An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions

Xiaolin Zhu [a], Jin Chen [a,*], Feng Gao [b], Xuehong Chen [a], Jeffrey G. Masek [b]

[a] State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875, China
[b] Biospheric Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA

ABSTRACT

Due to technical and budget limitations, remote sensing instruments trade spatial resolution and swath width. As a result not one sensor provides both high spatial resolution and high temporal resolution. However, the ability to monitor seasonal landscape changes at fine resolution is urgently needed for global change science. One approach is to "blend" the radiometry from daily, global data (e.g. MODIS, MERIS, SPOT-Vegetation) with data from high-resolution sensors with less frequent coverage (e.g. Landsat, CBERS, ResourceSat). Unfortunately, existing algorithms for blending multi-source data have some shortcomings, particularly in accurately predicting the surface reflectance of heterogeneous landscapes. This study has developed an enhanced spatial and temporal adaptive reflectance fusion model (ESTARFM) based on the existing STARFM algorithm, and has tested it with both simulated and actual satellite data. Results show that ESTARFM improves the accuracy of predicted fine-resolution reflectance, especially for heterogeneous landscapes, and preserves spatial details. Taking the NIR band as an example, for homogeneous regions the prediction of the ESTARFM is slightly better than the STARFM (average absolute difference [*AAD*] 0.0106 vs. 0.0129 reflectance units). But for a complex, heterogeneous landscape, the prediction accuracy of ESTARFM is improved even more compared with STARFM (*AAD* 0.0135 vs. 0.0194). This improved fusion algorithm will support new investigations into how global landscapes are changing across both seasonal and interannual timescales.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

Due to technical and budget limitations, remote sensing instruments trade spatial resolution and swath width. As a result it is difficult to acquire remotely sensed data with both high spatial resolution and frequent coverage (Price, 1994). For example, remotely sensed images acquired from Landsat series satellites, SPOT, and IRS with a spatial resolution from 6 to 30 m are usually the primary data source for land use/cover mapping and change detection (Woodcock & Ozdogan, 2004), monitoring ecosystem dynamics (Healey et al., 2005; Masek & Collatz, 2006; Masek et al., 2008), as well as biogeochemical parameter estimation (Cohen & Goward, 2004). However, the long revisit cycles of these satellites (Landsat/TM: 16-day; SPOT/HRV: 26-day; IRS: 24-day), frequent cloud contamination, and other poor atmospheric conditions (Asner, 2001; Jorgensen, 2000; Ju & Roy, 2008) have limited their use in detecting rapid surface changes associated with intraseasonal ecosystem variations (disturbance and phenology) and natural disasters (Gonzalez-Sanpedro et al., 2008; Ranson et al., 2003). On the other hand, the Moderate

Resolution Imaging Spectroradiometer (MODIS) on the Terra/Aqua, SPOT-Vegetation (SPOT-VGT), and NOAA Advanced Very High Resolution Radiometer (AVHRR) can provide high frequent (daily) observations, but with coarse spatial resolutions ranging from 250 m to 1000 m. This resolution is not sufficient for monitoring land cover and ecosystem changes within heterogeneous landscapes (Shabanov et al., 2003). Thus, combining remotely sensed data from different sensors is a feasible and less expensive way to enhance the capability of remote sensing for monitoring land surface dynamics (Camps-Valls et al., 2008; Gao et al., 2006; Marfai et al., 2008).

Traditional image fusion methods such as the intensity–hue–saturation (IHS) transformation (Carper et al., 1990), principal component substitution (PCS) (Shettigara, 1992), and wavelet decomposition (Yocky, 1996) focus on producing new multispectral images that combine high-resolution panchromatic data with multispectral observations acquired simultaneously at coarser resolution (Pohl & van Genderen, 1998; Zhang, 2004). They are useful for exploiting different spectral and spatial characteristics of multi-sensor data. However, they are not effective in enhancing spatial resolution and temporal coverage simultaneously because the panchromatic band is only helpful for enhancing the spatial resolution (up to a certain extent). However, enhancing spatial resolution and temporal coverage simultaneously is necessary for studying inter- and intra-

* Corresponding author. Tel.: +86 10 13522889711.
  E-mail address: chenjin@ires.cn (J. Chen).

annual vegetation dynamics. To simulate reflectance data with both high spatial resolution and frequent coverage, Gao et al. (2006) developed a spatial and temporal adaptive reflectance fusion model (STARFM) to blend Landsat and MODIS data for predicting daily surface reflectance at Landsat spatial resolution and MODIS temporal frequency. Testing using both simulated and actual data demonstrated the effectiveness of the STARFM for accurately predicting daily surface reflectance. Another downscaling algorithm was recently developed based on a linear mixing model to produce Landsat-like images having the spectral and temporal resolution provided by the Medium Resolution Imaging Spectrometer (MERIS) (Zurita-Milla et al., 2009). Nevertheless, this downscaling algorithm requires a high-resolution land use database for pixel unmixing, which may not be available for many applications. Compared with the downscaling algorithm, the STARFM method does not need any ancillary data. It also has been demonstrated and validated in a conifer-dominated area in central British Columbia, Canada, for which it produced daily surface reflectance at Landsat spatial resolution and in good agreement with actual Landsat data (Hilker et al., 2009b).

Although recent results of the STARFM method suggest new opportunities for producing remotely sensed data with both high spatial resolution and frequent coverage (Gao et al., 2006; Hilker et al., 2009a,b), it should be noted that the original STARFM method also has three limitations that need to be rectified before widespread application. First, the STARFM method cannot predict disturbance events if the changes caused by disturbance are transient and not recorded in at least one of the base Landsat images (Hilker et al., 2009a). Aiming to solve this limitation, a new fusion algorithm called as Spatial Temporal Adaptive Algorithm for mapping Reflectance Change (STAARCH) has been developed for the vegetated land surface based on the STARFM method (Hilker et al., 2009a). The STAARCH algorithm determines spatial changes from Landsat and temporal changes from MODIS, which allows the algorithm to choose an optimal Landsat base date and thus improve the accuracy of the synthetic Landsat images. Secondly, the STARFM method does not explicitly handle the directional dependence of reflectance as a function of the sun–target–sensor geometry described by the Bi-directional Reflectance Distribution Function (BRDF) (Roy et al., 2008). A semi-physical fusion approach was developed to solve this limitation, which uses the MODIS BRDF/Albedo land surface product and Landsat ETM+ data to predict ETM+ reflectance on the same, an antecedent, or subsequent date (Roy et al., 2008). Last, the quality of the predicted Landsat-like image depends on the geographic region of interest. The STARFM relies on temporal information from pure, homogeneous patches of land cover at the MODIS pixel scale. These "pure" pixels are identified by the homogeneity of Landsat pixels within the MODIS cell boundary. Simulations and predictions based on the actual Landsat and MODIS images show that STARFM can predict reflectance accurately if these coarse-resolution homogeneous pixels exist (Gao et al., 2006). However, prediction results degrade somewhat when used on heterogeneous fine-grained landscapes, including small-scale agriculture (Gao et al., 2006; Hilker et al., 2009b). The STAARCH algorithm (Hilker et al., 2009a) only chooses the optimal Landsat acquisitions for STARFM, so it has the same problem as STARFM for heterogeneous regions. The assumption of the semi-physical fusion approach (Roy et al., 2008) that the MODIS modulation term c is representative of the reflectance variation at Landsat ETM+ scale does not hold when reflectance change occurs in a spatially heterogeneous manner at scales larger than the 30 m Landsat pixels and smaller than the 500 m MODIS pixels (Roy et al., 2008). Thus it too has the same difficulty in predicting the reflectance of heterogeneous landscapes as the original STARFM method.

To solve the last limitation of the STARFM method, the accurate prediction of surface reflectance in heterogeneous landscapes, we developed an enhanced STARFM method (ESTARFM). The ESTARFM improves on the original STARFM algorithm by using the observed

reflectance trend between two points in time, and spectral unmixing theory, in order to better predict reflectance in changing, heterogeneous landscapes. The approach was validated by employing a small number of pairs (two or more) of fine spatial (e.g. Landsat) and coarse spatial resolution images (e.g. MODIS) acquired on the same day and a series of coarse spatial resolution images (e.g. MODIS) acquired on the desired prediction dates. In this paper, the theoretical basis of the ESTARFM method is first presented, and then results from the ESTARFM method based on simulated data and actual Landsat/ MODIS images are given and compared to the original STARFM method.

2. Theoretical basis of the ESTARFM method

For a given region, we assume that remotely sensed data from different satellite sensors acquired at the same date are comparable and correlated with each other after radiometric calibration, geometric rectification, and atmospheric correction. However, due to differences in sensor systems such as orbit parameters, bandwidth, acquisition time and spectral response function, there may be systematic biases in surface reflectance among different sensor images. The main idea of the ESTARFM is to make use of the correlation to blend multi-source data and meanwhile to minimize the system biases. According to the heterogeneity of land surfaces, pure pixels and mixed pixels are discussed below respectively. Moreover, for convenience in the discussion, we will call the image with low spatial resolution but frequent coverage as the "coarse-resolution" image, while the image with high spatial resolution but infrequent coverage will be identified as the "fine-resolution" image. We also suppose that the coarse-resolution sensor has similar spectral bands (e.g. band B) as the fine-resolution sensor.

2.1. Pure coarse-resolution pixel

Assume that the coarse-resolution reflectance image has been resampled to the same spatial resolution, size and extent of the fine-resolution image. For a pure, homogeneous coarse-resolution pixel which is covered by only one land type, the difference of reflectance between the resampled coarse-resolution pixel and fine-resolution pixel only results from the systematic biases mentioned above. Therefore, the relationship between the coarse-resolution reflectance and fine-resolution reflectance can be reasonably described by a linear model expressed as:

$$F(x, y, t_k, B) = a \times C(x, y, t_k, B) + b \qquad (1)$$

where F, C denote the fine-resolution reflectance and coarse-resolution reflectance respectively, (x, y) is a given pixel location for both fine-resolution and coarse-resolution images, t_k is the acquisition date, a and b are coefficients of the linear regression model for relative calibration between coarse and fine-resolution reflectance. For the pure coarse-resolution pixels, Eq. (1) should be stable considering the stability of sensors for extended periods. Considering differences of atmospheric condition, solar angle and altitude at different locations, and that radiometric calibration, geometric rectification and atmospheric correction cannot completely remove this variability, the coefficients a and b might change with location. Therefore, the coefficients a and b should be derived locally rather than using global coefficients.

Suppose we have a fine-resolution image and coarse-resolution image acquired at t_0 and another coarse-resolution image acquired at t_p. If the land cover type and sensor calibration do not change between t_0 and t_p, Eq. (1) can be written as Eq. (2) at t_0 and Eq. (3) at t_p:

$$F(x, y, t_0, B) = a \times C(x, y, t_0, B) + b \qquad (2)$$

$$F(x, y, t_p, B) = a \times C(x, y, t_p, B) + b. \tag{3}$$

From Eqs. (2) and (3), we can obtain:

$$F(x, y, t_p, B) = F(x, y, t_0, B) + a \times (C(x, y, t_p, B) - C(x, y, t_0, B)). \tag{4}$$

Eq. (4) shows that the fine-resolution reflectance at t_p equals the sum of the fine-resolution reflectance at t_0 and the scaled change of reflectance from t_0 to t_p given by coarse-resolution images from different dates. Because $F(x, y, t_0, B)$, $C(x, y, t_0, B)$, $C(x, y, t_P, B)$ all are known, accordingly, we can calculate the fine-resolution reflectance $F(x, y, t_p, B)$ at t_p when the coefficient a is known, even if there is no actual fine-resolution data. Here the conversion coefficient a is determined by the system biases between the sensors which can be considered stable for each pixel, if the acquisition time is close and thus atmospheric conditions are almost identical (or negligible after correction). If we can acquire two pairs of fine-resolution and coarse-resolution images at times t_m and t_n, we can obtain the coefficient a by linear regression of the fine-resolution reflectance against the coarse-resolution reflectance at t_m and t_n, and then calculate the fine-resolution reflectance at prediction time t_p. It is noted that the assumption of stable conversion coefficient a is strictly true only over non-changing surfaces such as deserts or water bodies after geometric rectification and atmospheric correction are performed. Except for these cases, the coefficient a may slightly change with time and could introduce some error in the Eq. (4) calculation.

