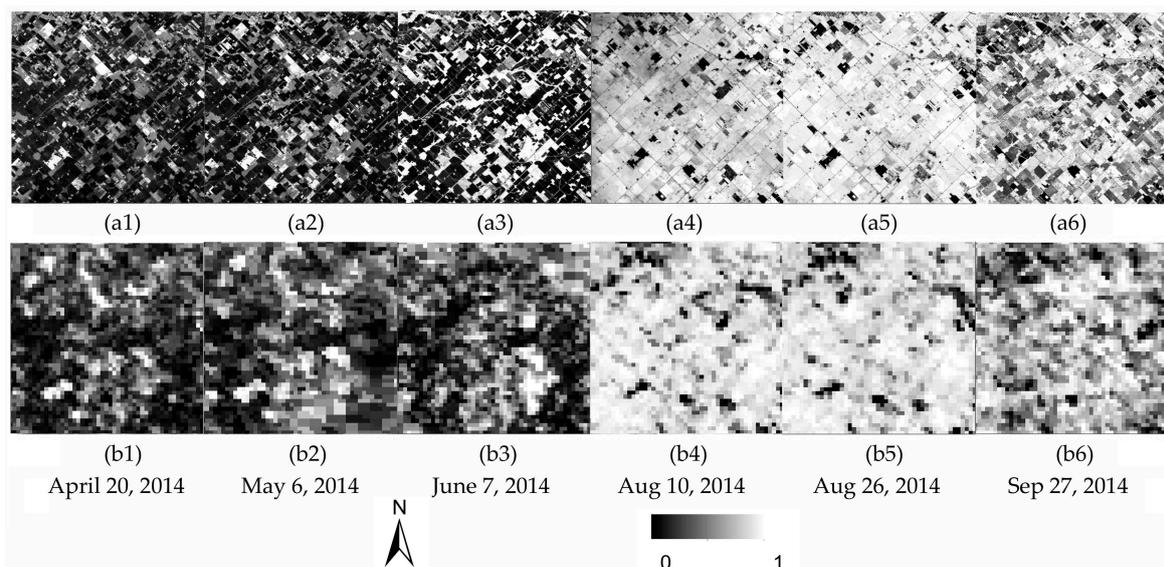


**Figure 5.** Landsat (upper row); and MODIS (lower row) NDVI images from three dates in 2014 over a cropland area in Xinjiang, China.

To assess the application of the proposed algorithm on time series data, a total of 18 cloud-free MODIS MOD09Q1 data and six Landsat-8 OLI data were acquired over London, Ontario throughout the growing season in 2014 (Figure 6). Figure 7 shows the six cloud-free NDVI images and corresponding MODIS NDVI images. The sub-images at the upper row are Landsat-8 NDVI images (fine-resolution, 30 m). The sub-images at the lower row are the eight-day MODIS NDVI images. From 6 May to 7 June, as the winter wheat grew, the NDVI increased greatly. On 10 August, the winter wheat had been harvested, and most of the land was covered by corn and soybean. Thus, there were great land cover changes from 7 June to 10 August. From 10 August to 26 August, a few winter wheat fields were covered by alfalfa, and the NDVI increased again. On 27 September, most corn and soybean were senescent and the NDVI decreased.



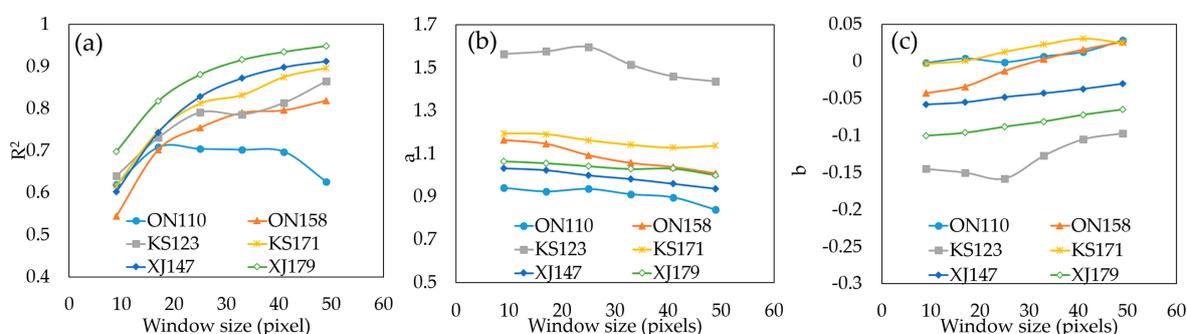
**Figure 6.** The dates for the available cloud free Landsat imagery and the MODIS time series data.



**Figure 7.** The Landsat and MODIS NDVI image pairs acquired throughout the growing season in Ontario, Canada.

### 3.2. Selection of Window Size for Deriving the Coefficients

Linear regression analysis was conducted between the fine- and coarse-resolution image pairs (Section 2.3.1) for different study sites using different sizes of moving window. The variations of the coefficient of determination ( $R^2$ ),  $a$  and  $b$  with the increasing window size are shown in Figure 8. It was illustrated that when 9 (the approximate size of one coarse-resolution pixel) was adopted as the window size, the correlation was much lower than for other choices. The reason for this is likely that there are errors introduced by rounding and the geometric correction process between fine-resolution and coarse-resolution images. In this way, the fine-resolution pixels may not be the pixels that are supposed to be within the original coarse-resolution pixel. With the increase of window size, the  $R^2$  values increased for most image pairs and plateaued when the window size is 17, which contains 4 MODIS pixels. This should be impacted by the MODIS point spread function (SPF) [49]. The value of  $a$  ranges from 0.9 to 1.5 for different study sites, and  $b$  ranges from  $-0.15$  to  $0.05$  for different study sites. Even for the same study site,  $a$  and  $b$  vary on different dates. With the increase of the window size, the variations of  $a$  and  $b$  are slight and smooth. It can be believed that both  $a$  and  $b$  are not sensitive to the change of window size. However, with a larger window, the increase rate of  $R^2$  becomes smaller, the sample points become less, and the significance of the correlation will be reduced. Accordingly, a  $33 \times 33$  moving window ( $4 \times 4$  coarse-resolution pixels) was used to obtain the coefficients.



**Figure 8.** The variations of:  $R^2$  (a);  $a$  (b); and  $b$  (c) with the increasing window size for fine- and coarse-resolution NDVI pairs over different study sites.

Besides the proposed method, the STARFM, ESTARFM and FSDAF methods were also used for comparison purpose. To select a reasonable window size for the algorithm implementation, the data fusion was conducted with different window sizes based on the STARFM, ESTARFM, FSDAF and STVIFM methods. Take the Ontario site as an example (Table 2). For the STVIFM, the accuracy is the highest when the window size is 25, and the computation time increases with the increase of window size. For the ESTARFM, the accuracy is the highest when the window size is 33. For the STARFM and the FSDAF, the larger window size gives higher accuracy (higher  $R^2$  and lower RMSE), but the increase of accuracy is very small. Therefore, we chose 33 as the window size (the same as the window size used in coefficients calculation) for the fusion models, after analyzing the  $R^2$ , RMSE, and computational efficiency of these algorithms using different window sizes.

**Table 2.** Regression results between Landsat-8 NDVI image and correspondent synthetic NDVI image based on different algorithm using different window size at the Ontario site.

