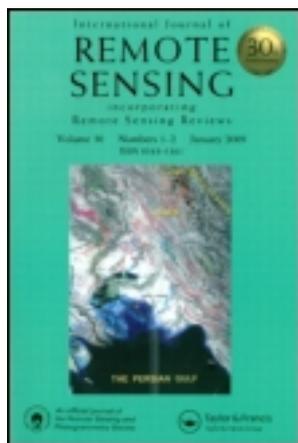


This article was downloaded by: [Wuhan University]

On: 21 March 2013, At: 17:48

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



International Journal of Remote Sensing

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/tres20>

A spatial and temporal reflectance fusion model considering sensor observation differences

Huanfeng Shen ^a, Penghai Wu ^a, Yaolin Liu ^a, Tinghua Ai ^a, Yi Wang ^b & Xiaoping Liu ^c

^a Department of Map Science and Geographical Information Engineering, School of Resource and Environmental Science, Wuhan University, Wuhan, China

^b Department of Geo-information Science and Technology, Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan, China

^c Department of Remote Sensing and Geographical Information Engineering, School of Geography and Planning, Sun Yat-sen University, Guangzhou, China

Version of record first published: 21 Mar 2013.

To cite this article: Huanfeng Shen, Penghai Wu, Yaolin Liu, Tinghua Ai, Yi Wang & Xiaoping Liu (2013): A spatial and temporal reflectance fusion model considering sensor observation differences, International Journal of Remote Sensing, 34:12, 4367-4383

To link to this article: <http://dx.doi.org/10.1080/01431161.2013.777488>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings,

demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

A spatial and temporal reflectance fusion model considering sensor observation differences

Huanfeng Shen^{a*}, Penghai Wu^a, Yaolin Liu^a, Tinghua Ai^a, Yi Wang^b, and Xiaoping Liu^c

^aDepartment of Map Science and Geographical Information Engineering, School of Resource and Environmental Science, Wuhan University, Wuhan, China; ^bDepartment of Geo-information Science and Technology, Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan, China; ^cDepartment of Remote Sensing and Geographical Information Engineering, School of Geography and Planning, Sun Yat-sen University, Guangzhou, China

(Received 16 March 2012; accepted 13 December 2012)

This article proposes a spatial–temporal expansion method for remote-sensing reflectance by blending observations from sensors with different spatial and temporal characteristics. Compared with the methods used in the past, the main characteristic of the proposed method is consideration of sensor observation differences between different cover types when calculating the weight function of the fusion model. The necessity of the temporal difference factor commonly used in spatial–temporal fusion is also analysed in this article. The method was tested and quantitatively assessed under different landscape situations. The results indicate that the proposed fusion method improves the prediction accuracy of fine-resolution reflectance.

1. Introduction

In the remote-sensing community, there is an increasing need to integrate Earth observation data with complementary spatial, temporal and spectral characteristics (Zurita-Milla et al. 2009). As a result, image fusion techniques have received more and more attention, and a large number of data fusion methods have been developed (for a review, see Ehlers 1991; Hall and Llinas 1997; Luo, Chih-chen, and Kuo Lan 2002; Pohl and Van Genderen 1998). Classical remote-sensing image fusion techniques include panchromatic/multispectral (PAN/MS) fusion (Thomas et al. 2008; Wang et al. 2005), multispectral/hyperspectral (MS/HS) fusion (Eismann and Hardie 2005), and multitemporal fusion (Shen et al. 2009). These fusion techniques are aimed at extracting information and enhancing visual effects. However, they are not effective in enhancing temporal resolution and spatial resolution simultaneously and cannot provide high spatial resolution observations with a dense time series.

Gao et al. (2006) developed a filter-based framework to blend Landsat and MODIS (Moderate Resolution Imaging Spectroradiometer) data for predicting daily surface reflectance at Landsat spatial resolution and MODIS temporal frequency. Their framework is called the spatial and temporal adaptive reflectance fusion model (STARFM). The STARFM algorithm has been shown to be a relatively reliable model for generating synthetic Landsat images, which may then be used to investigate vegetation dynamics in

*Corresponding author. Email: shenhf@whu.edu.cn

different land-cover types (Hilker et al. 2009b). Watts et al. (2011) improved the classification accuracy of conservation tillage adoption by incorporating STARFM-based synthetic data sets into the classification model. The STARFM fusion method has also been applied to the generation of gross primary productivity (Singh 2010), the analysis of dryland forest phenology (Walker et al. 2011), the examination of virus dissemination (Liu and Weng 2011), and the estimation of daily evapotranspiration (Anderson et al. 2011). Hilker et al. (2009a) and Zhu et al. (2010) extended the STARFM method for cases where there are two or more base image pairs as inputs to the simulation of surface disturbance and heterogeneous fine-grained landscapes, respectively.