2.2. Mixed (heterogeneous) coarse-resolution pixel

Due to the complexity of the land surface, most of the pixels in coarse-resolution images are mixed pixels (i.e. covered by multiple land cover types). In this case, the relationship between the coarse-resolution reflectance and fine-resolution reflectance may not exist as described in Eq. (4). Supposing that the reflectance of a mixed pixel can be modeled as a linear combination of the reflectance of the different land cover components present in that pixel weighted by their fractional area coverage (Adams et al., 1985), the changes in reflectance of a mixed pixel between two dates represent the weighted sum of changes in reflectance for each land cover type within the pixel. Assuming that the proportions of land cover types contained in the mixed coarse-resolution pixel are not changed from date t_m to t_n, the reflectance of a mixed coarse-resolution pixel can be described as following according to linear mixture model:

$$C_m = \sum_{i=1}^{M} f_i \left(\frac{1}{a} F_{im} - \frac{b}{a} \right) + \varepsilon$$
$$C_n = \sum_{i=1}^{M} f_i \left(\frac{1}{a} F_{in} - \frac{b}{a} \right) + \varepsilon \tag{5}$$

where C_m, C_n are reflectance of mixing coarse-resolution pixel at date t_m and t_n respectively, f_i is fraction of ith land type (ith endmember), F_{im} and F_{in} are reflectance of ith endmember at date t_m and t_n obtained in fine-resolution image respectively, M is the total number of endmembers, and ε is the residual error, a, b are the coefficients of relative calibration between coarse and fine-resolution reflectance as described in Section 2.1. All of the fine-resolution pixels contained within the mixed coarse-resolution pixels can be regarded as endmembers of the coarse-resolution pixel, so F_{im} and F_{in} are the reflectance of fine-resolution pixels of different land types. However, owing to the bias between the fine- and coarse-resolution reflectance described in Section 2.1, F_{im} and F_{in} must be calibrated to be the endmember reflectance of coarse-resolution mixing pixel. From

Eq. (5), we can get the changes of coarse-resolution reflectance from t_m to t_n:

$$C_n - C_m = \sum_{i=1}^{M} \frac{f_i}{a} (F_{in} - F_{im}). \tag{6}$$

We also suppose the change of reflectance of each endmember is linear from t_m to t_n, then the reflectance of ith endmember at t_n can be described by the reflectance at t_m:

$$F_{in} = h_i \times \Delta t + F_{im} \tag{7}$$

where $\Delta t = t_n - t_m$, and h_i is the change rate and can be thought as stable during a period. The assumption that the reflectance linearly changes from t_m to t_n is reasonable during a short time period. Admittedly, the reflectance change might not be linear in some situations, such as phenological change of vegetation. In such cases, linear approximation is a tradeoff choice because the accurate nonlinear model is unknown, although this approximation could add some error. Then Eq. (6) can be rewritten as:

$$C_n - C_m = \Delta t \sum_{i=1}^{M} \frac{f_i h_i}{a}. \tag{8}$$

If the reflectance of kth endmember at date t_m and t_n is known, the Δt can also be represented from Eq. (7) as:

$$\Delta t = \frac{F_{kn} - F_{km}}{h_k} \tag{9}$$

where h_k is the change rate of kth endmember reflectance. Taking Eq. (9) into Eq. (8), it can be rewritten as:

$$\frac{F_{kn} - F_{km}}{C_n - C_m} = \frac{h_k}{\sum_{i=1}^{M} \frac{f_i h_i}{a}} = v_k. \tag{10}$$

The right part of Eq. (10) is a constant given our prior assumptions that the proportion of each endmember and the reflectance change rate of each endmember are stable. Thus v_i indicates the ratio of the change of reflectance for kth endmember to the change of reflectance for a mixed coarse-resolution pixel. For consistency with the pure coarse-resolution pixel, we also call v_i a conversion coefficient below. From Eq. (10), we can find that there is a linear relationship for reflectance change between the endmember and mixed coarse-resolution pixel. When the endmember is taken as fine-resolution pixels (x, y) within a mixed coarse-resolution pixel, we can obtain the conversion coefficient $v(x, y)$ by linearly regressing the reflectance changes of fine-resolution pixels of the same endmember and coarse-resolution pixel.

Similarly, if one pair of fine-resolution and coarse-resolution images at t_0 and another coarse-resolution image at t_p has been acquired, the unknown reflectance of fine-resolution pixel at t_p can be predicted according to Eq. (11):

$$F(x, y, t_p, B) = F(x, y, t_0, B) + v(x, y) \times (C(x, y, t_p, B) - C(x, y, t_0, B)). \tag{11}$$

Although Eqs. (4) and (11) have the same form, their meanings are different. Eq. (4) represents the relative normalization of pure pixels between different resolution images and it is reasonable for each date, so the prediction of reflectance of fine-resolution pixel is more accurate, while Eq. (11) describes the relationship of reflectance change between an endmember (fine-resolution pixel) and mixed coarse-resolution pixel according to a series of assumptions. It is a reasonable assumption during a relatively short period in which the proportion of each endmember and the change rate of reflectance of each endmember are stable. Note that Eq. (4) is a special case of

Eq. (11) when the coarse-resolution pixel is dominated by just one endmember.

It is obvious that Eqs. (4) and (11) only use information from a single pixel to predict fine-resolution reflectance. Considering neighboring same-class pixels with similar reflectance changes, a moving window method described by Gao et al. (2006) is used to take full advantage of the information from neighboring pixels. Here the moving window is used to search similar pixels within the window and information of similar pixels is then integrated into fine-resolution reflectance calculation as described in Eq. (12). In detail, assuming w is the search window size, the fine-resolution reflectance of the central pixel $(x_{w/2}, y_{w/2})$ at date t_p can be calculated as follows according to Eq. (4) or Eq. (11):

$$F(x_{w/2}, y_{w/2}, t_p, B) = F(x_{w/2}, y_{w/2}, t_0, B) + \sum_{i=1}^{N} W_i \times V_i \times (C(x_i, y_i, t_p, B) - C(x_i, y_i, t_0, B))$$

$$(12)$$

where N is the number of similar pixels including the central "prediction" pixel, (x_i, y_i) is the location of ith similar pixel, and W_i is the weight of ith similar pixel. Here, neighboring pixels with the same land cover type as the central pixel are called "similar" pixels. Thus they have spectral characteristics similar to the central pixel obtained from the fine-resolution image. Because we can predict the fine-resolution reflectance more accurately from a pure coarse-resolution pixel according to the above theoretical discussion, pure coarse-resolution pixels should be given larger weight value. V_i is the conversion coefficient of ith similar pixel. The search window size w is determined by the homogeneity of surface. If the regional landscape is more homogeneous, then w can be smaller. Eq. (12) means that the fine-resolution reflectance of prediction date equals the fine-resolution reflectance observed at one time (base date) added to the reflectance changes that are predicted from all similar pixels within the window in the resampled coarse-resolution image.

3. Process of ESTARFM implementation

Fig. 1 presents a flowchart of the ESTARFM. This algorithm requires at least two pairs of fine- and coarse-resolution images acquired at the same date and a set of coarse-resolution images for desired prediction dates. Before implementing the ESTARFM, all the images must be preprocessed to georegistered surface reflectance. There are four major steps in the ESTARFM algorithm implementation. First, two fine-resolution images are used to search for pixels similar to the central pixel in a local window. Second, the weights of all similar pixels (W_i) are calculated. Third, the conversion coefficients V_i are determined by linear regression. Finally, W_i and V_i are used to calculate the fine-resolution reflectance from the coarse-resolution image at the desired prediction date. All of the steps will be discussed in detail below.

3.1. Data preprocessing

Both coarse-resolution and fine-resolution images need to be preprocessed geometrically and radiometrically before using in the ESTARFM. For this study, Landsat ETM+ data were co-registered and orthorectified using the automated registration and orthorectification package (AROP) (Gao et al., 2009). Digital numbers from Landsat level 1 product were calibrated and atmospherically corrected using Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) (Masek et al., 2006). MODIS daily surface reflectance (MOD09GA) data were reprojected and resampled to the Landsat resolution and extent using MODIS Reprojection Tools (MRT). As LEDAPS uses similar atmospheric correction approach (6S approach) to the MODIS surface reflectance product, reflectance from two sensors was found consistent and comparable (Masek et al., 2006).
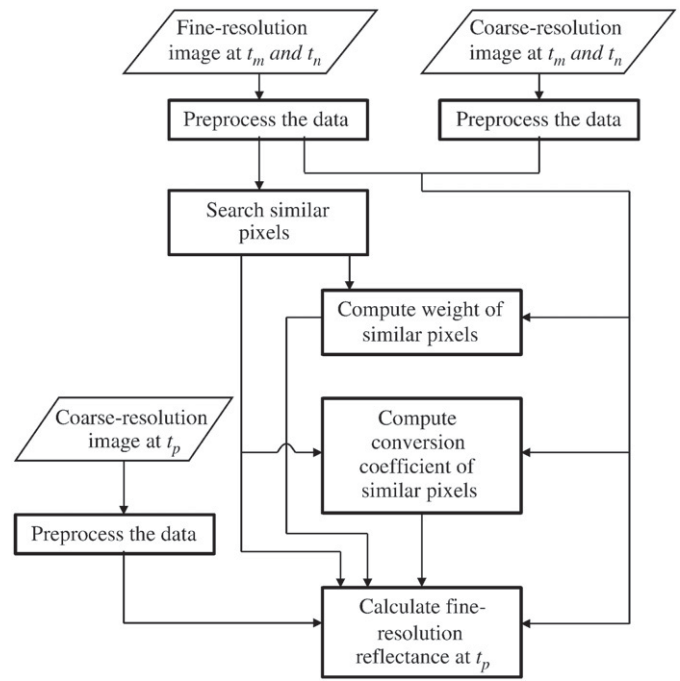

Fig. 1. The flowchart of the ESTARFM algorithm.