Window Size	STARFM			ESTARFM			FSDAF			STVIFM		
	$R^2$	RMSE	Time (s)	$R^2$	RMSE	Time (s)	$R^2$	RMSE	Time (s)	$R^2$	RMSE	Time (s)
9	0.668	0.099	2.08	0.673	0.120	34.67	0.782	0.081	52.04	0.739	0.099	13.6
17	0.661	0.098	4.07	0.704	0.116	40.57	0.804	0.077	55.56	0.824	0.071	20.69
25	0.659	0.096	7.21	0.717	0.113	46.71	0.816	0.076	58.68	0.83	0.070	31.5
33	0.659	0.096	11.52	0.723	0.112	52.2	0.824	0.075	68.6	0.826	0.071	46.4
41	0.659	0.095	16.31	0.722	0.112	62.01	0.828	0.075	67.45	0.818	0.074	64.18
49	0.659	0.094	22.8	0.721	0.113	73.59	0.832	0.074	78.7	0.811	0.076	87.08

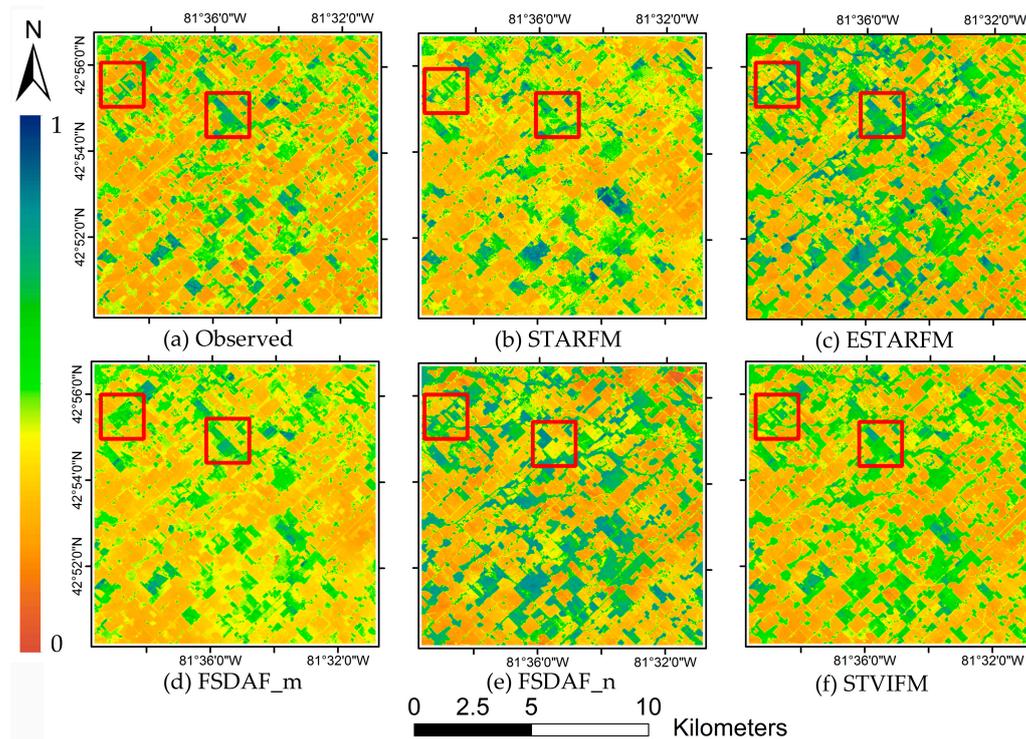
$p < 0.01$ .

### 3.3. Algorithm Tests in Regions with Different Landscapes

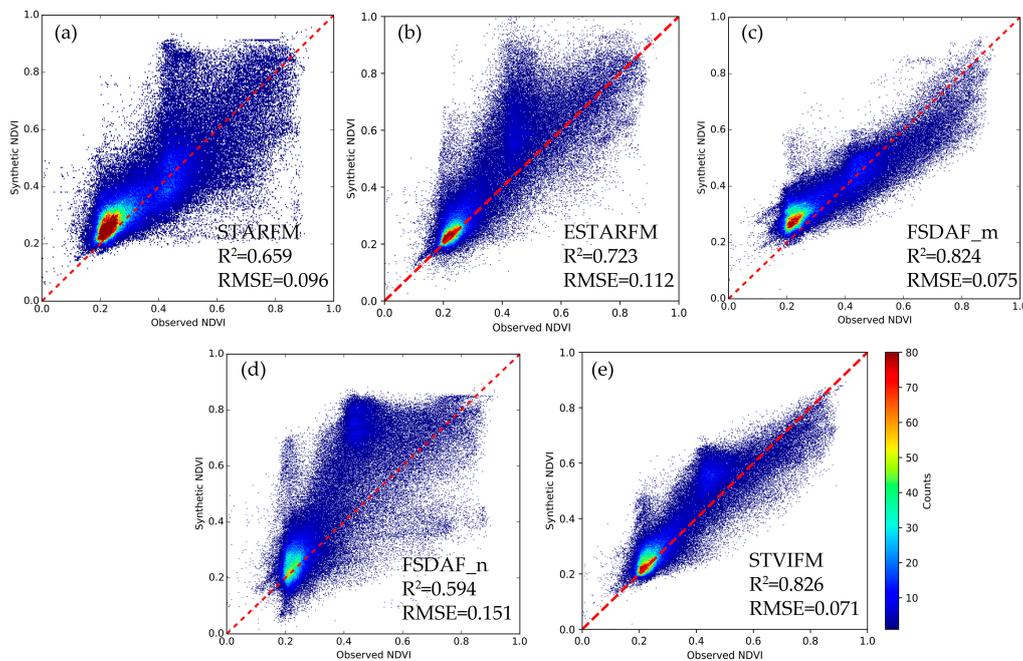
As the FSDAF only needs one pair of fine- and coarse-resolution images, the pair of images acquired on the date before and after the prediction date were used, respectively, for all three sites, and FSDAF\_m and FSDAF\_n were used to represent the two predictions hereafter. The performances of the four algorithms were assessed by visual comparison and regression analyses. The coefficient of determination, Mean Absolute Difference (MAD) and Mean Difference (MD) between the observed NDVI and predicted NDVI images were also calculated to assess the accuracies of the four algorithms.

Figures 9–14 show the original Landsat-8 OLI NDVI image and the synthetic NDVI images predicted by the four algorithms, and the scatter plots between the synthetic and the original NDVI values of Landsat OLI image acquired at  $t_p$  at the three study sites. Table 3 shows the  $R^2$ , RMSE, MAD, MD and computation time of the four algorithms at different test sites. The performance of the STVIFM is better than the STARFM and ESTARFM methods, and similar with the best prediction of the FSDAF at all the three study sites according to  $R^2$ , RMSE, and MAD. In addition, the STVIFM is more computationally efficient than the ESTARFM and FSDAF when a window of  $33 \times 33$  pixels was adopted.