In other frameworks, Roy et al. (2008) proposed a semi-physical fusion framework, which uses the MODIS bidirectional reflectance distribution function (BRDF)/albedo land surface product and Landsat Enhanced Thematic Mapper Plus (ETM+) data to predict ETM+ reflectance on the same, an antecedent, or subsequent date. Hansen et al. (2008) used regression trees to fuse Landsat and MODIS data based on the 500 m, 16 day MODIS BRDF/albedo land surface characterization product. The method demands a single ‘best’ image to map forest-cover status for a given year or decade. Zurita-Milla et al. (2009) proposed an unmixing-based Landsat Thematic Mapper (TM) and Medium Resolution Imaging Spectrometer MERIS full resolution (FR) data fusion algorithm, which requires a high-resolution land-use database for pixel unmixing.

In this article, an extended spatial–temporal reflectance fusion method is presented. Unlike the works of Hilker et al. (2009a) and Zhu et al. (2010), which are based on two base image pairs, we concentrate on improving the STARFM method in the single-pair case. The main characteristic of this method is the consideration of sensor observation differences based on within-class regression to improve fusion accuracy. This study also validates removal of the measure of temporal difference from the spatial–temporal fusion model.

The rest of this article is structured as follows. Section 2 reviews the original STARFM framework. The proposed method is presented in Section 3. Experimental results and quantitative evaluations are shown in Section 4. In Section 5, we provide some concluding remarks.

2. STARFM framework

The premise of STARFM is that the preprocessed and corrected reflectances of different sensors are consistent and comparable (Gao et al. 2006; Masek et al. 2006). For a homogenous coarse-resolution (MODIS) pixel, the corresponding fine-resolution (ETM+) reflectance can be expressed as (Gao et al. 2006)

$$F(x_i, y_j, t_0) = C(x_i, y_j, t_0) + \varepsilon_0, \quad (1)$$

where F and C denote fine- and coarse-resolution reflectance, respectively, (x_i, y_j) denotes the pixel location, t_0 is the acquisition date, and ε_0 is the residual. Similarly, the fine-resolution (ETM+) reflectance at predicted date t_1 can be expressed as

$$F(x_i, y_j, t_1) = C(x_i, y_j, t_1) + \varepsilon_1. \quad (2)$$

Supposing the residual $\varepsilon_0 = \varepsilon_1$, the following equation can be obtained:

$$F(x_i, y_j, t_1) = C(x_i, y_j, t_1) + F(x_i, y_j, t_0) - C(x_i, y_j, t_0). \quad (3)$$

Considering neighbouring same-class pixels with similar reflectance changes, a moving window method is used to take full advantage of the information from neighbouring pixels. The fine-resolution (ETM+) reflectance is calculated with a weight function (Gao et al. 2006),

$$F(x_{w/2}, y_{w/2}, t_1) = \sum_{i=1}^w \sum_{j=1}^w W_{ij} (C(x_i, y_j, t_1) + F(x_i, y_j, t_0) - C(x_i, y_j, t_0)), \quad (4)$$

where w is the size of the moving window and $(x_{w/2}, y_{w/2})$ is the central pixel. It should be noted that only spectrally similar and cloud-free pixels within the moving window are selected to calculate reflectance.

The spatial weighting function W_{ij} determines by how much each similar pixel contributes to the central pixels. Ensuring the sum of all weights is 1, W_{ij} can be normalized as

$$W_{ij} = \frac{1}{E_{ij}} \bigg/ \sum_i^w \sum_j^w (1/E_{ij}), \quad (5)$$

where E_{ij} may be related to spectral difference, temporal difference, and geometric distance to the central pixels in the moving window and can be expressed as

$$E_{ij} = \ln(S_{ij}A + 1) \ln(T_{ij}A + 1) D_{ij}, \quad (6)$$

where A is a scale factor (equal to 10,000 when using MODIS reflectance products, which linearly scale reflectance from 0 to 10,000) (Gao et al. 2006), S_{ij} represents the spectral difference, T_{ij} represents the temporal difference, and D_{ij} represents the geometric distance. Equations (5) and (6) show that one pixel may be assigned a higher weight W_{ij} if the pixel has lower values of S_{ij} , T_{ij} , and D_{ij} .

3. Proposed method

3.1. Calculating spectral difference S_{ij}

Although the observation data can be preprocessed before the fusion process, systematic observation differences may still exist between the different sensors. The observation differences may come from differences in spectral range and/or calibration algorithms. To minimize the effects of observation differences on fusion results, this study proposes to fit regression curves for different cover types when measuring spectral difference.