3.2. Selection of similar neighbor pixels

The pixels within the window with the same land cover type as the central pixel ("similar" pixels) provide specific temporal and spatial information to compute the fine-resolution reflectance for the central pixel. There are mainly two methods to search for similar pixels: (1) an unsupervised clustering algorithm is applied to the fine-resolution image and neighboring pixels belonging to the same cluster as the central pixel are identified; (2) the reflectance difference is computed between neighboring pixels and the central pixel in the fine-resolution image, and thresholds are used to identify similar pixels. The thresholds can be determined by the standard deviation of a population of pixels from the fine-resolution image and the estimated number of land cover classes of the image (Gao et al., 2006). If all the bands for the ith neighbor pixel satisfy Eq. (13), the ith neighbor pixel is selected as a similar pixel:

$$\left| F(x_i, y_i, t_k, B) - F(x_{w/2}, y_{w/2}, t_k, B) \right| \leq \sigma(B) \times 2/m \qquad (13)$$

where $\sigma(B)$ is the standard deviation of reflectance for band B, m is the estimated number of classes. Using a larger number of classes represents a stricter condition for selection similar pixels from fine-resolution images. Both approaches select pixels within the window with similar spectral characteristics as the central pixel. However, the clustering method applies the same clustering rules over the whole image and any misclassification has an adverse impact on all pixels in the image. On the contrary, the threshold method is applied within a local window. Even if individual pixels are incorrectly identified as spectrally "similar", the impact of the misclassification is restricted to the region within the local window. Therefore, we employed the threshold method to select similar pixels within the search window.

It is notable that the reflectance of some land cover types may change significantly from date 1 to date 2, resulting in some uncertainty in selecting the similar pixels if we only use one image date. For example, there may be bare soil and crop vegetation pixels in a search window with the central pixel covered by crop. If we use the image acquired when the crop has not greened up, the selected similar pixels may be bare soil because the spectral characteristics of cropland matches bare soil at that time. On the other hand, if we use

the image acquired when the crop has grown for a period, the selected similar pixels are likely green crops. Accordingly, in contrast to the original STARFM, we used the fine-resolution images acquired at t_m and t_n to select the similar pixels, respectively, and then extract the intersection of the two results to obtain a more accurate set of similar pixels. It is true that in some cases a central pixel may have no spectrally similar pixels within the search window However, even in the extreme case that no similar pixels exist except for the central pixel itself, the weight of the central pixel is set to 1.0 and the conversion coefficient can be computed according to the algorithm.

3.3. Calculation of weight for similar pixels

The weight W_i decides the contribution of ith similar pixel to predicting reflectance change at the central pixel. It is determined by the location of the similar pixel and the spectral similarity between the fine- and coarse-resolution pixel. Higher similarity and smaller distance of the similar pixel to the central pixel produce a higher weight (i.e. greater contribution) for the similar pixel. Here, spectral similarity is determined by correlation coefficient between each similar pixel and its corresponding coarse-resolution pixel as Eq. (14).

$$R_i = \frac{E[(\mathbf{F}_i - E(\mathbf{F}_i))(\mathbf{C}_i - E(\mathbf{C}_i))]}{\sqrt{D(\mathbf{F}_i)} \cdot \sqrt{D(\mathbf{C}_i)}} \qquad (14)$$

$$\mathbf{F}_i = \{F(x_i, y_i, t_m, B_1), ..., F(x_i, y_i, t_m, B_n), F(x_i, y_i, t_n, B_1), ..., F(x_i, y_i, t_n, B_n)\}$$

$$\mathbf{C}_i = \{C(x_i, y_i, t_m, B_1), ..., C(x_i, y_i, t_m, B_n), C(x_i, y_i, t_n, B_1), ..., C(x_i, y_i, t_n, B_n)\}$$

where R_i is the spectral correlation coefficient between fine- and coarse-resolution pixel for ith similar pixel, $\mathbf{F}_i$, $\mathbf{C}_i$ is the spectral vector containing the reflectance of each band at t_m and t_n for ith fine-resolution similar pixel and its corresponding coarse-resolution pixel, $E(\cdot)$ is the expected value, and $D(\mathbf{F}_i)$, $D(\mathbf{C}_i)$ is the variance of $\mathbf{F}_i$ and $\mathbf{C}_i$ respectively. The value of R varies from -1 to 1, and a larger R denotes a higher spectral similarity. The reason for combining the spectra of two different dates to compute the spectral similarity is that the spectral characteristics of some land cover types change through time; combining more spectral information from different times can provide a more accurate measure of similarity between the fine- and coarse-resolution pixel.

The geographic distance d_i between the ith similar pixel and central pixel can be calculated according to Eq. (15).

$$d_i = 1 + \sqrt{(x_{w/2} - x_i)^2 + (y_{w/2} - y_i)^2} / (w / 2) \qquad (15)$$

where w is the width of searching window that is used to normalize the distance, ensuring that the distance range for similar pixels in different search windows extends from 1 to $1 + 2^{0.5}$.

Combining spectral similarity and distance, a synthetic index D can be computed that combines spectral and geographic distance as:

$$D_i = (1 - R_i) \times d_i. \qquad (16)$$

As described above, a similar pixel with a larger D value should contribute less to computing the reflectance change for the central pixel, so we have used the normalized reciprocal of D as the weight W_i:

$$W_i = (1 / D_i) / \sum_{i=1}^{N} (1 / D_i). \qquad (17)$$

The range of W_i is from 0 to 1, and the total weight of all similar pixels is 1. For a special situation, when there are P similar pixels among all similar pixels whose corresponding coarse-resolution pixels are pure ($R = 1$), we defined the weight for these similar pixels

is $1/P$ and the weights of the rest of the similar pixels are 0, that is, all the change information is given by the pure coarse-resolution pixels with equal weight.

3.4. Calculation of conversion coefficient

It is desirable to calculate the conversion coefficient V_i by linear regression analysis for each similar pixel in a search window. Theoretically, for each similar pixel (x_i, y_i), its conversion coefficient can be computed from the fine- and coarse-resolution reflectance at the base time (t_m and t_n). However, since the preprocessing cannot remove all the contamination of the images and it is very hard to make the geometrical position of fine- and coarse-resolution images coincide accurately, only using each similar pixel to compute the conversion coefficient might cause a large uncertainty. Therefore, we take full advantage of the information from neighboring similar pixels to compute the conversion coefficient. Reflecting the fact that the spectral characteristics of the similar pixels within the same coarse pixel are more consistent with each other, it follows that they should have the same conversion coefficient. Consequently, we apply linear regression model to the fine- and coarse-resolution reflectance of the similar pixels within the same coarse pixel to obtain the conversion coefficients either for the pure coarse-resolution pixel or for the mixed coarse-resolution pixel. As Fig. 2 shows, the search window covers 4 intact and 8 partial coarse-resolution pixels. For similar pixels located in the intact coarse-resolution pixel, their conversion coefficients can be obtained from the regression of the fine-resolution reflectance against the coarse-resolution reflectance of these similar pixels. For the similar pixels located in the partial coarse-resolution pixel, the number of similar pixels sometimes is too few to build a reliable regression model. As Fig. 2 shows, there are only three similar
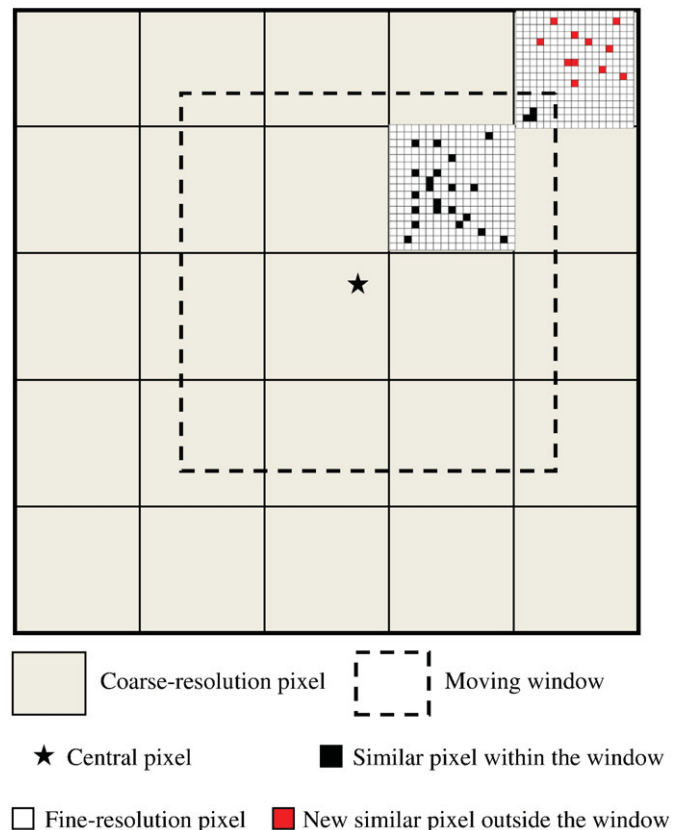

Fig. 2. Schematic diagram of the similar pixels within a same coarse-resolution pixel.

pixels (marked in black) in the upper-right partial coarse-resolution pixel. Therefore, according to the assumption that the similar pixels within the same coarse pixel have the same conversion coefficient, all the similar pixels in this partial coarse-resolution pixel are also selected additionally (marked in red color in Fig. 2). All the similar pixels within this partial coarse-resolution pixel are used to calculate conversion coefficient by regression analysis. It should be noted that that these new similar pixels outside the search window (the pixels marked in red color in Fig. 2) are only used to compute the conversion coefficient V, but are not used to predict the reflectance in the following step because they are outside the search window.

As an example of the linear regression analysis (Fig. 3), there are 34 similar pixels within a coarse-resolution pixel, and the two dashed rectangles label the reflectance of these pixels at t_m and t_n respectively. From these points in Fig. 3, a linear regression model can be built ($R^2 = 0.925$, $P<0.001$) and its slope corresponds to conversion coefficient V for all similar pixels within the coarse-resolution pixel ($V = 1.115$).

As a special case, if the linear regression model with statistical significance cannot be built, V has to be set as 1 even if it introduces some errors.

3.5. Calculation of reflectance of the central pixel

After the weight and conversion coefficient has been calculated according to Eq. (12), the fine-resolution reflectance at t_p can be predicted based on the fine-resolution reflectance at the base date and the resampled coarse-resolution reflectance observed at t_p. Also according to Eq. (12), either the fine-resolution reflectance at t_m or t_n can be used as the reflectance at base date to compute the fine-resolution reflectance of prediction date t_p, and the results are marked as F_m $(x_{w/2}, y_{w/2}, t_p, B)$ and F_n $(x_{w/2}, y_{w/2}, t_p, B)$ respectively. A more accurate reflectance at t_p can be obtained by a weighted combination of the two prediction results. Fine-resolution samples closer in date to the prediction date should show closer reflectance values, so it is reasonable to set a larger temporal weight to the fine-resolution reflectance input in this case. Thus the temporal weight can be calculated according to the change magnitude detected by resampled coarse-resolution reflectance between the time t_k ($k = m$ or n) and the prediction time t_p:

$$T_k = \frac{1 \Big/ \left| \sum_{j=1}^{w} \sum_{l=1}^{w} C(x_j, y_l, t_k, B) - \sum_{i=1}^{w} \sum_{l=1}^{w} C(x_j, y_l, t_p, B) \right|}{\sum_{k=m,n} \left(1 \Big/ \left| \sum_{j=1}^{w} \sum_{l=1}^{w} C(x_j, y_l, t_k, B) - \sum_{i=1}^{w} \sum_{l=1}^{w} C(x_j, y_l, t_p, B) \right| \right)}, (k = m, n).$$
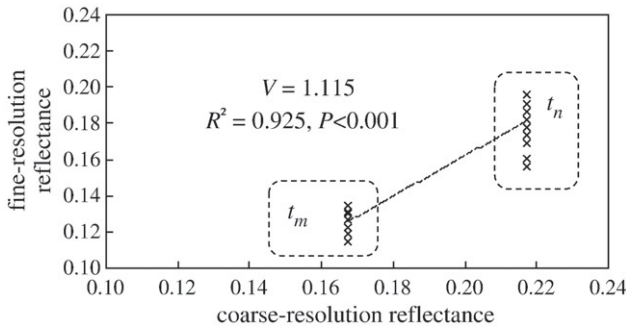
(18)

Fig. 3. An example of calculating the transition coefficient V.