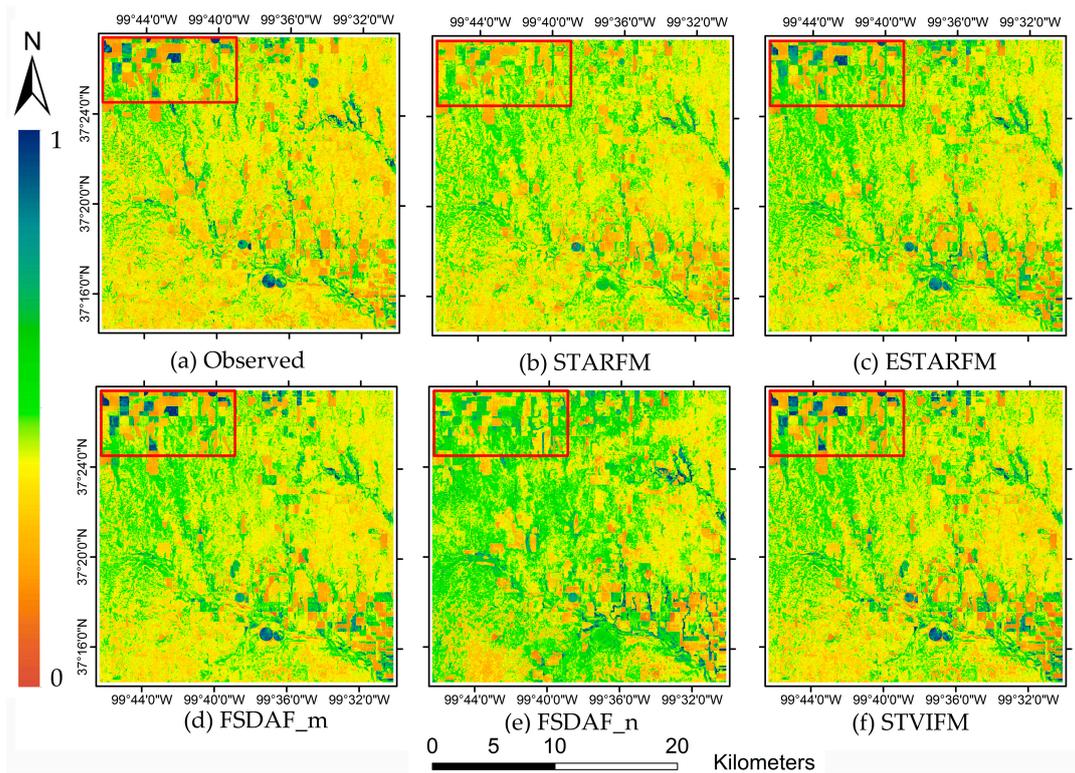
The Ontario site has many small-area croplands (the width of the fields is less than 250 m), which resulted in many mixed pixels in the coarse-resolution image. Figure 10 shows that the predicted NDVI images generated by the FSDAF\_m and STVIFM are more similar to the original image. However, the small-area land cover changes shown in the red boxes in Figure 10 were not accurately predicted by the FSDAF. Actually, the predicted NDVI image obtained by the FSDAF is more similar to the input fine-resolution image. As there are less land cover changes between 20 April and 6 May than between 6 May and 7 June at this site, the predicted result using the FSDAF (FSDAF\_m) is more similar to the observed NDVI. The STARFM, ESTARFM, and STVIFM are all able to predict the land cover change by making use of two image pairs, whereas the ESTARFM overestimated the NDVI.



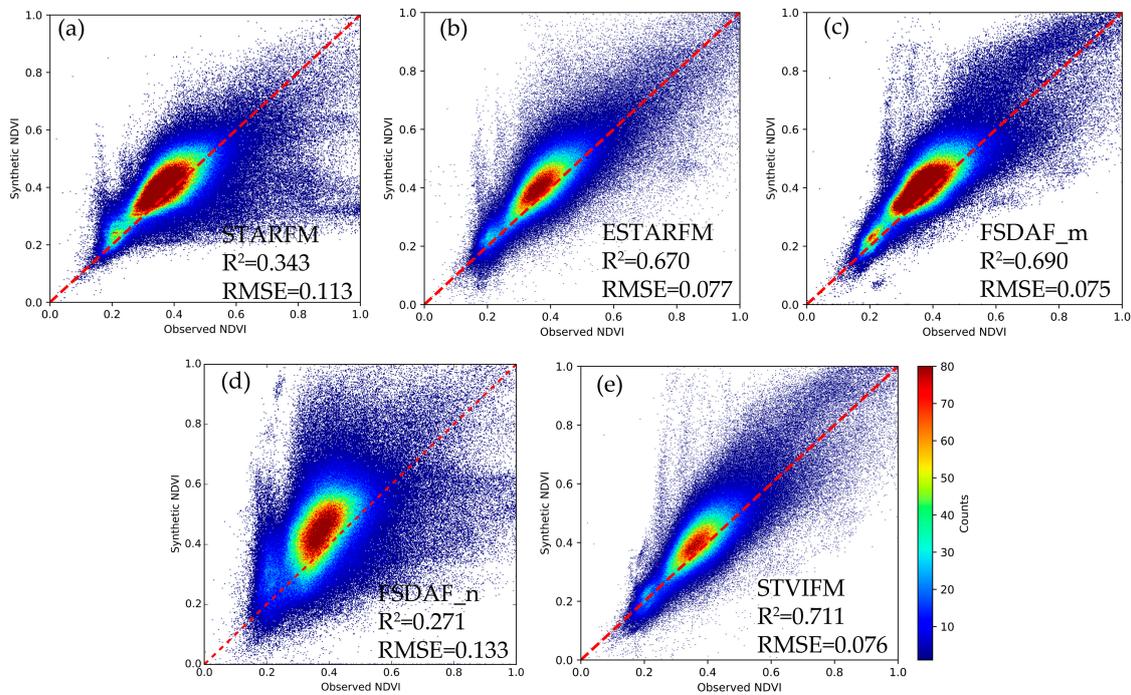
**Figure 9.** Comparison of the observed Landsat image (a); and the synthetic images based on: spatial and temporal adaptive reflectance fusion model (STARFM) (b); enhanced spatial and temporal adaptive reflectance fusion model (ESTARFM) (c); flexible spatiotemporal data fusion (FSDAF) (d,e); and spatio-temporal vegetation index image fusion model (STVIFM) (f), in Ontario, Canada. Red boxes show small-area land cover changes occurred on Landsat-8 images acquired at  $t_p$  (6 May 2014) and  $t_n$  (7 June 2014).



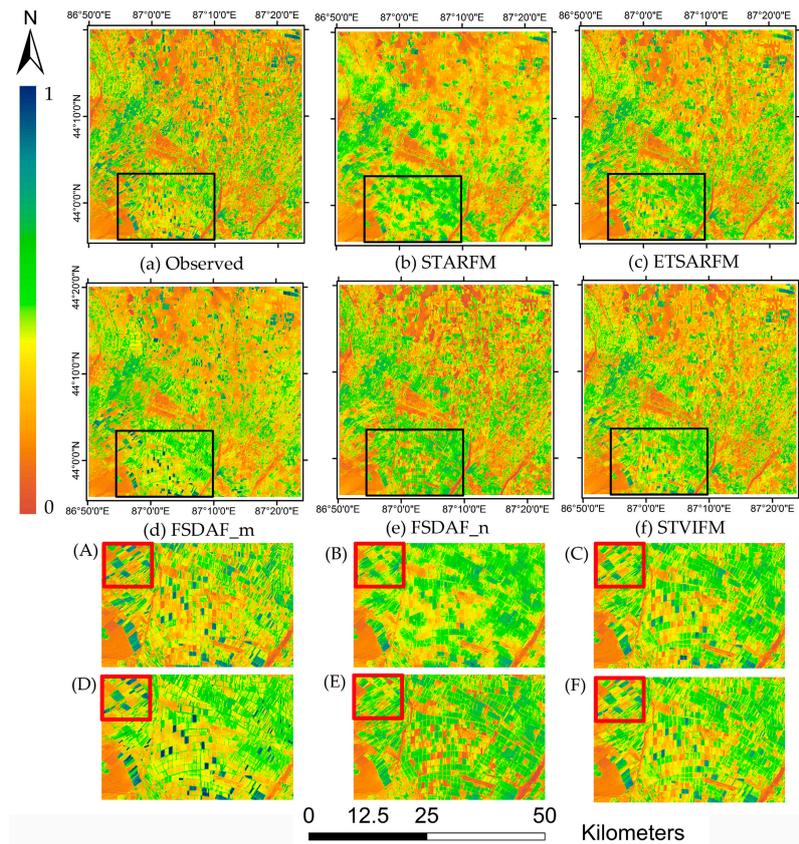
**Figure 10.** Scatter plots of the observed and synthetic Landsat NDVI produced by: STARFM (a); ESTARFM (b); FSDAF\_m (c); FSDAF\_n (d); and STVIFM (e) in Ontario, Canada.