The original STARFM calculates the spectral difference as

$$S_{ij} = |F(x_i, y_j, t_0) - C(x_i, y_j, t_0)|. \quad (7)$$

It would be assigned a higher weight when a pixel has a lower value of S_{ij} , because this means the fine-resolution pixel has spectral features closer to the averaged surrounding pixels. However, (7) fails to consider the observation differences of fine- and coarse-resolution sensors, which may lead to inaccurate estimation. For example, assume that the reflectances of river water retrieved by the fine- and coarse-resolution sensors are, respectively, 0.1 and 0.09 under normal conditions. Furthermore, assume that there are two pixels in the moving window, and that their spectral differences are 0.005 and 0.01, respectively. According

to (7), the pixel whose spectral difference is 0.005 will be assigned a higher weight. In fact, since the value of 0.01 is the normal difference between the two sensors, the pixel whose spectral difference is 0.01 should be assigned a higher weight. In order to consider observation differences, a simple linear regression relationship between coarse-resolution (MODIS) and the corresponding fine-resolution (ETM+) reflectance at time t can be described as

$$C(x_i, y_j, t) = aF(x_i, y_j, t) + b, \quad (8)$$

where a is the gain parameter of the linear regression function between coarse- and fine-resolution pixels and b is the bias. Thus, the spectral difference can be calculated as

$$S_{ij} = |aF(x_i, y_j, t_0) + b - C(x_i, y_j, t_0)|. \quad (9)$$

It is worth noting that the linear regression coefficients (a and b) between the two sensors may vary significantly for different cover types. To confirm this, an experiment was performed. Figure 1(a) shows a Landsat ETM+ sub-image with NIR-red-green composites, and Figure 1(b) shows the corresponding MODIS image. From the two images, three cover types of water, vegetation, and bare area were manually selected to perform a linear fitting for each band, and the fitted parameters are shown in Table 1. We can recognize that the linear relationship between MODIS and ETM+ reflectance is significantly related to land-cover type.

To solve this problem, the spectral difference Equation (9) can be rewritten as

$$S_{ijc} = |a_c F_c(x_i, y_j, t_0) + b_c - C_c(x_i, y_j, t_0)|, \quad (10)$$

where the subscript c represents land-cover types, which can be obtained by classifying the fine-resolution image. The classification provides the possibility of using the regression coefficients for different cover types. Our experiments indicated that the proposed

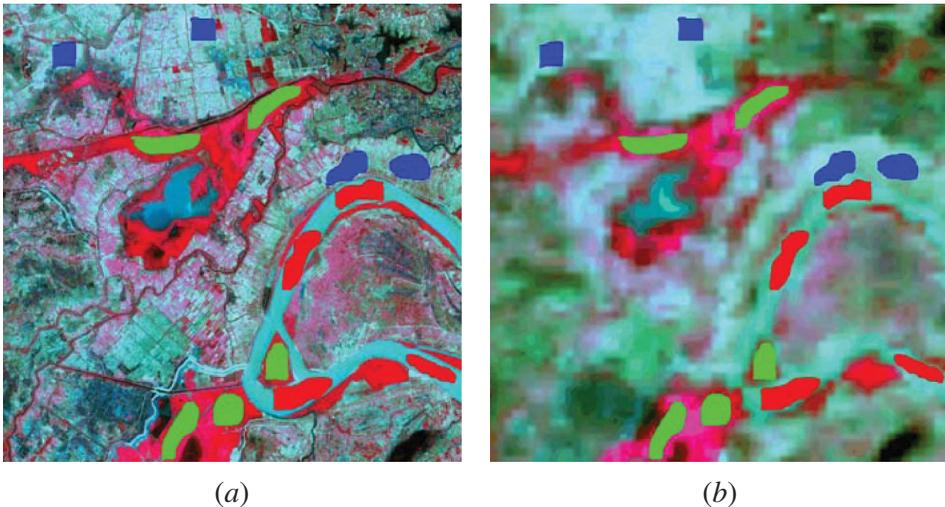


Figure 1. Images from selected regions for the linear fitting test. (a) ETM+ image, (b) MODIS image (resampled).

Table 1. Quantitative linear fitting results of the ETM+ and MODIS images for different cover types and spectral bands. The results represent three cover types for the NIR, red, and green bands, respectively.

	Water		Vegetation		Bare	
	a	b	a	b	a	b
NIR	0.11	0.09	0.47	0.17	0.83	0.02
Red	0.13	0.11	0.76	0.05	0.60	0.05
Green	-0.15	0.15	0.56	0.08	0.49	0.07

algorithm is not sensitive to the choice of classification algorithm. It is also not sensitive to class number when the number is greater than 5 (see the experiment part). For convenience, the unsupervised IsoData classification method was employed, and the number of classes was set as seven in all experiments described in this article.