Then the final predicted fine-resolution reflectance at the prediction time t_p is calculated as:

$$F(x_{w/2}, y_{w/2}, t_p, B) = T_m \times F_m(x_{w/2}, y_{w/2}, t_p, B) + T_n \times F_n(x_{w/2}, y_{w/2}, t_p, B).$$

(19)

4. Algorithm tests

4.1. Tests with simulated data

The ESTARFM algorithm was tested with simulated reflectance data, which helps to understand its accuracy and reliability. In order to compare with original STARFM algorithm, we used the same simulated data as Gao et al. (2006). In detail, a series of 153×153-pixel fine-resolution images were first simulated by assigning each pixel a positive value in the range 0 to 1 denoting the reflectance of each pixel. The coarse-resolution images were produced by scaling up the fine-resolution images (i.e., each cluster of 17×17 neighboring pixels in the fine-resolution images was aggregated to create a coarse-resolution pixel).

We tested four cases: changing reflectance, changing shapes, small objects and linear objects. The spatial resolutions of fine- and coarse-resolution simulated images are identical as those of Landsat/ETM and MODIS. Specifically, three pairs of fine- and coarse-resolution images acquired at same date were simulated, then the first and last pairs and the coarse-resolution image of the second pair were used to predict the fine-resolution image of the second pair, and the predicted and the real image were compared to assess the accuracy of the new algorithm. Our results show that the performance of the ESTARFM is the same as the STARFM for the cases of changing reflectance and changing shapes.

For the case of small objects, we simulated a series of triplet images with changing size of the circular objects from 90 to 480 m (from 3 fine-resolution pixels to 16 fine-resolution pixels). As an example of one image series (Fig. 4), a circular object had constant reflectance of 0.05 and its radius was 150 m (5 fine-resolution pixels). The background changed reflectance from 0.1 (date 1) to 0.2 (date 2) and then 0.4 (date 3) (Fig. 4(a), (b), (c)). The coarse-resolution images were aggregated from the fine-resolution images (Fig. 4(d), (e),(f)). Fig. 4(g) and (h) are predicted version of Fig. 4(b) using the ESTARFM and the STARFM respectively. It is obvious that both the algorithms can predict well the shape of the small circular object. Moreover, the predicted reflectance of the small circular object from the ESTARFM is closer to that in Fig. 4(b) compared with the STARFM and the relative errors of ESTARFM and STARFM are 0% and 131% respectively. As the circular object changes radius from 480 to 90 m, the reflectance predicted by ESTARFM remains accurate, while the prediction error by the original STARFM increases as the object becomes smaller (Fig. 5). Furthermore, the reflectance predicted by the original STARFM remains identical to that in Fig. 4(b) until the object size reaches the coarse-resolution pixel size (object radius $= 500*2^{0.5}/2 = 353$ m). These results suggest that theoretically the ESTARFM can predict the reflectance of an object accurately regardless of object's size. The original STARFM has a significant limitation in dealing with small objects with a characteristic size less than that of the coarse-resolution pixel size if "pure" homogeneous coarse-resolution pixels do not exist in the search window.

For the linear objects case, the simulated fine-resolution images contained three objects: the background with the reflectance changing from 0.1 (date 1), to 0.2 (date 2) and then 0.4 (date 3), a circular object with a constant reflectance of 0.05, and a linear object with a constant reflectance of 0.5 (Fig. 6(a), (b), (c)). The coarse-resolution images were aggregated from the fine-resolution images (Fig. 6(d), (e), (f)). Fig. 6(g) and (h) is the predicted version of Fig. 6
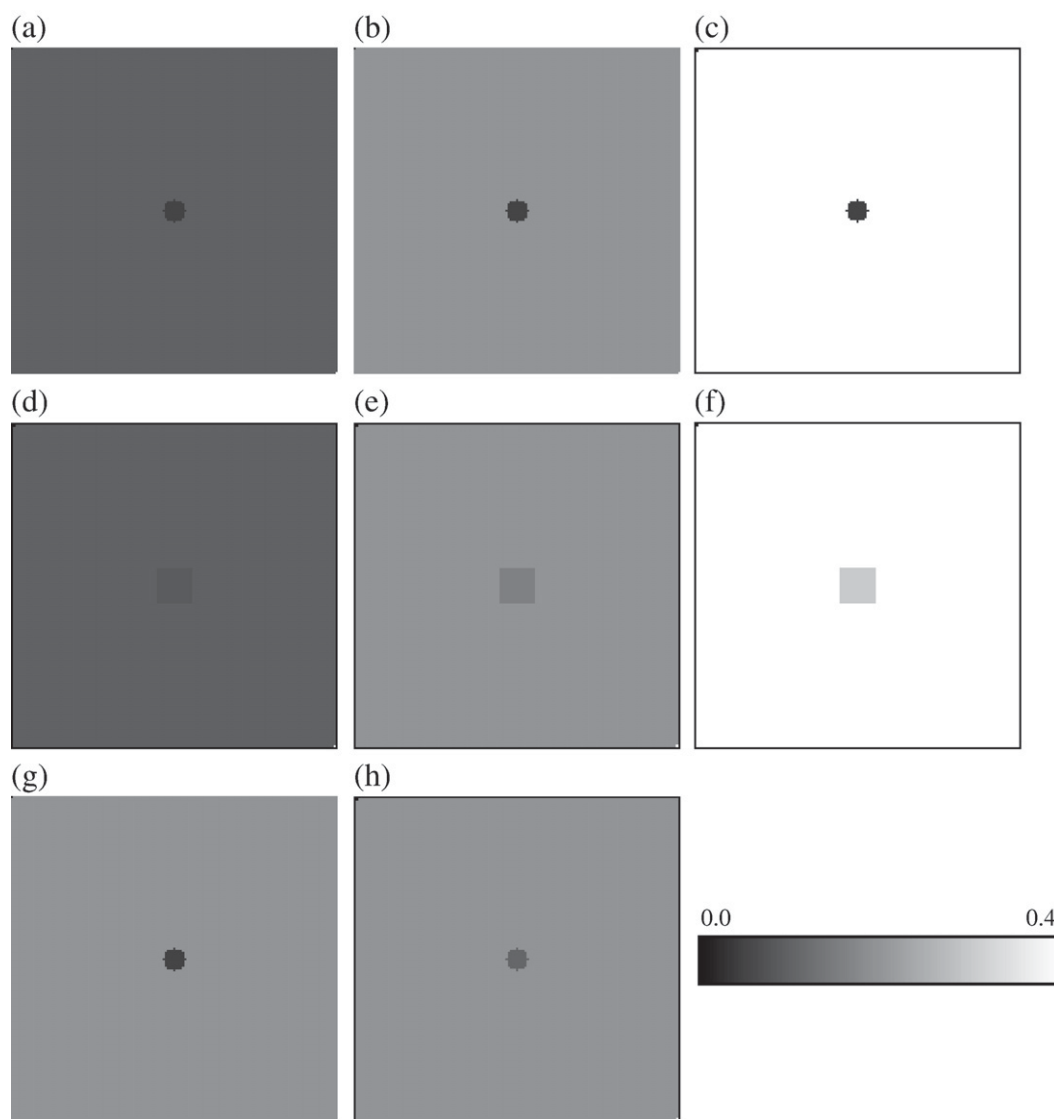
Fig. 4. Simulation test on a small object (radius of the circular object is 150 m). The coarse-resolution images (d), (e), (f) are aggregated from the fine-resolution images (a), (b), (c). Image (g) and (h) are predicted by the ESTARFM and STARFM respectively for comparison with image (b).

(b) using the ESTARFM and the STARFM respectively. The result shows that both ESTARFM and STARFM can predict the shape of the linear object well. However, from the quantitative comparison of the predicted linear object reflectance between the ESTARFM and the STARFM (Fig. 7), it can be seen that the ESTARFM can predict the reflectance of the linear object exactly, while the original STARFM only predicts that part of the linear object within the circle accurately

Fig. 5. The relationship between the prediction errors and the actual object size for the ESTARFM and STARFM.

while the segments outside the circle are predicted with increasing error.

4.2. Tests with satellite data

The ESTARFM algorithm was applied to the actual Landsat-7 ETM+ and MODIS images. In order to compare to the original STARFM algorithm, we also used the same preprocessed satellite images same as Gao et al. (2006) (http://ledaps.nascom.nasa.gov/ledaps/Tools/StarFM.htm). The resolution of Landsat-7 ETM+ and MODIS is 30 and 500 m respectively, and the bands are green, red, and NIR, which correspond with bands 2, 3, and 4 of Landsat-7 ETM+ and bands 4, 1, and 2 of MODIS.

4.2.1. Seasonal changes over forested region
The images are located around 54°N and 104°W, where the growing season is short and phenology changes rapidly (Gao et al., 2006). Three pairs of Landsat-7 ETM+ and MODIS images were acquired on May 24, 2001, July 11, 2001, and August 12, 2001 respectively. Fig. 8 shows the scenes of the images using a NIR–red–green as red–green–blue composite and identical linear stretches. The two pairs of Landsat-7 ETM+ and MODIS images acquired at May 24,
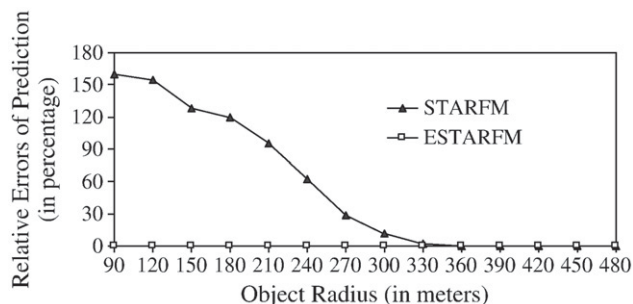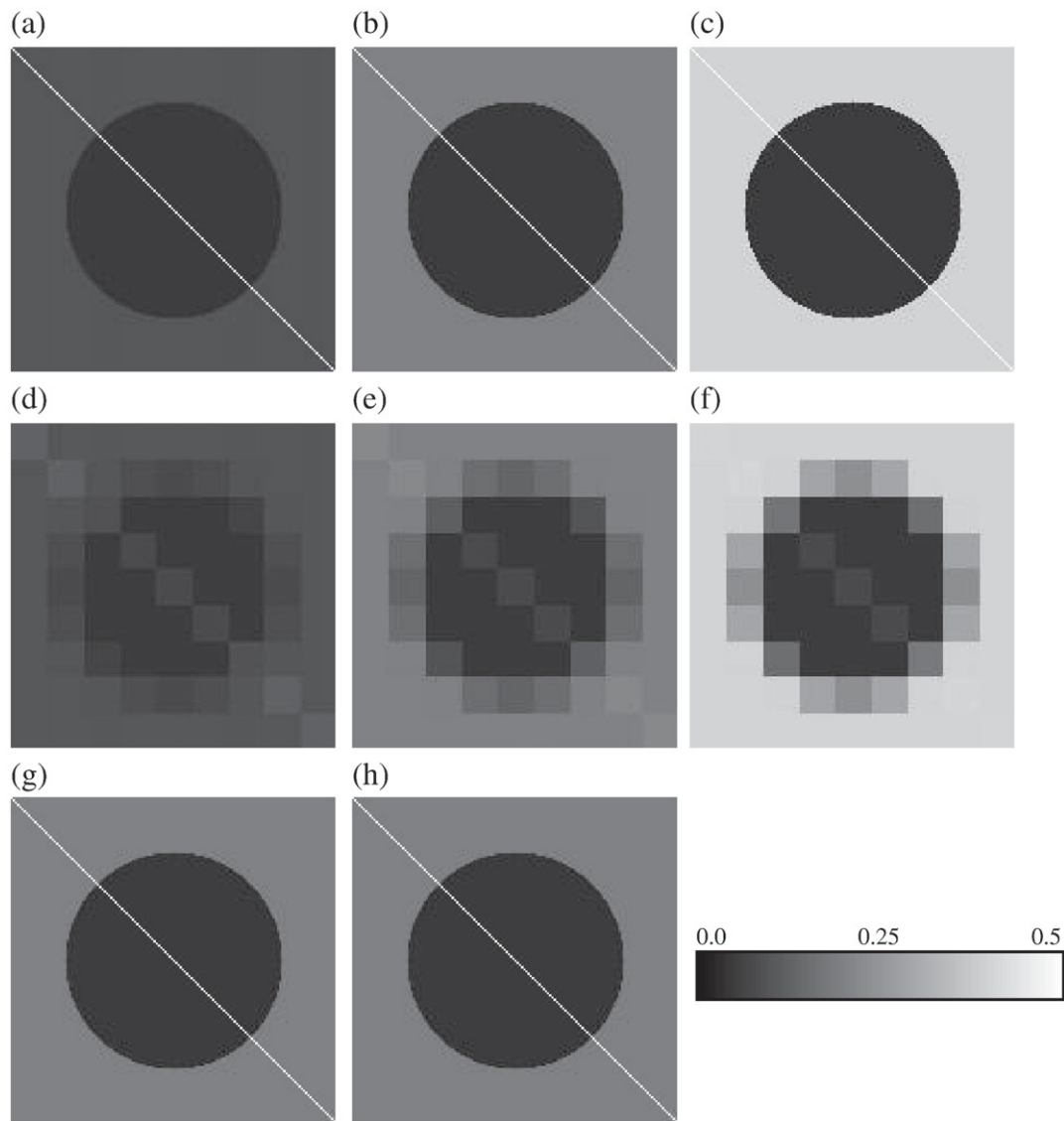
Fig. 6. Simulation test on a linear object. The coarse-resolution images (d), (e), (f) are aggregated from the fine-resolution images (a), (b), (c). Images (g) and (h) are predicted by the ESTARFM and STARFM respectively for comparison with image (b).