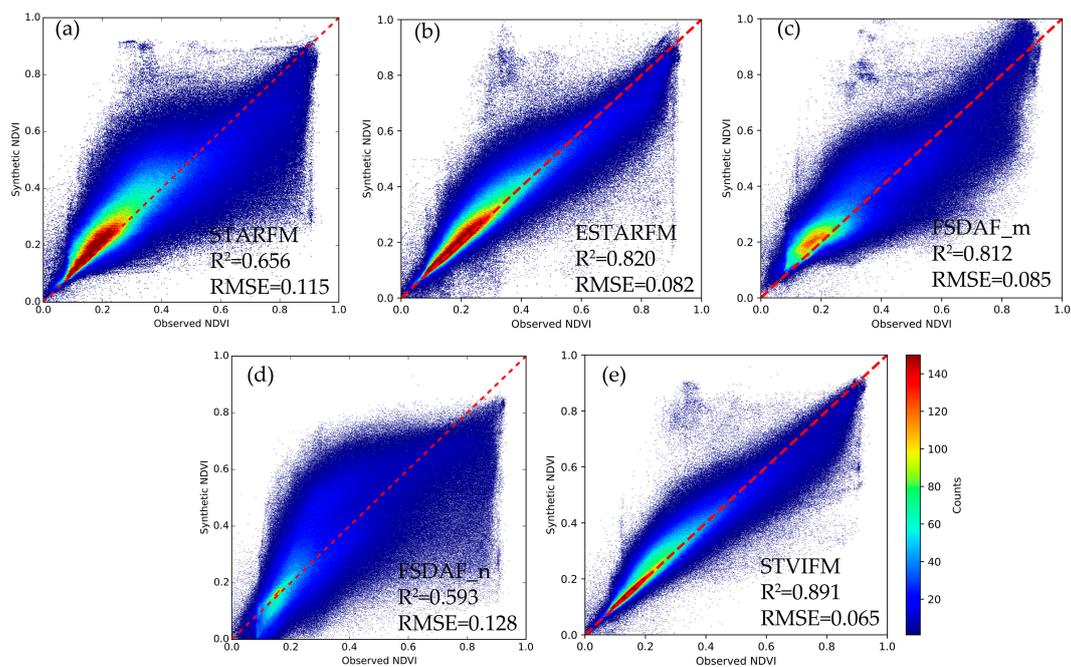
**Figure 11.** Comparison of the observed Landsat image (a); and the synthetic images based on: STARFM (b); ESTARFM (c); FSDAF (d,e); and STVIFM (f), in Kansas, U.S. Red boxes show that the harvesting of crops appeared in Landsat-8 images acquired at  $t_n$  (20 June 2014) are accurately predicted in the synthetic NDVI image based on the STVIFM and ESTARFM.



**Figure 12.** Scatter plots of observed and synthetic Landsat NDVI by: STARFM (a); ESTARFM (b); FSDAF\_m (c); FSDAF\_n (d); and STVIFM (e) in Kansas, U.S.



**Figure 13.** Comparison of the observed Landsat image (a); and the synthetic images based on: STARFM (b); ESTARFM (c); FSDAF\_m (d); FSDAF\_n (e); and STVIFM (f) in Xinjiang, China. (A–F) Zoom-in images shown in the black boxes on the original NDVI and the all the results generated by the four methods. Red boxes show the senescent fields.



**Figure 14.** Scatter plot of observed and synthetic Landsat NDVI by: STARFM (a); ESTARFM (b); FSDAF\_m (c); FSDAF\_n (d); and STVIFM (e) in Xinjiang, China.

**Table 3.** Statistical parameters of the linear regression analysis between synthetic and original Landsat NDVI image.

Study Site	Image Size	Methods	R <sup>2</sup>	RMSE	MAD	MD	Time
Ontario, Canada	400 × 400 (12 km × 12 km)	STARFM	0.659	0.096	0.066	0.020	12 s
		ESTARFM	0.723	0.112	0.075	0.052	52 s
		FSDAF_m	0.824	0.075	0.061	0.026	69 s
		FSDAF_n	0.594	0.151	0.109	0.072	65 s
		STVIFM	0.826	0.071	0.052	0.026	46 s
Kansas, United States	800 × 800 (24 km × 24 km)	STARFM	0.343	0.113	0.073	0.005	40 s
		ESTARFM	0.67	0.077	0.057	0.031	3 min 25 s
		FSDAF_m	0.69	0.075	0.056	0.032	9 min 21 s
		FSDAF_n	0.271	0.133	0.102	0.059	9 min
		STVIFM	0.711	0.076	0.055	0.028	3 min 9 s
Xinjiang, China	1500 × 1500 (45 km × 45 km)	STARFM	0.656	0.115	0.072	−0.02	2 min 42 s
		ESTARFM	0.82	0.082	0.048	0	12 min 29 s
		FSDAF_m	0.812	0.085	0.060	0.010	41 min 57 s
		FSDAF_n	0.593	0.128	0.095	0.001	40 min 16 s
		STVIFM	0.891	0.065	0.045	0.006	11 min 15 s

$p < 0.01$ .

As indicated by the scatter plots shown in Figure 11 and the assessment indices shown in Table 3, the accuracy of the predicted NDVI image using the STVIFM is the best ( $R^2$ : 0.826, RMSE: 0.071) and slightly better than the accuracy of the predicted NDVI image using FSDAF\_m ( $R^2$ : 0.824, RMSE: 0.075). The STARFM performed better than the ESTARFM and FSDAF\_n in terms of the RMSE and MAD. Most NDVI values were overestimated by the ESTARFM and FSDAF\_n. From the perspective of computational efficiency, the STVIFM consumed less time than the ESTARFM and the FSDAF. For the sub-image of 400 × 400 pixels, the ESTARFM and FSDAF consumed 52 s and 69 s, respectively, and the STVIFM consumed about 46 s.

The Kansas site is mostly covered by grassland and it is more homogeneous than the other two sites. The overall accuracy using the proposed STVIFM method ( $R^2$ : 0.711, RMSE: 0.076) is the best when compared with the STARFM and ESTARFM. The RMSE and MAD of the result produced by the STVIFM are comparable with that of the result produced by the FSDAF\_m, but the  $R^2$  of the former result is higher than the later one. Figure 12 reveals that the ESTARFM, FSDAF\_m, and STVIFM performed better in the cropland area shown in the red box, while all the methods seem to overestimate the NDVI in the grassland area. The RMSEs for these three methods are similar (RMSE: 0.077 vs. 0.075 vs. 0.076) but the  $R^2$  for the STVIFM is the highest ( $R^2$ : 0.711). As the land cover for most areas barely changed from the date of the first image pair (3 May 2014) to the prediction date (19 May 2014), the NDVI predicted by the FSDAF\_m shows higher accuracy than the NDVI predicted by the FSDAF\_n. In terms of computational efficiency, the ESTARFM and FSDAF consumed 3 min 29 s and 9 min, respectively, whereas the STVIFM consumed 3 min 9 s for an image of 800 × 800 pixels.

At the Xinjiang site, the predicted NDVI image produced by the STVIFM shows good agreement with the observed NDVI ( $R^2$ : 0.891, RMSE: 0.065). Most of the fields were in the growing stage during this period and a few fields were in the senescent stage. The accuracy of the synthetic NDVI images produced by the ESTARFM and FSDAF\_m are comparable but lower than that of the STVIFM ( $R^2$ : 0.820 vs. 0.812, RMSE: 0.082 vs. 0.085). As shown in the zoom-in area in the black box, the predicted NDVI image obtained by the STARFM shows “blurred” field boundaries, and the overall accuracy is much lower than the ESTARFM. The temporal NDVI changes of the senescent fields shown in the red box in Figure 14 were more accurately captured using the STVIFM when compared with the STARFM and ESTARFM. For the FSDAF, the predicted NDVI image is more accurate using the image pair acquired on 27 May than using the image pair acquired on 28 June. Even though the time intervals between the prediction date (12 June 2014) and dates of the two base image pairs were the same, the land cover changed significantly from 12 June to 28 June. Therefore, the NDVI prediction using the image pair acquired on 27 May is more accurate than using the image pair acquired on 28 June. In terms of the

computational efficiency, the ESTARFM and FSDAF consumed about 12 min and 40 min respectively for an image of  $1500 \times 1500$  pixels, whereas the STVIFM consumed about 11 min.