Once the cover map is determined using the classification algorithm, the linear regression coefficients for each cover type and spectral band can be solved by various optimization estimation methods. The simplest method is the least squares approach, although this is sensitive to the distribution assumption of noise and the existence of outliers. When noise distribution is abnormal and there are outliers in the observations, the estimation may be unsatisfactory (Wu and Tam 2001). In order to improve the accuracy of the parameters estimated, we used the M-estimation method with the Huber cost function (Huber 1964) to solve the coefficients.

Let e_{ij} be the residual of the ij_{th} data point:

$$e_{ij} = a_c F_c(x_i, y_j, t_0) + b_c - C_c(x_i, y_j, t_0). \tag{11}$$

The standard least squares method tries to minimize $\sum_{ij} e_{ij}^2$, but M-estimators try to reduce the effect of outliers by replacing the squared residuals e_{ij}^2 by another function of the residuals (Fox and Weisberg 2011):

$$\min \sum_{ij} \rho(e_{ij}), \tag{12}$$

where $\rho(\bullet)$ is a symmetric, positive-definite cost function with a unique minimum of zero. There are many different kinds of cost function, but here we chose the Huber cost function:

$$\rho(e_{ij}) = \begin{cases} e_{ij}^2/2 & |e_{ij}| \leq k \\ k|e_{ij}| - k^2/2 & |e_{ij}| > k \end{cases}, \tag{13}$$

where k is the Huber parameter.

In order to facilitate the optimization process, the minimization problem of (12) can be equivalently converted into an iteratively reweighted least squares problem:

$$\min \sum_{ij} (\omega(e_{ij}) e_{ij})^2, \tag{14}$$

where

$$\omega(e_{ij}) = \frac{\rho'(e_{ij})}{e_{ij}} \tag{15}$$

represents the weight and $\rho'(e_{ij})$ is the derivative of the Huber function $\rho(e_{ij})$. By this means, the coefficients can be solved by an iterative procedure (Fox and Weisberg 2011).

Consideration of sensor observation differences when calculating spectral differences is a primary contribution of this article. Here, the observation differences are described using the gain and bias parameters, a_c and b_c , as shown in (10). It should be noted that although the method described in Zhu et al. (2010) also considered gain and bias parameters between coarse and fine sensors, the meaning is very different. We used gain and bias parameters to improve the solution precision of the weights of neighbouring pixels. However, Zhu et al. considered gain and bias parameters when setting up the relational models (Zhu et al. 2010), and used (16) and (17) to replace (1) and (2):

$$F(x_i, y_j, t_0) = a_0 C(x_i, y_j, t_0) + \varepsilon_0, \quad (16)$$

$$F(x_i, y_j, t_1) = a_1 C(x_i, y_j, t_1) + \varepsilon_1. \quad (17)$$

Suppose $\varepsilon_0 = \varepsilon_1$ and $a_0 = a_1$, the following equation can be obtained:

$$F(x_i, y_j, t_1) = F(x_i, y_j, t_0) + a_1(C(x_i, y_j, t_1) - C(x_i, y_j, t_0)), \quad (18)$$

where $a = a_0 = a_1$. To deal with mixed pixels, the conversion coefficient a can be obtained by linearly regressing the reflectance changes of fine- and coarse-resolution pixels of the same end member (Zhu et al. 2010). Equations (16)–(18) have been validated as being effective for the two-pairs case in which there are two high-quality input images (Zhu et al. 2010). It should be noted that although (16) and (17) are theoretically more reasonable due to their additional consideration of gain parameter at each observation date, an additional condition, $a_0 = a_1$, is added when we deduce (18). In the single-pair case, the gain parameter a is often very large or very small in linear regression because the change can be balanced by the bias parameter ε . Once the parameter ε is removed, as in (18), the solution stability would be considerably affected. Therefore, in this article, we used (1)–(3) rather than (16)–(18).

3.2. Calculation function

Besides the spectral difference S_{ij} , there are two other parameters in (6), i.e. the distance measurement D_{TJ} and the temporal difference T_{ij} . Here, D_{ij} can be easily obtained by (Gao et al. 2006)

$$D_{ij} = 1 + \sqrt{(x_{w/2} - x_i)^2 + (y_{w/2} - y_j)^2} / (w/2), \quad (19)$$

where w is used to normalize distance, ensuring that the distance range for similar pixels in different search windows changes from 1 to $1 + 2^{0.5}$ (Gao et al. 2006). As for T_{ij} , it is defined as (Gao et al. 2006)

$$T_{ij} = |C(x_i, y_j, t_1) - C(x_i, y_i, t_0)|. \quad (20)$$

A smaller value of T_{ij} means less phenological change between time t_1 and t_0 for a pixel, and it would be assigned a higher weight in the STARFM algorithm (Gao et al. 2006). In most single-pair applications of the STARFM method, the temporal difference is retained (Anderson et al. 2011; Liu and Weng 2011; Walker et al. 2011; Watts et al. 2011). In the paper by Hilker et al. (2009b), however, this factor was excluded. There have been no comparisons between these two choices. Here, we analyse the ambiguity of the two cases.