2001 and August 12, 2001 and the MODIS image acquired at July 11, 2001 were used to predict the image at Landsat spatial resolution at July 11, 2001. Then, the predicted image was compared with an actual Landsat ETM+ image acquired on July 11, 2001 to assess the performance of our algorithm.

Considering the landscape and land cover types of this area, the size of the search window was set as 3 MODIS pixels (50 ETM+ pixels). The number of land cover classes was set as 4. All input control

Fig. 7. The spatial profile of reflectance of the linear objects predicted by the ESTARFM and STARFM.

parameters were identical to the work in Gao et al. (2006). Fig. 9 shows the Landsat ETM+ images predicted by the ESTARFM and STARFM respectively. It is clear that the image predicted by ESTARFM (Fig. 9(b)) is very similar to the actual image (Fig. 9(a)) and contains most of the spatial details, while the image predicted by the original STARFM (Fig. 9(c)) seems somewhat "blurry" and has lost some spatial details.

Scatter plots in Fig. 10 show the relationship of reflectance between the predicted and the actual values in the July 11, 2001 Landsat ETM+ image for the NIR, red and green bands respectively. All the data in the scatter plots fall close to the 1:1 line, indicating that both STARFM and ESTARFM can capture the reflectance changes caused by phenology. In order to assess the accuracy quantitatively, the average absolute difference (*AAD*) and average difference (*AD*) of May 24, August 12 and predicted reflectance compared to real reflectance of July 11 were calculated (Table 1). The predicted surface reflectance from both algorithms has a smaller difference compared to the actual July 11 image compared to those from bracketing dates, indicating that both algorithms have successfully incorporated change information from MODIS observations to estimate the July 11 ETM+ reflectance. For the green band, the value of *AAD* is comparable for two algorithms and the prediction error of both algorithms almost has no
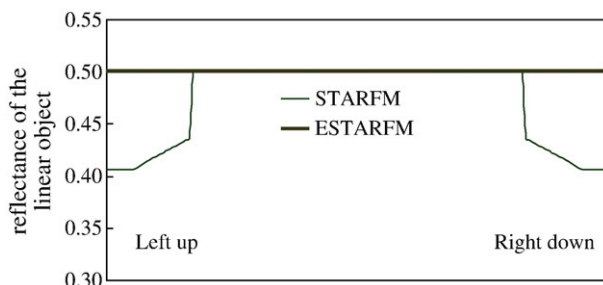
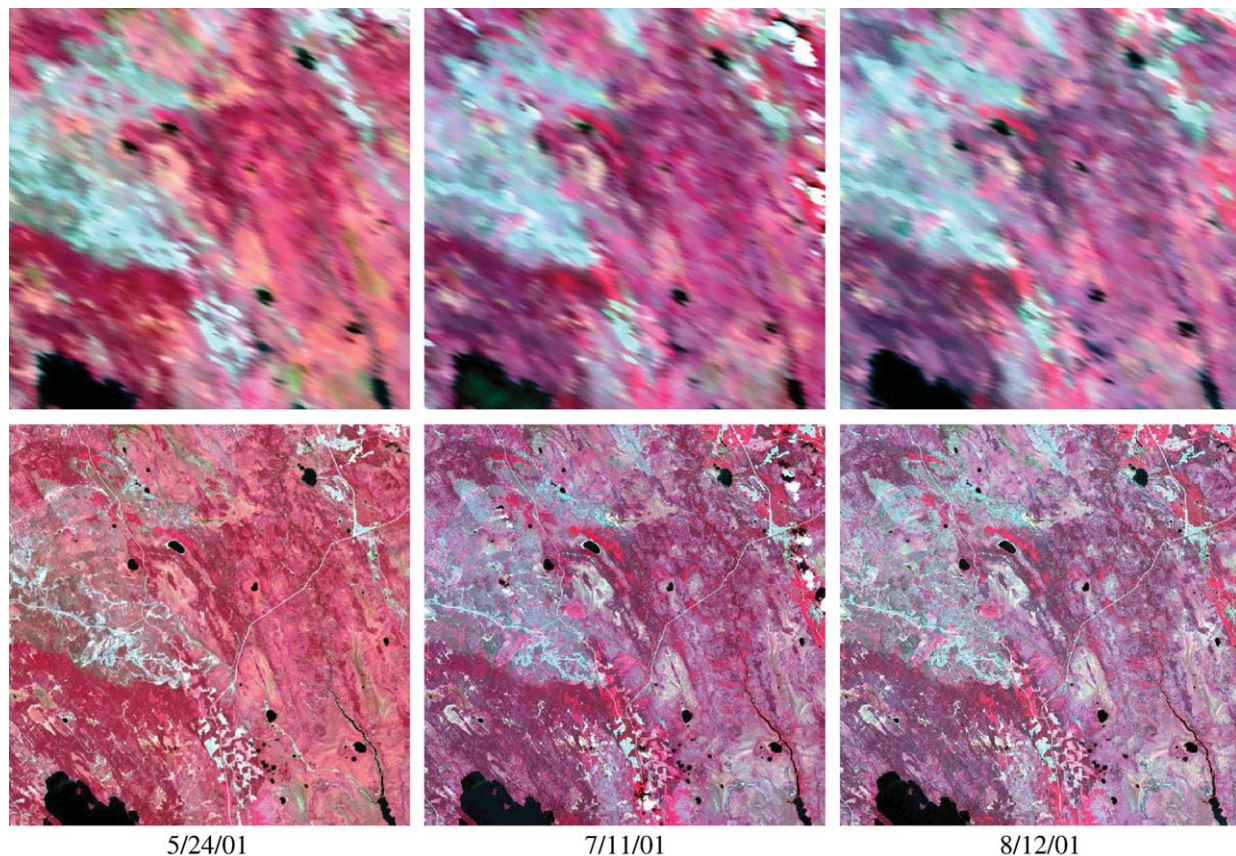5/24/01　　　　　　　　　　7/11/01　　　　　　　　　　8/12/01

Fig. 8. NIR–green–blue composites of MODIS surface reflectance (upper row) and Landsat ETM+ surface reflectance (lower row) images on May 24, 2001, July 11, 2001, and August 12, 2001 respectively.

bias (AD: −0.0002 vs. −0.0009). For the red band, the prediction error of the ESTARFM ($AAD = 0.0032$) is smaller than that of the original STARFM ($AAD = 0.0044$), meanwhile ESTARFM nearly has no bias ($AD = 0.0002$) compared with the prediction of STARFM ($AD = 0.0012$). For the NIR band, the prediction of the ESTARFM is also slightly better than the STARFM (AAD 0.0106 vs. 0.0129). However, the prediction of ESTARFM is slightly overestimated compared to that of STARFM (AD: −0.0041 vs. −0.0030). Generally, the image at Landsat ETM+ resolution predicted by the enhanced STARFM is more accurate than the original STARFM.

4.2.2. Heterogeneous (mixed) region

Heterogeneous landscapes are challenges for any data fusion algorithm. To test the enhanced STARFM model, we used a series of images covering a complex region located in central Virginia around 37°N and 77°W. The major land cover types are forest, bare soil, water and urban land. Fig. 11 shows the Landsat and MODIS images acquired on January 25, 2002, February 26, 2002, and May 17, 2002 respectively. These Landsat images were acquired during spring green-up of the vegetation. Table 2 shows that the reflectance values are more similar between January 25 and February 26, compared to the May 17 data, suggesting that the phenology of February 26 was very similar to that of January 25. From Fig. 11, we can see that there are many small patches of different land types, including forest, bare soil, and residential patches.

Fig. 12 shows the predicted images of February 26, 2002 using the two pairs of Landsat and MODIS images acquired at January 25, 2002 and May 17, 2002 and the MODIS image acquired at February 26, 2002. The fine-resolution image predicted by ESTARFM (Fig. 12(b)) successfully captures almost all of the reflectance changes for small land patches, and seems to be very close to the actual image. Since reflectance observed February 26 was close to the reflectance

observed on January 25 due to similar phenology between two dates, STARFM works well with one input date pair of January 25 as the base data (Fig. 12(c)). However the image predicted by STARFM using two input pairs resulted in an unrealistic image due to the large differences between two input pairs (Fig. 12(d)). This large difference caused a temporal "smoothing" of reflectance.