When we compare the accuracies for the three study sites, it is obvious that both the Ontario site and Xinjiang site are more heterogeneous than the Kansas site and there are many crop fields with small areas. This is possibly the reason that the STARFM and ESTARFM performed better at Kansas site than Ontario site and Xinjiang site in terms of the RMSE and MAD. However, the STVIFM performed better for Ontario site and Xinjiang site than Kansas site. Therefore, it can be concluded that the STVIFM performs better than the STARFM and ESTARFM not only in homogeneous regions but also in heterogeneous regions.

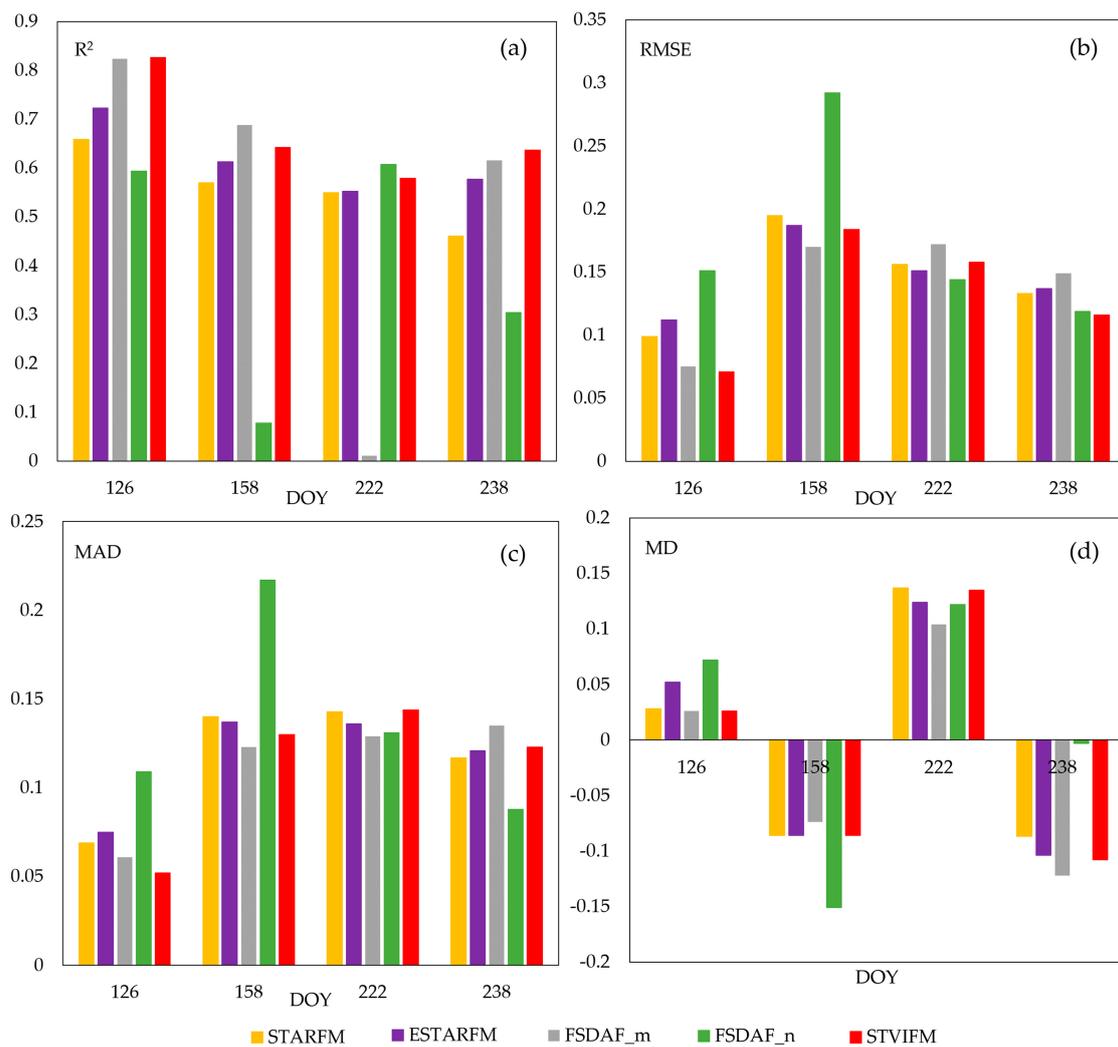
### 3.4. Tests with Time Series Data

In this test, all six available Landsat NDVI images and the corresponding MODIS NDVI images acquired throughout the whole growing season over Ontario site were tested using the four methods for the four predictions shown in Table 4. For the FSDAF, both the image pairs before and after the prediction date were used to implement this method. (a1)–(a6) and (b1)–(b6) are the images shown in Figure 7.

**Table 4.** Images used for the four methods in the four experiments.

Prediction Date (DOY)	Input Images	Images for Validation
121	FSDAF_m: b1, a1, b2 FSDAF_n: b3, a3, b2 STARFM: b1, a1, b3, a3, b2 ESTARFM: b1, a1, b3, a3, b2 STVIFM: b1, a1, b3, a3, b2	a2
158	FSDAF_m: b2, a2, b3 FSDAF_n: b4, a4, b3 STARFM: b2, a3, b4, a4, b3 ESTARFM: b2, a3, b4, a4, b3 STVIFM: b2, a3, b4, a4, b3	a3
222	FSDAF_m: b3, a3, b4 FSDAF_n: b5, a5, b4 STARFM: b3, a3, b5, a5, b4 ESTARFM: b3, a3, b5, a5, b4 STVIFM: b3, a3, b5, a5, b4	a4
238	FSDAF_m: b4, a4, b5 FSDAF_n: b6, a6, b5 STARFM: a4, b4, a6, b6, a5 ESTARFM: a4, b4, a6, b6, a5 STVIFM: a4, b4, a6, b6, a5	a5

The results were validated with the original Landsat NDVI images, and the assessment indices including  $R^2$ , RMSE, MAD, and AD for different methods were shown in Figure 15. For the first prediction, the two Landsat images were acquired on 20 April and 7 June. The main crop was winter wheat, which was growing steadily during this period. There were some small-area land cover changes in images acquired on 6 May and 7 June (Figure 7 (a3)). As presented in Section 3.3 (Ontario site), the STVIFM performed the best when compared with the other three methods during this period.



**Figure 15.** The accuracy of predicted NDVI obtained by different methods on four different dates in Ontario site during the growing season: (a)  $R^2$ ; (b) RMSE; (c) MAD; and (d) MD.

For the second prediction, the two Landsat images were acquired on 6 May and 10 August and the prediction date was 7 June. From 6 May to 7 June, corn and soybeans were planted, then wheat was harvested and alfalfa was planted in the harvested wheat fields from 7 June to 10 August. The STVIFM performed better than the STARFM and ESTARFM (RMSE: 0.184 vs. 0.195 vs. 0.187). The predicted NDVI image generated by the FSDAF using the image pair acquired on 6 May shows much higher accuracy than using the image acquired on 10 August and the predicted NDVI generated by the other two methods.