Case 1. The predicted central pixel showed a little or no phenological change. In this situation, the central pixel to be predicted at time t_1 should have spectral features closer to the corresponding pixel at time t_0 , and all similar neighbouring pixels used for prediction should also show a few phenological changes. Therefore, it is reasonable to set higher weights for pixels with small temporal differences T_{ij} .

Case 2. The predicted central pixel showed a considerable phenological change. In this case, similar pixels of values lower than of T_{ij} are unable to indicate whether the predicted central pixel will show certain phenological changes. Setting with greater weight would yield poor results.

In reality, it is difficult to know in advance whether the predicted central pixel will show significant reflectance change. Since our aim was simply to predict reflectance changes, the factor of temporal difference T_{ij} should be removed. Even when there are two or more fine-resolution input images, the aspect of temporal difference should use other more suitable expressions, as described in papers by Hilker et al. (2009a) and Zhu et al. (2010).

Based on the foregoing, the spatial weighting function E_y is related only to spectral difference and geometric distance and can be expressed as

$$E_{ijc} = \ln(S_{ijc}(A + 1))D_{ij}. \quad (21)$$

In order to ensure that sum of all weights is 1, W_{JC} can be normalized as

$$W_{ijc} = \left(1/E_{ijc} / \sum_1^w (1/E_{ijc}) \right). \quad (22)$$

Equations (21) and (22) mean that a difference in regard to reduced distance and spectral difference between the similar pixel and the central pixel should yield a greater weight. The fine-resolution image can now be predicted using

$$F(x_{w/2}, y_{w/2}, t_1) = \sum_{i=1}^w \sum_{j=1}^w W_{ij} (C(x_i, y_j, t_1) + F(x_i, y_j, t_0) - C(x_i, y_j, t_0)). \quad (23)$$

Note that only spectrally similar pixels in the neighbourhood are selected to calculate reflectance. The neighbouring pixels may be selected using

$$\left| F(x_i, y_j, B) - F(x_{w/2}, y_{w/2}, B) \right| \leq d \times 2^{F(x_{w/2}, y_{w/2}, B)}, \quad (24)$$

where $F(x_{w/2}, y_{w/2}, B)$ is central pixel reflectance, $F(x_i, y_j, B)$ is neighbour pixel reflectance, with B being the band number, and d is a free parameter. For different sensors, the parameter d may be slightly different. However, this may be distributed within a very limited range for

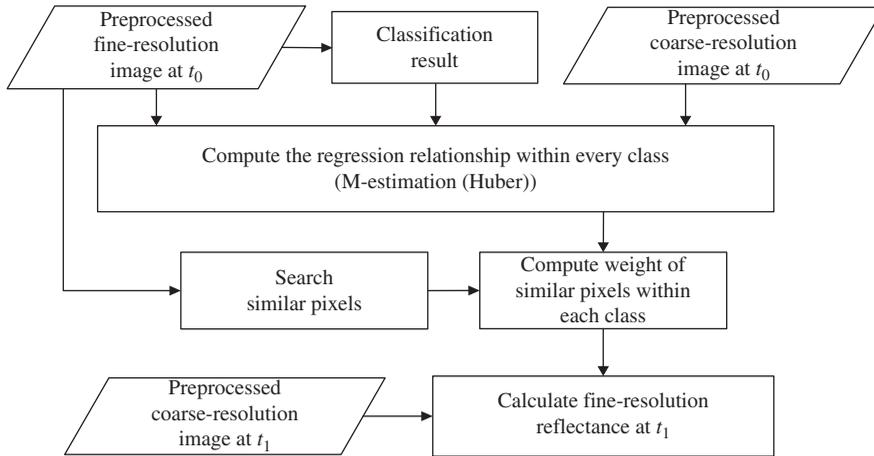


Figure 2. Flowchart of the proposed method.

a given sensor, and in this article, it was set to 0.01 for the ETM+ case. Figure 2 presents a flowchart of the proposed method.

4. Results and analysis

In this section, we tested the performance of the proposed method on several different data sets. The data were obtained from both USGS LPDAAC portal (<https://lpdaac.usgs.gov/>) and NASA GSFC portal (<http://modis.gsfc.nasa.gov/>). Moderate spatial resolution MODIS surface reflectance data (MOD09) and 30 m spatial resolution Landsat ETM+ data were used. Both MOD09 and ETM+ data were preprocessed geometrically and radiometrically. The ETM+ data (L1T product) downloaded provided systematic radiometric and geometric accuracy by incorporating ground control points while employing a digital elevation model (DEM) for topographic accuracy. More details on L1T data currently available can be found online (Gutman and Masek 2012). In order to fit with ETM+ data, MOD09 data were re-projected to Universal Transverse Mercator (UTM) mode using the MODIS re-projection tool and were resampled to 30 m spatial resolution using the cubic interpolation approach.