Scatter plots in Fig. 13 illustrate the difference between the predicted surface reflectance and the actual observations. We can see that the predicted surface reflectance by STARFM using one input pair and ESTARFM using two input pairs more closely match the actual observations (1:1 line) than of the results of STARFM using two input pairs. For all the three bands, the predicted errors of ESTARFM are slightly larger than those of STARFM using one input pair (AAD values: green band: 0.0068 vs. 0.0058; red band: 0.0095 vs. 0.0073; NIR band: 0.0135 vs. 0.0132), but in whole the two results predicted by the two methods are comparable. On the other hand, the prediction errors of ESTARFM are obviously smaller than STARFM using two input pairs (AAD values: green band: 0.0068 vs. 0.0075; red band: 0.0095 vs. 0.0111; NIR band: 0.0135 vs. 0.0194). For the whole image, all the predictions underestimate the surface reflectance slightly, as all the average differences (AD) of predictions are positive values (Table 2). The reason that the prediction of STARFM using one input pair is relatively more accurate than that of ESTARFM is that phenology and land cover changes are small between the prediction date and the date of input data, thus the prediction from STARFM using one input pair has less confusion than two input pairs especially when two input images change dramatically. However, if STARFM uses two pairs to predict the fine-resolution reflectance, the reflectance changes cannot be allocated well to the internal fine-resolution pixels when the coarse-resolution pixels are mixed, which causes more errors in the predicted reflectance by STARFM. On the contrary, as shown in the previous simulation tests, ESTARFM can predict the reflectance of
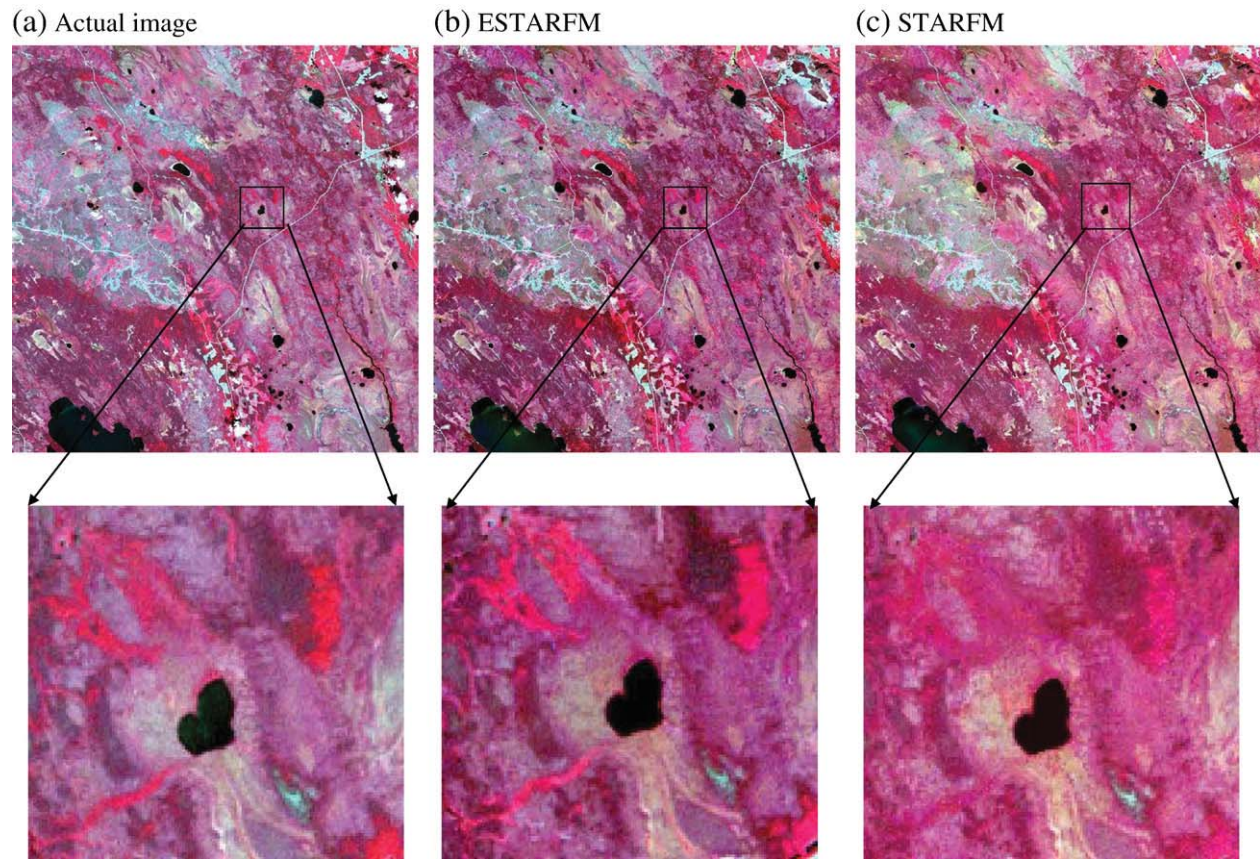
Fig. 9. The actual image observed on July 11, 2001 (a) and its prediction images by the ESTARFM (b) and STARFM (c). The lower row images are the amplifying scenes of the area marked in the upper row images.
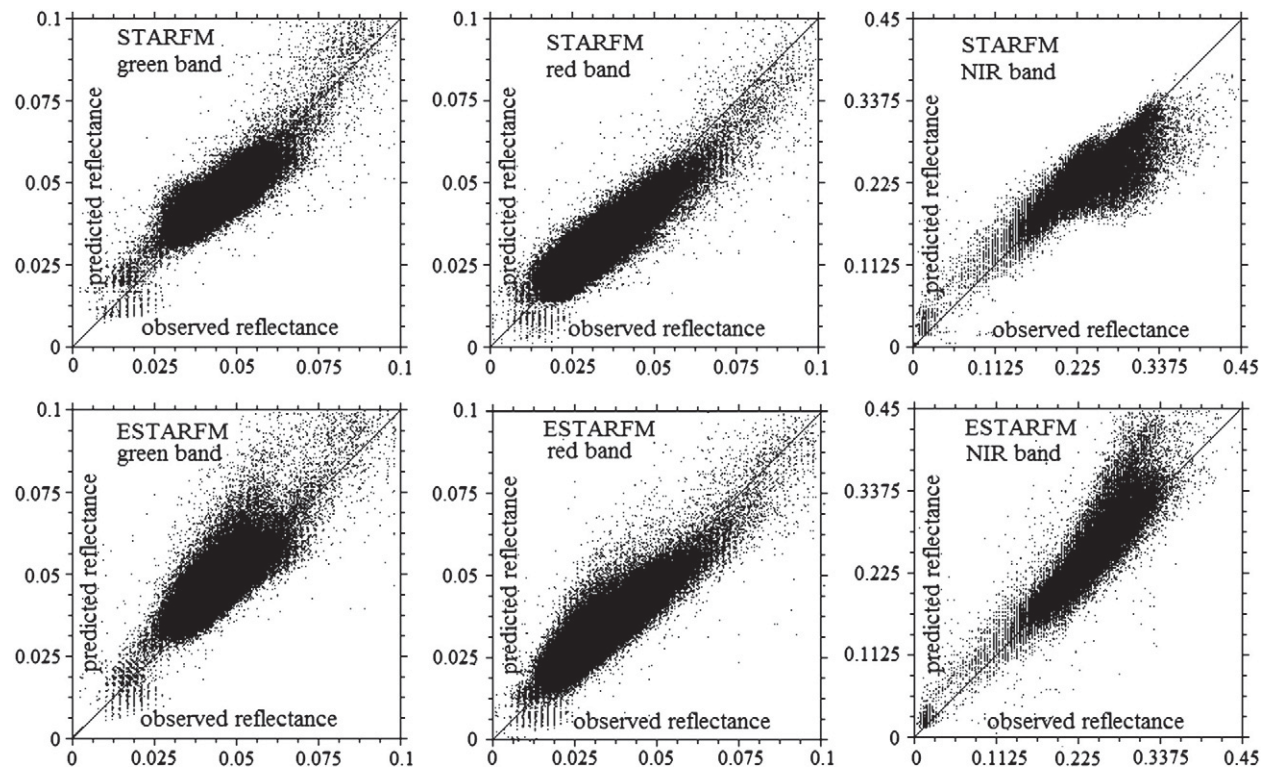

Fig. 10. Scatter plots of the real reflectance and the predicted ones product by the STARFM and ESTARFM for green, red and NIR-infrared band.

Table 1
Average absolute difference and difference of May 24, August 12 and predicted reflectance compared to real reflectance of July 11 for forested region.

ETM+	Average absolute difference (AAD)				Average difference (AD)			
Band	5/24/01	8/12/01	Prediction		5/24/01	8/12/01	Prediction	
			STARFM	ESTARFM			STARFM	ESTARFM
Green	0.0043	0.0071	0.0035	0.0035	−0.0014	0.0070	−0.0002	−0.0009
Red	0.0114	0.0058	0.0044	0.0032	−0.0111	0.0053	0.0012	0.0002
NIR	0.0443	0.0155	0.0129	0.0106	0.0441	0.0140	−0.0030	−0.0041

small objects correctly. Although many landscape patches are smaller than one MODIS pixel, ESTARFM can adjust the reflectance changes of mixed MODIS pixels to the reflectance changes of internal Landsat pixels.

Fig. 14 shows the temporal weight difference between the two input pairs (calculated from Eq. (19)) in ESTARFM, in which the negative value presents higher weight contributed by the pair of May 17, and positive value denotes that January 25 pair contributes more weight. For red and NIR band, the predicted reflectance of most pixels is mainly dependent on the information from January 25. This is consistent with the fact that there is smaller phenological difference between the reflectance observed on January 25 and February 26 (Table 2). For green band, the temporal weights from the two input pairs are comparable, which are consistent with the results shown in Table 2 (0.0119 vs. 0.0107). It is evident that the enhanced STARFM decides the contribution from two input pairs automatically and can be operational with two input pairs, while the selection of one input pair in original STARFM lacks of quantitative measures, so the prediction errors of ESTARFM are smaller than the original STARFM using two input pairs.

5. Conclusion and discussion

This study described the theoretical basis, implementation process and performance of an enhanced STARFM (ESTARFM) fusion algorithm to blend the multi-source remotely sensed data. Compared to the original STARFM algorithm, this improved algorithm can produce a synthetic fine-resolution reflectance product more accurately, especially for heterogeneous landscapes.