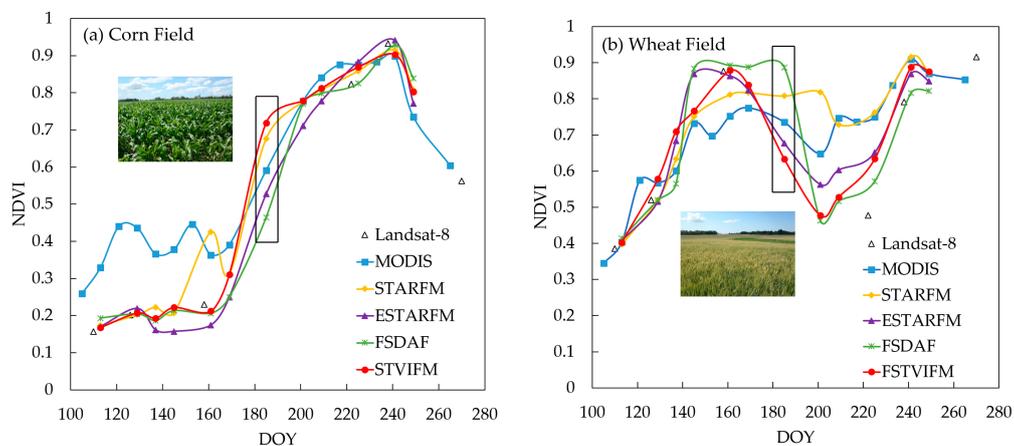
For the third prediction, the two Landsat images were acquired on 6 June and 26 August and the prediction date was 10 August. As mentioned above, the wheat fields changed greatly from 6 June to 10 August, whereas the land cover seldom changed from 10 August to 26 August. The correlation of determination of the predicted NDVI using the STVIFM is higher than the STARFM and ESTARFM ( $R^2$ : 0.579 vs. 0.550 vs. 0.552), whereas the RMSE is higher than the STARFM and ESTARFM (RMSE: 0.158 vs. 0.151 vs. 0.156). This may be because of the large land cover changes for wheat fields from 6 June to 26 August, and the NDVI for wheat field is near the valley of the NDVI profile on 10 August. As mentioned earlier, larger inaccuracy would be produced by the STVIFM if the peak or valley in the NDVI profile between the two dates of acquired Landsat images needs to be predicted, and the NDVI change is not captured by the NDVI difference of the two Landsat images. This inaccuracy can

be reduced if more fine-resolution images can be acquired during this period. The prediction of the FSDAF using the image pair acquired on 26 August shows a higher accuracy than using the image pair acquired on the 6 June.

For the fourth prediction, the two Landsat images were acquired on 10 August and 27 September, and the prediction date was 26 August. Corn and soybeans were senescent during this period. The STVIFM performed better than the STARFM and ESTARFM (RMSE: 0.116 vs. 0.133 vs. 0.137). The accuracy of the image predicted by the FSDAF using image acquired on 27 September is higher than the accuracy of the NDVI predicted by the FSDAF using image acquired on 10 August in terms of the RMSE and MAD.

In addition, a total of 12 Landsat-like NDVI images were predicted using the six acquired Landsat and MODIS image pairs. By using more Landsat images, the prediction accuracy would be improved. However, there was no more available Landsat image to assess the prediction results. The NDVI time series were assessed by analyzing the temporal variations over the cornfield and winter wheat field and compare with the phenology information and photos collected in the field work. For the STARFM, ESTARFM and STVIFM, two temporally closest Landsat and MODIS NDVI image pairs and one MODIS NDVI image acquired between the two dates were used to predict the Landsat-like NDVI image each time. For the FSDAF, the MODIS NDVI image on the prediction date and one Landsat and MODIS NDVI image pair acquired closer to the prediction date was used to predict the Landsat-like NDVI image each time.

Figure 16 shows the temporal profiles of the average NDVI of an area ( $360\text{ m} \times 360\text{ m}$ ) from a healthy cornfield and an area ( $360\text{ m} \times 360\text{ m}$ ) from a healthy winter wheat field between DOY 113 and DOY 249, which were generated by the STARFM, ESTARFM, FSDAF, and the STVIFM. These two fields were surrounded by different crop types; therefore, the MODIS pixels contain mixed Landsat pixels. The average NDVI extracted from the original MODIS NDVI time series images and five Landsat-8 NDVI images are also presented as comparisons.



**Figure 16.** Time series of the average NDVI of: the cornfield (a); and the wheat field (b), generated by the STARFM, ESTARFM, FSDAF, and STVIFM algorithms. The predictions shown in the black box (DOY 185) present large difference between the FSDAF and STVIFM. The pictures were collected two days before that date (DOY 183).

The corn and winter wheat show two distinct temporal patterns due to the difference of their growing seasons. The corn was generally seeded in May and harvested in October. The winter wheat was generally seeded in October of the previous year, started to ripen from the beginning of July and was harvested by the end of July, then alfalfa was planted. Before the emergence of corn, the MODIS NDVI shows higher and fluctuated values in the cornfield due to the influence of neighboring wheat fields. All the methods can generate reasonable predictions during this period. Between 7 June and 10 August, only two Landsat images were acquired, and the predictions of the STVIFM and

FSDAF show large difference on 4 July (DOY 185). As the image pair closer to the prediction date was always used, the profile generated by the FSDAF shows peak and valley, but large inaccuracy was still produced on some dates when the land cover changed greatly. During this period, corn was in its growing stage whereas winter wheat was harvested and alfalfa was planted. From the general survey pictures collected in 2014, corn had reached to the stem elongation stage (BBCH-scale 33) and most fields were covered by corn leaves, whereas the color of wheat started to turn yellow (BBCH-scale 79). Therefore, the STVIFM seems to generate more reasonable temporal profiles for corn and winter wheat. The ESTARFM also shows reasonable temporal profiles for the two types of crops, however, the prediction of the land cover change for the wheat field is less accurate than the prediction produced by the STVIFM when compared with the nearby Landsat NDVI values.

## 4. Discussion

### 4.1. Advantages of the STVIFM

The algorithm tests at three study sites illustrated that the STVIFM algorithm performed better than the STARFM, ESTARFM and FSDAF at the three study sites. According to the results of the above tests, the performance of the FSDAF greatly depends on the degree of land cover change between the two dates of the input data as it uses only one pair of fine- and coarse-resolution images as input, which agrees with the findings stated in [31]. It performs better than other methods which use one image pair [22] and it is flexible when only one fine-resolution image can be acquired. However, it is less robust than methods using two fine-resolution images as inputs in the land cover change prediction. Even though the FSDAF can predict NDVI with higher accuracy when the time interval between the two dates of the input images are close enough or land covers are similar enough, the STVIFM still performs better than the FSDAF during the growing stage or senescent stage. In addition, the STVIFM is more computationally efficient and performs about three times faster than the FSDAF at Kansas site and Xinjiang site, where the sizes are  $24 \text{ km} \times 24 \text{ km}$  and  $45 \text{ km} \times 45 \text{ km}$ , respectively.

Since the inputs of the STVIFM are same as the inputs of the STARFM and ESTARFM, the theoretical comparisons are made between the three methods. Compared with the STARFM and ESTARFM algorithm, the STVIFM has made several improvements. Firstly, the STVIFM builds a relationship between the mean NDVI change of fine-resolution pixels and mean NDVI change of coarse-resolution pixels within a moving window. It attempts to detect the mean fine-resolution NDVI change calculated from the coarse-resolution NDVI images and to seek each fine-resolution pixel's contribution to the total NDVI change by calculating the weight of each fine-resolution pixel. In contrast, the STARFM and ESTARFM build a relationship between the NDVI change of single fine-resolution pixels and single coarse-resolution pixels, therefore accurate geometric correction between the fine- and coarse-resolution images is required in order to achieve more accurate results [36].

Secondly, the ESTARFM assumes that the relationships between the fine-resolution and coarse-resolution image pairs are the same on all dates. However, due to the difference of weather conditions, the relationship between the two images may be different on different dates. The STVIFM attempts to obtain the coefficients between the fine-resolution and coarse-resolution image pairs on different dates using linear regression analyses. However, it is difficult to obtain the coefficients between the images on the prediction date, due to the unavailability of the fine-resolution image. The STVIFM adopts the weights calculated from the correlation coefficients between the coarse-resolution images to obtain the coefficients between the fine- and coarse-resolution images on the prediction date.