Two MOD09 data sets were acquired to correspond with each ETM+ scene: daily reflectance (MOD09GA) and 8 day composite reflectance data (MOD09A1). The daily reflectance data set comprised intra-day data collected by the MODIS sensor on the Terra satellite. However, MOD09GA data may be affected by BRDF effects (Gao et al. 2010). The 8 day composite comprises the best possible observation obtained from the Terra satellite during an 8 day period, as selected on the basis of high observation coverage, low view angle, absence of clouds or cloud shadow, and aerosol loading (Walker et al. 2011). To illustrate the BRDF effects associated with MODIS daily observations, two MOD09GA sub-images were chosen in the following experiments. One was chosen only when viewing zenith angles were between 0° and 8° (near-nadir), a range similar to ETM+ data. The other was chosen by selecting a nearby date, but with a different viewing zenith angle (off-nadir) and/or different resolution. In addition, the influence of sampling resolution on fusion results was also analysed.

In the first series of experiments, the proposed method was tested on heterogeneous regions under a long time interval (144 days) in Wuhan, China. The second experiment and the third experiment were tested in Beijing, China and Qinghai-Tibet Plateau, China,

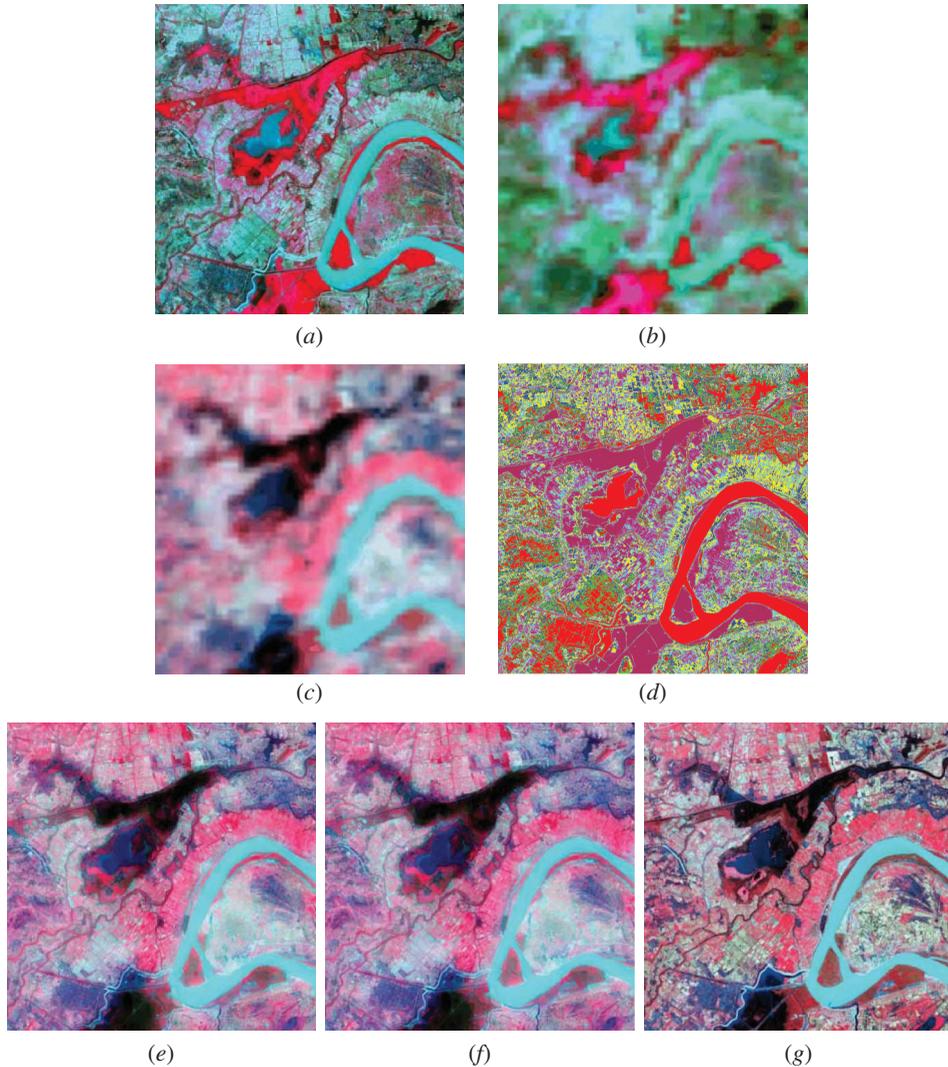


Figure 3. Prediction results from the first experiment (band combination: NIR, red, and green). (a) ETM+ image acquired on 3 May 2001, (b) MOD09GA image (0–5° near-nadir view) acquired on 3 May 2001, (c) MOD09GA image (0–5° near-nadir view) acquired on 24 September 2001, (d) classified image of (a), (e) predicted synthetic ETM+ image by the STARFM method, (f) predicted synthetic ETM+ image by the proposed method, and (g) observed ETM+ image on 24 September 2001.