The ESTARFM has made several improvements to the original STARFM algorithm. The most significant improvement of the ESTARFM is using a conversion coefficient to enhance the accuracy of prediction for heterogeneous landscapes. The STARFM has a limitation in predicting the reflectance of objects with size significantly smaller than the coarse-resolution pixel if homogeneous coarse-resolution pixels cannot be found in the search window, since the STARFM assumes that the reflectance change of coarse-resolution pixel equals the change of the fine-resolution pixels within it. This assumption is reasonable when the coarse-resolution pixels are pure (homogeneous), but may not hold when the coarse-resolution pixels are mixed. In reality, it may be hard to find homogeneous coarse-resolution pixels for
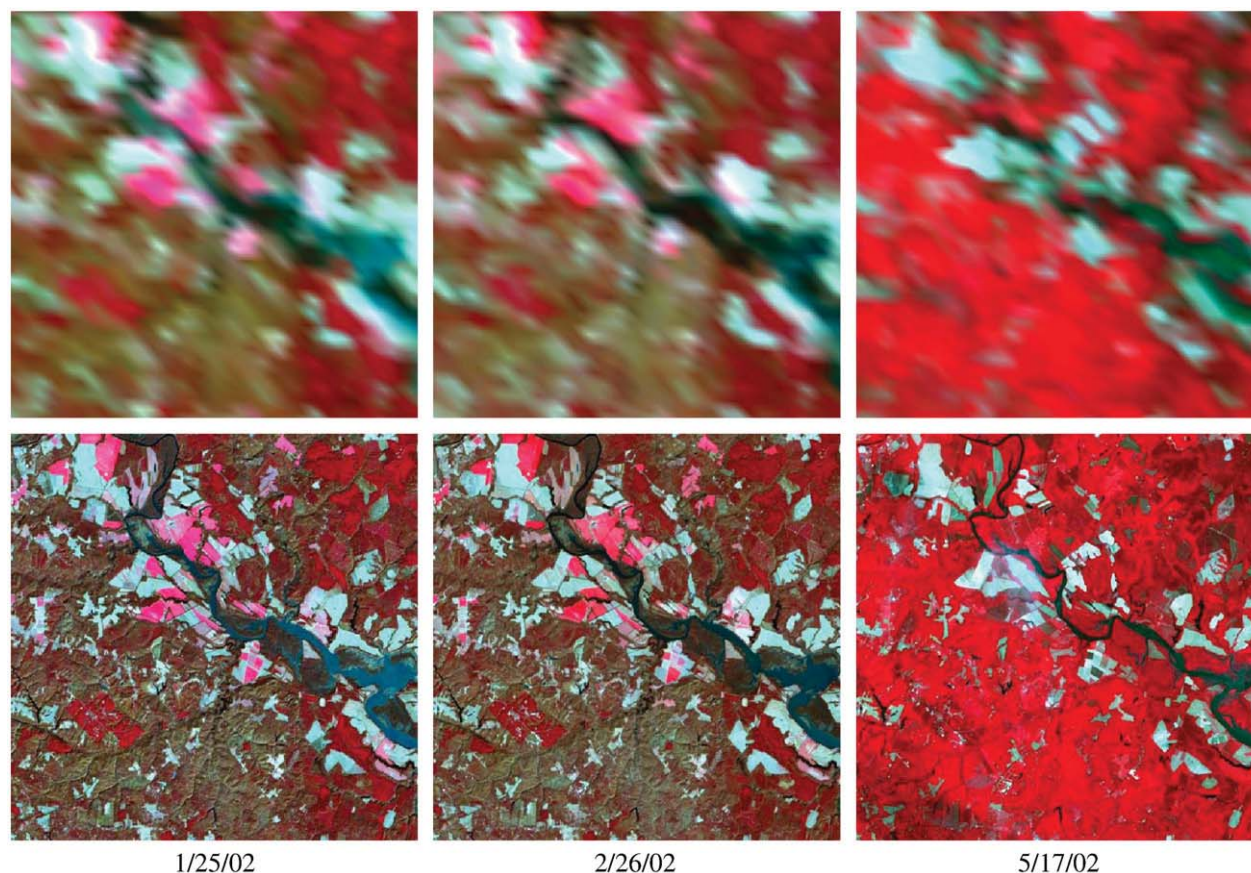

Fig. 11. NIR–green–blue composites of MODIS surface reflectance (upper row) and Landsat ETM+ surface reflectance (lower row) images on January 25, February 26, and May 17, 2002 respectively.

Table 2
Average absolute difference and difference of January 25, May 17 and predicted reflectance compared to real reflectance on February 26 for the complex mixture region.

ETM+	Average absolute difference (AAD)					Average difference (AD)				
Band	1/25/02	5/17/02	Prediction			1/25/02	5/17/02	Prediction		
			STARFM[a]	STARFM[b]	ESTARFM			STARFM[a]	STARFM[b]	ESTARFM
Green	0.0119	0.0107	0.0058	0.0075	0.0068	0.0113	−0.0031	0.0007	0.0026	0.0028
Red	0.0150	0.0283	0.0073	0.0111	0.0095	0.0143	0.0218	0.0013	0.0040	0.0021
NIR	0.0279	0.1774	0.0132	0.0196	0.0135	0.0269	−0.1722	0.0019	0.0060	0.0022

[a] The prediction of STARFM using only one pair images on January 25.
[b] The prediction of STARFM using the two pair images.

the 500×500 m MODIS pixel size. The ESTARFM employs a conversion coefficient to convert the reflectance changes of a mixed coarse-resolution pixel to the fine-resolution pixels within it, ensuring an accurate prediction of reflectance for small objects and linear objects.

Secondly, the ESTARFM improves the accuracy of selecting similar pixels. Selecting similar pixels in a search window is an important step both in the ESTARFM and STARFM algorithm, because their information will be included in the prediction of reflectance for the central pixel. In the original STARFM algorithm, only two bands (red and NIR) are used to identify the similar pixels and the similar pixels are selected from the fine-resolution images acquired at different dates independently. In the ESTARFM algorithm, all the bands are used to select the similar pixels and the intersection of the selected results from all fine-resolution images are extracted. The spectra of many objects on the land surface changes through time, leading to spectral features that may be confused with other objects at various times. Extracting the intersection of the selected results of different date can ensure that the right similar pixel will be selected with same spectral trajectory.

Thirdly, for the weight calculation of each similar pixel, the ESTARFM uses spectral similarity (correlation coefficient) between fine- and coarse-resolution pixel to represent the homogeneity of a coarse-resolution pixel instead of spectral distance in the original STARFM. Using the correlation coefficient of the spectra between fine and coarse resolution at all observed dates to identify the homogeneity of a coarse-resolution pixel not only can avoid some errors in absolute reflectance value calculation stemming from radiometric calibration and atmospheric correction, but can also introduce information on the spectral trajectory into the weight calculation.

Last, the STARFM uses weights to average the prediction of the fine-resolution reflectance of all similar pixels in the search window to obtain the reflectance of central pixel, while the ESTARFM uses weights to combine the reflectance trajectories of all similar pixels. This change is added to the fine-resolution reflectance observed at
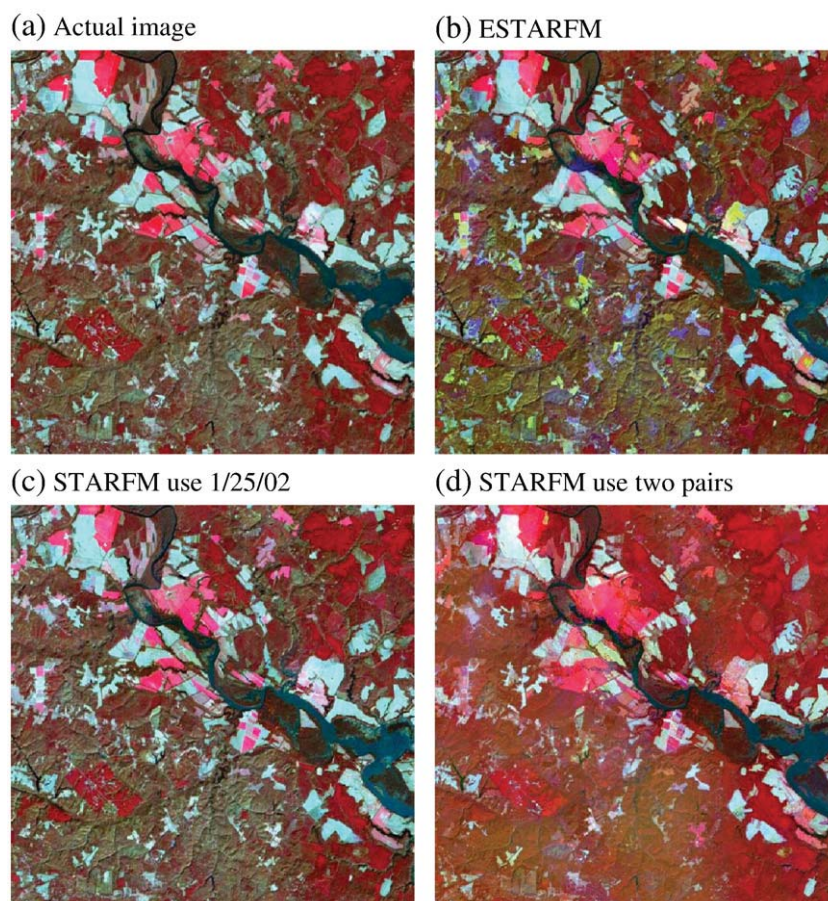
(a) Actual image

(b) ESTARFM

(c) STARFM use 1/25/02

(d) STARFM use two pairs

Fig. 12. The actual Landsat image observed on February 26, 2002 (a) and its prediction images by the ESTARFM (b), STARFM using one input pair from January 25 (c) and STARFM using two input pairs (d).
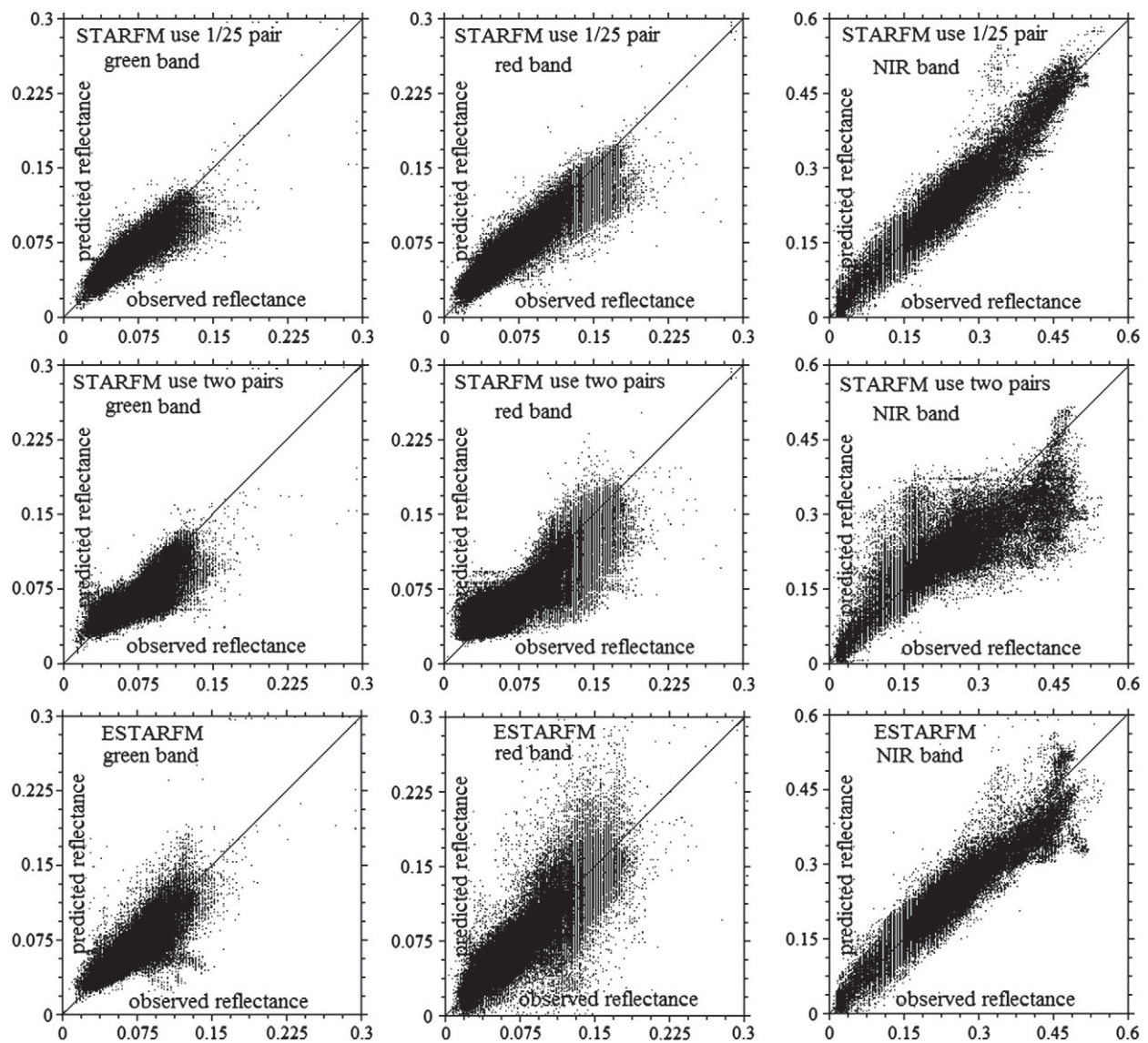
Fig. 13. Scatter plots of the real reflectance and the predicted ones product by the STARFM and ESTARFM for green, red and NIR-infrared band.
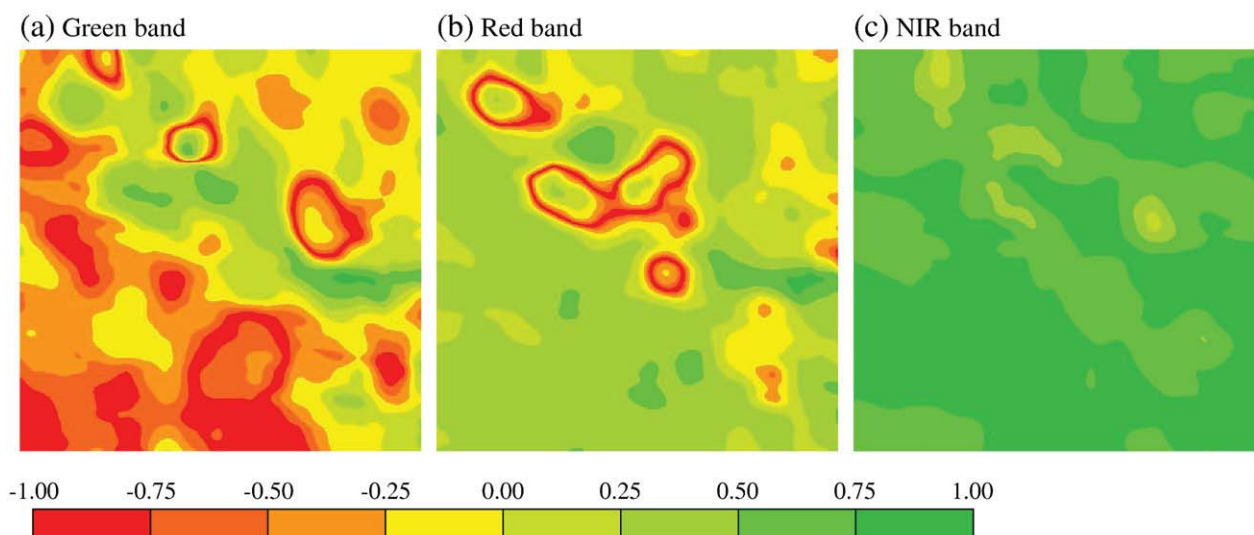