Thirdly, each of the STVIFM, STARFM and ESTARFM applied a weighting system to calculate the NDVI of the central pixel, but the meaning of the weighting system for STVIFM and the weighting system for STARFM and ESTARFM are different. The weight in the STARFM or ESTARFM means the similarity between the central pixel and the surrounding similar pixels within the moving window, whereas the weight in the STVIFM means the variation in contributions of fine-resolution pixels to

the total NDVI change within the moving window. The STVIFM considers the change rate variation at both spatial scale and temporal scale, which is more reasonable for non-evergreen vegetation. It attempts to calculate the spatial variation of NDVI change (spatial weight) of each fine-resolution pixel at any prediction date by incorporating the weights calculated based on one base fine-resolution image and the temporal NDVI change of the two fine-resolution images. These two elements are incorporated according to the land cover similarity between the prediction date and the two base dates. However, the ESTARFM assumes that the change rate is stable during a short period. This assumption is reasonable if the vegetation is evergreen or if the period between the two input image pairs is short enough (e.g., one day), but it would be unreasonable if the period is longer (e.g., more than 10 days) [18].

Lastly, the two predictions obtained from the two base dates are combined using a temporal weight for the ESTARFM and a similarity weight for the STVIFM. The ESTARFM calculated the temporal weight using the mean absolute difference between two coarse-resolution pixels within the moving window [18]. However, this is not the best selection to determine the land cover similarity in heterogeneous region. For instance, the mean absolute difference may be the same for area where has large land cover change (NDVI decrease mixed with NDVI increase), and area where has the same land cover but with NDVI decrease or increase. Therefore, the STVIFM adopts the correlation of determination for heterogeneous areas and the mean absolute difference for homogeneous areas to calculate the similarity weight.

Due to the advantages mentioned above, the STVIFM can make more accurate NDVI predictions in heterogeneous regions than the STARFM and ESTARFM when the land cover or NDVI changes were captured by the two pairs of fine- and coarse-resolution images. The accuracy improvements of the STVIFM are more obvious for Ontario site and Xinjiang site, which are characterized by heterogeneous cropland areas. Accordingly, the STVIFM can generate more reasonable NDVI time series for winter wheat and corn, which have different growing season.

#### 4.2. Limitations and Uncertainties of the STVIFM

In addition to the advantages mentioned above, it is worth noting that the STVIFM has its limitations. The following three aspects are the theoretical limitations of the STVIFM algorithm. Firstly, the STVIFM algorithm assumes that the NDVI is spatially additive. This linear assumption for the NDVI may lead to minor inaccuracies since the NDVI is not a linear combination of reflectance. Secondly, the relationship between the fine-resolution and coarse-resolution images acquired at  $t_p$  is calculated from the relationship between the two input image pairs. The result obtained in this way may be slightly different from the real relationship. Additionally, the STVIFM adopts the same coefficients for the whole image, but the coefficients may vary at different locations. More efforts should be made in further work to obtain the coefficients using a more accurate way.

There are some practical limitations with the STVIFM. Firstly, the small-area (width or length is less than one coarse-resolution pixel) abrupt disturbances that occurred between  $t_m$  and  $t_p$  or between  $t_p$  and  $t_n$ , may not be accurately detected because the influence of other land covers in one coarse-resolution pixel. Secondly, if the dates  $t_m$  and  $t_n$  are in the growing and senescent period of the vegetation, respectively, and the NDVI change at  $t_p$  is not captured by the images acquired at  $t_m$  and  $t_n$ , the performance of the STVIFM is slightly worse than the other methods since the NDVI change from  $t_m$  to  $t_n$  cannot reflect the NDVI change from  $t_m$  to  $t_p$ . In this case, more frequent high spatial resolution images that can cover the important vegetation phenology will be helpful. Another possible way to improve the prediction accuracy is to integrate the fusion model with a vegetation growth model for different types of vegetation.

#### 4.3. Applications of the STVIFM

The STVIFM uses two pairs of Landsat-8 OLI and MODIS images acquired before and after the prediction date and one coarse-resolution image on the prediction date as inputs, to predict the

fine-resolution NDVI image on the prediction date. This algorithm can be applied to regions with different landscapes such as grassland, forest and cropland areas. It can also be applied to other vegetation indices, but the thresholds may need to be adjusted accordingly. Besides the Landsat-8 OLI and MODIS data, other high spatial resolution data such as the SPOT, RapidEye, Sentinel-2, and high temporal frequency data such as AVHRR, MERIS can also be used. There are four parameters that could be set in the STVIFM, the window size for coefficients deriving and the window size for STVIFM implementation, the NDVI value for the maximum change rate of vegetation and the variance of the change rate index. The window size should be the odd rounding value of the integer multiple of the resolution ratio between the coarse- and fine-resolution images. The suggested window size for coefficients deriving and algorithm implementation is 25 or 33 for Landsat and MODIS data. However, for images with different spatial scales, the window size may need to be adjusted. Since the theoretical NDVI values for vegetation pixels range from 0 to 1, the median value 0.5 is suggested as the value of  $d$ , but the value can be adjusted according to the actual NDVI range of vegetation for special vegetation cover types. The suggested value for  $\sigma^2$  is 0.1–0.2 to obtain the change rate index with a dynamic range from 0 to 1. For the NDVI time series generation, different spatio-temporal data fusion methods may need to be incorporated to improve accuracy.

## 5. Conclusions

In this study, a spatio-temporal vegetation index image fusion model (STVIFM) was developed to fuse high spatial resolution and high temporal frequency NDVI images. The STVIFM algorithm considers the differences between fine-resolution and coarse-resolution pixel values on different dates. It also considers the variations of change rate at the spatial scale and temporal scale by using a temporal weight calculated from the correlation coefficients between two temporally adjacent coarse-resolution images. The STVIFM outperforms in NDVI prediction than the STARFM and ESTARFM when the land cover or NDVI changes are captured by the two pairs of fine- and coarse-resolution images. For the results predicted by STVIFM, the  $R^2$  varied between 0.711 and 0.891 and the RMSE varied between 0.065 and 0.76 for three study sites with different landscapes, which shows a higher NDVI prediction accuracy than the STARFM and ESTARFM. The STVIFM is more robust than the FSDAF when there are large land cover changes between the prediction date and the date of the image pairs. In addition, the STVIFM is more computationally efficient than the FSDAF. The STVIFM enhances the capability for generating both high spatial resolution and high temporal frequency NDVI images in heterogeneous regions. More efforts are needed in the future for the calculation of coefficients between different sensor images obtained under different weather conditions and geographic locations, and for the prediction of land cover changes that are not captured in the two fine-resolution images.

**Acknowledgments:** This work was funded by Ontario Trillium Scholarship (OTS) for Chunhua Liao's Ph.D. research and partially funded by an NSERC Discovery Grant to Jinfei Wang. The authors would like to acknowledge Xiaolin Zhu for providing the ESTARFM and FSDAF packages and Feng Gao for providing the STARFM package. The authors thank the USGS EROS data center for providing free Landsat data and MODIS science team for providing free MODIS products. The authors would also like to thank Taifeng Dong from AAFC, and the anonymous reviewers.

**Author Contributions:** Chunhua Liao developed the algorithm and the code for this algorithm, and prepared the first draft of the paper. Jinfei Wang, Ian Pritchard, Jianguo Liu, and Jiali Shang contributed to the design of the algorithm test and results interpretation part and edited the paper.