respectively. Figure 3 shows the first experimental results with a 30×30 km subset. Figure 3(a) shows the ETM+ image acquired on 3 May 2001, and Figures 3(b) and (c) show the MOD09GA sub-images (near-nadir) acquired on 3 May and 24 September 2001, respectively. Figure 3(d) shows the class map using the IsoData algorithm. The predicted synthetic ETM+ images by STARFM and the proposed algorithm are shown in Figures 3(e) and (f). The observed ETM+ image in Figure 3(g) can be used to evaluate the predicted result. From visual inspection, it can be seen that the predicted images in Figures 3(e) and (f) are very similar to the observed image in Figure 3(g), which indicates that both methods can capture reflectance changes to some extent. However, it can be seen that the results of

both methods appear more ‘hazy’ than the observed images. One reason for this is that the large resolution difference between ETM+ and MODIS sensors makes it difficult for fusion algorithms to capture changes in small-grained ground objects with arbitrary distribution. The other reason is that the long time interval (144 days) between the observed and predicted images further aggravates the blurring. When spatial distribution is more homogeneous and temporal interval is shorter, the problem can be relieved to some extent (see the second experiment).

Figure 4 shows the scatter plots between the observed and predicted ETM+ reflectance for the NIR, red, and green bands. Figures 4(a), (c), and (e) show the correlation between the actual reflectance images for 24 September and 3 May for the NIR, red, and green bands, respectively. All scatter plots deviate from the 1:1 line, which indicates there are distinct changes between the two temporal images. Figures 4(b), (d), and (f) show the correlation between the observed 24 September image and the predicted reflectance images of the proposed method. It can be seen that these data all fall closer to the 1:1 line, indicating that the proposed method captured the changes between the May and September images. After the fusion process, the determination coefficients R^2 greatly increased from 0.05, 0.36, and 0.38 to 0.66, 0.65, and 0.63 for the NIR, red, and green bands, respectively.

Similar experiments were also conducted using the corresponding MOD09A1 data sets and MOD09GA data at a nearby date, but with different view angle and resolution, respectively. In order to quantitatively compare the proposed method with STARFM using the different MOD09 data sets, the evaluation index of average absolute difference (AAD) was adopted (Gao et al. 2006; Walker et al. 2011; Zhu et al. 2010). AAD represents the average absolute difference between predicted and observed actual ETM+ images. Furthermore, the IR (improvement ratio) index

$$IR = \frac{|AAD_{STARFM} - AAD_{Proposed}|}{AAD_{STARFM}} \times 100\% \quad (25)$$

is defined to evaluate the degree of improvement in the proposed method in comparison with STARFM. Quantitative values are shown in Table 2. For all the data sets, the MOD09GA sub-images observed at a near-nadir view yielded the lowest AAD values. The quantitative results became significantly worse when the MOD09GA sub-images were observed at an off-nadir view, which can be explained by BRDF effects associated with the MOD09GA data sets. Although a coarser resolution ($2 \text{ km} \times 2 \text{ km}$) may reduce the uncertainty of surface reflectance at an off-nadir view ($55\text{--}60^\circ$) (Gao et al. 2010), its influence on fusion results is very small. In addition, AAD values from the MOD09GA sub-images at a near-nadir view were slightly better than those from the MOD09A1 sub-images for all three bands. The reason for this may be that a small number of pixels of the 8 day composite MOD09A1 were composited during periods of phenological change. However, when the region of interest is subject to BRDF effects, 8 day composite MOD09A1 data sets may be more applicable than MOD09GA.

On the other hand, in all cases, ADD values from the proposed method were lower than those from the original STARFM, which proves that the proposed method is applicable to both MOD09GA and MOD09A1. In particular, improvement ratios from the MOD09A1 data sets were 8.1%, 4.5%, and 5.9% for the NIR, red, and green bands, respectively. In general, because of the factor of sensor observation differences, the proposed method produced better quantitative evaluation results than the STARFM algorithm.