Fig. 14. The temporal weight difference between the two input pairs in ESTARFM: (a) green band, (b) red band, and (c) NIR band.

base date to predict the fine-resolution reflectance of central pixel at date t_p. The weight averaging function in STARFM leads to predicted fine-resolution images that seem "hazy", and smoothes some spatial details (Fig. 9(c)). On the contrary, the ESTARFM is successful in keeping the spatial details (Fig. 9(b)), because the main part of the predicted reflectance is provided by the observed fine-resolution reflectance of the central pixel at the base date, and thus the contrast between the central pixel and the neighbor pixels can be preserved.

There are several limitations and constraints while using the enhanced STARFM approach. First, similar to STARFM, it cannot accurately predict objects for which shape changes with time and will thus blur the changing boundary. Secondly, it also cannot accurately predict short-term, transient changes that are not recorded in any of the bracketing fine-resolution images, therefore combining ESTARFM with the STAARCH algorithm (Hilker et al., 2009a) may be a feasible way to enhance the capability of the new algorithm. Thirdly, sensors with different spectral band passes may lead to nonlinear relationships. In our study, MODIS and Landsat data show good linear agreements. However, extra attention may need to be paid when the ESTARFM is used for other sensors. Also, the assumption that the rate of reflectance linear change is constant might be not appropriate in some situation, especially during a long period. Therefore, it is better to use the images acquired near the prediction time to retrieve the unknown fine-resolution reflectance. Fourth, there are two parameters that should be set in ESTARFM, the size of moving window and the number of classes, which might limit automated processing. In practical applications, we can set the parameters according to the homogeneity of land surface observed from Landsat images. If we need massive processing (e.g., at global scale), global land cover maps will be helpful to determine the parameters adaptively. Lastly, compared with the original STARFM algorithm, the ESTARFM may be more computationally intensive and require at least two pairs of fine- and coarse-resolution images acquired at the same date, which is more than the required by the original STARFM. Accuracy of prediction may depend on the selection of input image pairs. More frequent intra-annual imageries to bracket all vegetation phenology changes are helpful. However, in some cloudy regions where it is difficult to acquire two high-quality input pairs simultaneously, the original STARFM may be more appropriate. Though we demonstrated that a single input pair can produce accurate prediction from original STARFM, the accuracy of these predictions depends on the similarity of single input pair to the prediction date. The enhanced STARFM can weight input pairs based on the similarity to the coarse-resolution data on the prediction date and thus is more robust when two input pairs are used.

In conclusion, the ESTARFM algorithm advances the capability for producing remotely sensed data products with both high spatial resolution and frequent coverage from multi-source satellite data. Such a capability is helpful for monitoring intra-annual land surface and ecological dynamics at the spatial scales most relevant to human activities.

Acknowledgements

References

Adams, J. B., Smith, M. O., & Johnson, P. E. (1985). Spectral mixture modeling: A new analysis of rock and soil types at the Viking Lander I site. *Journal of Geophysical Research-Atmosphere*, *91*, 8089–8112.

Asner, G. P. (2001). Cloud cover in Landsat observations of the Brazilian Amazon. *International Journal of Remote Sensing*, *22*, 3855–3862.

Camps-Valls, G., Gomez-Chova, L., Munoz-Mari, J., Rojo-Alvarez, J. L., & Martinez-Ramon, M. (2008). Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection. *IEEE Transactions on Geoscience and Remote Sensing*, *46*, 1822–1835.

Carper, W. J., Lilles, T. M., & Kiefer, R. W. (1990). The use of intensity–hue-saturation transformations for merging SPOT panchromatic and multispectral image data. *Photogrammetric Engineering and Remote Sensing*, *56*, 459–467.

Cohen, W. B., & Goward, S. N. (2004). Landsat's role in ecological applications of remote sensing. *Bioscience*, *54*, 535–545.

Gao, F., Masek, J., Schwaller, M., & Hall, F. (2006). On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance. *IEEE Transactions on Geoscience and Remote Sensing*, *44*, 2207–2218.

Gao, F., Masek, J., & Wolfe, R. (2009). An automated registration and orthorectification package for Landsat and Landsat-like data processing. *Journal of Applied Remote Sensing*, *3*, 033515. doi:10.1117/1.3104620

Gonzalez-Sanpedro, M. C., Le Toan, T., Moreno, J., Kergoat, L., & Rubio, E. (2008). Seasonal variations of leaf area index of agricultural fields retrieved from Landsat data. *Remote Sensing of Environment*, *112*, 810–824.

Healey, S. P., Cohen, W. B., Yang, Z. Q., & Krankina, O. N. (2005). Comparison of Tasseled Cap-based Landsat data structures for use in forest disturbance detection. *Remote Sensing of Environment*, *97*, 301–310.

Hilker, T., Wulder, M. A., Coops, N. C., Linke, J., McDermid, G., Masek, J., et al. (2009). A new data fusion model for high spatial- and temporal-resolution mapping of forest based on Landsat and MODIS. *Remote Sensing of Environment*, *113*, 1613–1627.

Hilker, T., Wulder, M. A., Coops, N. C., Sritz, N., White, J. C., Gao, F., et al. (2009). Generation of dense time series synthetic Landsat data through data blending with MODIS using a spatial and temporal adaptive reflectance fusion model. *Remote Sensing of Environment*, *113*, 1988–1999.

Jorgensen, P. V. (2000). Determination of cloud coverage over Denmark using Landsat MSS/TM and NOAA-AVHRR. *International Journal of Remote Sensing*, *21*, 3363–3368.

Ju, J. C., & Roy, D. P. (2008). The availability of cloud-free Landsat ETM plus data over the conterminous United States and globally. *Remote Sensing of Environment*, *112*, 1196–1211.

Marfai, M. A., Almohammad, H., Dey, S., Susanto, B., & King, L. (2008). Coastal dynamic and shoreline mapping: Multi-sources spatial data analysis in Semarang Indonesia. *Environmental Monitoring and Assessment*, *142*, 297–308.

Masek, J. G., & Collatz, G. J. (2006). Estimating forest carbon fluxes in a disturbed southeastern landscape: Integration of remote sensing, forest inventory, and biogeochemical modeling. *Journal of Geophysical Research-Biogeosciences*, *111*.

Masek, J. G., Huang, C. Q., Wolfe, R., Cohen, W., Hall, F., Kutler, J., et al. (2008). North American forest disturbance mapped from a decadal Landsat record. *Remote Sensing of Environment*, *112*, 2914–2926.

Masek, J. G., Vermote, E. F., Saleous, N. E., Wolfe, R., Hall, F. G., Huemmrich, F., et al. (2006). A Landsat surface reflectance data set for North America, 1990–2000. *IEEE Geoscience and Remote Sensing Letters*, *3*(1), 69–72.

Pohl, C., & Van Genderen, J. L. (1998). Multisensor image fusion in remote sensing: Concepts, methods and applications. *International Journal of Remote Sensing*, *19*, 823–854.

Price, J. C. (1994). How unique are spectal signatures? *Remote Sensing of Environment*, *49*, 181–186.

Ranson, K. J., Kovacs, K., Sun, G., & Kharuk, V. I. (2003). Disturbance recognition in the boreal forest using radar and Landsat-7. *Canadian Journal of Remote Sensing*, *29*, 271–285.

Roy, D. P., Ju, J., Lewis, P., Schaaf, C., Gao, F., Hansen, M., et al. (2008). Multi-temporal MODIS-Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data. *Remote Sensing of Environment*, *112*, 3112–3130.

Shabanov, N. V., Wang, Y., Buermann, W., Dong, J., Hoffman, S., Smith, G. R., et al. (2003). Effect of foliage spatial heterogeneity in the MODIS LAI and FPAR algorithm over broadleaf forests. *Remote Sensing of Environment*, *85*, 410–423.

Shettigara, V. K. (1992). A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set. *Photogrammetric Engineering and Remote Sensing*, *58*, 561–567.

Woodcock, C. E., & Ozdogan, M. (2004). Trends in land cover mapping and monitoring. In Gutman (Ed.), *Land Change Science* (pp. 367–377). New York: Springer.

Yocky, D. A. (1996). Multiresolution wavelet decomposition image merger of Landsat Thematic Mapper and SPOT panchromatic data. *Photogrammetric Engineering and Remote Sensing*, *62*, 1067–1074.

Zhang, Y. (2004). Understanding image fusion. *Photogrammetric Engineering and Remote Sensing*, *70*, 657–661.

Zurita-Milla, R., Kaiser, G., Clevers, J. G. P. W., Schneider, W., & Schaepman, M. E. (2009). Downscaling time series of MERIS full resolution data to monitor vegetation seasonal dynamics. *Environmental Monitoring and Assessment*, *113*, 1874–1885.