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

## References

1. 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](#)]
2. Fensholt, R. Earth observation of vegetation status in the Sahelian and Sudanian West Africa: Comparison of terra MODIS and NOAA AVHRR satellite data. *Int. J. Remote Sens.* **2004**, *25*, 1641–1659. [[CrossRef](#)]

3. 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](#)]
4. Jönsson, 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](#)]
5. Kang, S.; Running, S.W.; Lim, J.H.; Zhao, M.; Park, C.R.; Loehman, R. A regional phenology model for detecting onset of greenness in temperate mixed forests, Korea: An application of MODIS leaf area index. *Remote Sens. Environ.* **2003**, *86*, 232–242. [[CrossRef](#)]
6. 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](#)]
7. 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](#)]
8. Olexa, E.M.; Lawrence, R.L. Performance and effects of land cover type on synthetic surface reflectance data and NDVI estimates for assessment and monitoring of semi-arid rangeland. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *30*, 30–41. [[CrossRef](#)]
9. Schmidt, M.; Udelhoven, T.; Gill, T.; Röder, 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*, 63512. [[CrossRef](#)]
10. Tewes, A.; Thonfeld, F.; Schmidt, M.; Oomen, R.; Zhu, X.; 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](#)]
11. 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](#)]
12. Watts, J.D.; Powell, S.L.; Lawrence, R.L.; Hilker, T. Improved classification of conservation tillage adoption using high temporal and synthetic satellite imagery. *Remote Sens. Environ.* **2011**, *115*, 66–75. [[CrossRef](#)]
13. Gaulton, R.; Hilker, T.; Wulder, M.A.; Coops, N.C.; Stenhouse, G. Characterizing stand-replacing disturbance in western Alberta grizzly bear habitat, using a satellite-derived high temporal and spatial resolution change sequence. *For. Ecol. Manag.* **2011**, *261*, 865–877. [[CrossRef](#)]
14. 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](#)]
15. Dong, T.; Liu, J.; Qian, B.; Zhao, T.; Jing, Q.; Geng, X.; Wang, J.; Huffman, T.; Shang, J. Estimating winter wheat biomass by assimilating leaf area index derived from fusion of Landsat-8 and MODIS data. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *49*, 63–74. [[CrossRef](#)]
16. Meng, J.; Du, X.; Wu, B. Generation of high spatial and temporal resolution NDVI and its application in crop biomass estimation. *Int. J. Digit. Earth* **2011**, *6*, 203–218. [[CrossRef](#)]
17. Anderson, M.C.; Kustas, W.P.; Norman, J.M.; Hain, C.R.; Mecikalski, J.R.; Schultz, L.; González-Dugo, M.P.; Cammalleri, C.; D'Urso, G.; Pimstein, A.; et al. Mapping daily evapotranspiration at field to continental scales using geostationary and polar orbiting satellite imagery. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 223–239. [[CrossRef](#)]
18. 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](#)]
19. Gao, F.; Masek, J.; Schwaller, M.; Hall, F. On the Blending of the MODIS and Landsat ETM+ Surface Reflectance. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2207–2218. [[CrossRef](#)]
20. Hazaymeh, K.; Hassan, Q.K. Spatiotemporal image-fusion model for enhancing the temporal resolution of Landsat-8 surface reflectance images using MODIS images. *J. Appl. Remote Sens.* **2015**, *9*, 96095. [[CrossRef](#)]
21. Marfai, M.A.; Almohammad, H.; Dey, S.; Susanto, B.; King, L. Coastal dynamic and shoreline mapping: Multi-sources spatial data analysis in Semarang Indonesia. *Environ. Monit. Assess.* **2008**, *142*, 297–308. [[CrossRef](#)] [[PubMed](#)]
22. 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](#)]

23. Jarihani, A.; McVicar, T.; Van Niel, T.; Emelyanova, I.; Callow, J.; 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](#)]
24. Mathieu, P.P.; O’Neill, A. Data assimilation: From photon counts to Earth System forecasts. *Remote Sens. Environ.* **2008**, *112*, 1258–1267. [[CrossRef](#)]
25. Kempeneers, P.; Sedano, F.; Piccard, I.; Eerens, H. Data Assimilation of PROBA-V 100 and 300 m. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 3314–3325. [[CrossRef](#)]
26. Sedano, F.; Kempeneers, P.; Hurtt, G. A kalman filter-based method to generate continuous time series of medium-resolution NDVI images. *Remote Sens.* **2014**, *6*, 12381–12408. [[CrossRef](#)]
27. Zhukov, B.; Oertel, D.; Lanzl, F. Unmixing-based multisensor multiresolution image fusion. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 1212–1226. [[CrossRef](#)]
28. Wu, M.; Wang, H.; Niu, Z.; Zhao, Y.; Wang, C. A model for spatial and temporal data fusion. *J. Infrared Millim. Waves* **2012**, *31*, 80–84. [[CrossRef](#)]
29. 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](#)]
30. 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](#)]
31. 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](#)]
32. Zurita-Milla, R.; Kaiser, G.; Clevers, J.G.P.W.; Schneider, W.; Schaepman, M.E. Downscaling time series of MERIS full resolution data to monitor vegetation seasonal dynamics. *Remote Sens. Environ.* **2009**, *113*, 1874–1885. [[CrossRef](#)]
33. Huang, B.; Song, H. Spatiotemporal reflectance fusion via sparse representation. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3707–3716. [[CrossRef](#)]
34. Song, H.; Huang, B. Spatiotemporal satellite image fusion through one-pair image learning. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 1883–1896. [[CrossRef](#)]
35. Wu, B.; Huang, B.; Zhang, L.; Member, S. An Error-Bound-Regularized Sparse Coding for Spatiotemporal Reflectance Fusion. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 6791–6803. [[CrossRef](#)]
36. Chen, B.; Huang, B.; Xu, B. Comparison of spatiotemporal fusion models: A review. *Remote Sens.* **2015**, *7*, 1798–1835. [[CrossRef](#)]
37. Yang, J.; Wright, J.; Huang, T.S.; Ma, Y. Image Super-Resolution Via Sparse Representation. *IEEE Trans. Image Process.* **2010**, *19*, 2861–2873. [[CrossRef](#)] [[PubMed](#)]
38. 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](#)]
39. Bisquert, M.; Bordogna, G.; Bégué, A.; Candiani, G.; Teisseire, M.; Poncelet, P. A Simple Fusion Method for Image Time Series Based on the Estimation of Image Temporal Validity. *Remote Sens.* **2015**, *7*, 704–724. [[CrossRef](#)]
40. Xie, D.; Zhang, J.; Zhu, X.; Pan, Y.; Liu, H.; Yuan, Z.; Yun, Y. An improved STARFM with help of an unmixing-based method to generate high spatial and temporal resolution remote sensing data in complex heterogeneous regions. *Sensors (Basel)* **2016**, *16*, 207. [[CrossRef](#)] [[PubMed](#)]
41. Cheng, Q.; Liu, H.; Shen, H.; Wu, P.; Zhang, L. A Spatial and Temporal Nonlocal Filter-Based Data Fusion Method. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4476–4488. [[CrossRef](#)]
42. Settle, J.J.; Drake, N.A. Linear mixing and the estimation of ground cover proportions. *Int. J. Remote Sens.* **1993**, *14*, 1159–1177. [[CrossRef](#)]
43. Kerdiles, H.; Grondona, M.O. NOAA-AVHRR NDVI decomposition and subpixel classification using linear mixing in the Argentinean Pampa. *Int. J. Remote Sens.* **1995**, *16*, 1303–1325. [[CrossRef](#)]
44. Sun, L.; Schulz, K. Spatio-temporal LAI modelling by integrating climate and MODIS LAI data in a mesoscale catchment. *Remote Sens.* **2017**, *9*, 144. [[CrossRef](#)]
45. Lambers, H.; Chapin, F.S., III; Pons, T.L. Ecosystem and Global Processes: Ecophysiological Controls. In *Plant Physiological Ecology*, 2nd ed.; Springer: New York, NY, USA, 2008; p. 559. [[CrossRef](#)]
46. Gandhi, G.M.; Parthiban, S.; Thummalu, N.; Christy, A. NDVI: Vegetation change detection using remote sensing and gis—A case study of Vellore District. *Procedia Comput. Sci.* **2015**, *57*, 1199–1210. [[CrossRef](#)]

47. Zhang, X.; Liao, C.; Li, J.; Sun, Q. Fractional vegetation cover estimation in arid and semi-arid environments using: HJ-1 satellite hyperspectral data. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *21*, 506–512. [[CrossRef](#)]
48. Vermote, E.F.; Kotchenova, S.Y.; Ray, J.P. MODIS Surface Reflectance User's Guide. 2011. Available online: [http://modis-sr.ltdri.org/guide/MOD09\\_UserGuide\\_v1.4.pdf](http://modis-sr.ltdri.org/guide/MOD09_UserGuide_v1.4.pdf) (accessed on 1 March 2016).
49. 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](#)]



© 2017 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/>).