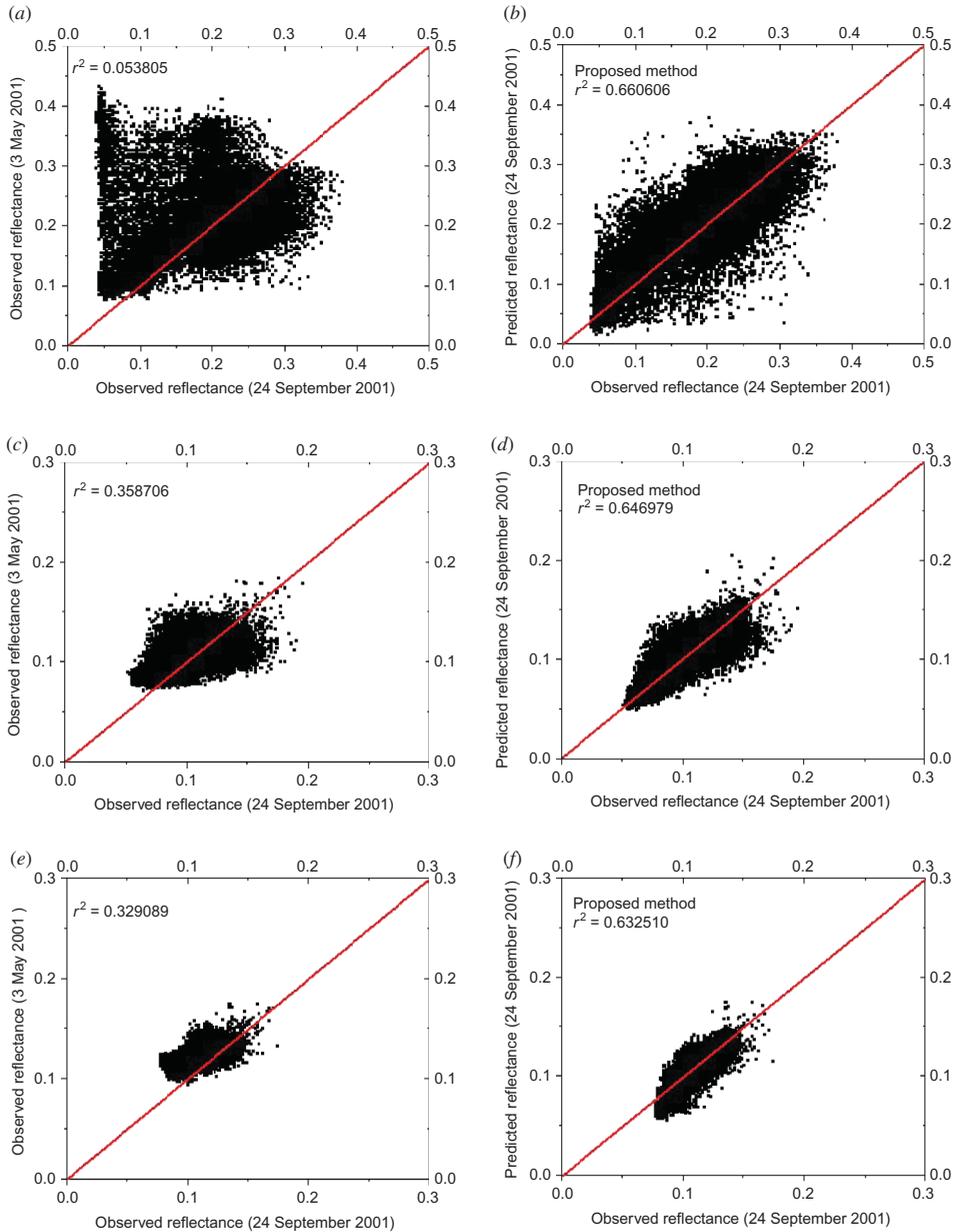


Figure 4. Scatter plots of the first experiment. Left column: comparison between ETM+ reflectance observed on 24 September and ETM+ reflectance observed on 3 May, for NIR (a), red (c), and green (e) bands. Right column: comparison between ETM+ reflectance observed at 24 September and predicted ETM+ reflectance, for NIR (b), red (d), and green (f) bands. The plots shown in the right column show a better fit with the 1:1 line than those in the left.

The sensitivity of the proposed method in regard to class number and different classification algorithms was evaluated, and the results are shown in Table 3. Using the unsupervised IsoData algorithm, fusion accuracy is poor when the class number is less

Table 2. Quantitative evaluation results (AAD) from the first experiment.

		STARFM	Proposed method	IR (%)
MOD09GA (0–5°)	NIR	0.033182	0.031218	5.9
	Red	0.010914	0.010674	2.0
	Green	0.006925	0.006628	4.2
MOD09GA (55–60°) (0.5 km × 0.5 km)	NIR	0.052045	0.050202	3.5
	Red	0.013250	0.013141	0.82
	Green	0.009352	0.009100	2.7
MOD09GA (55–60°) (2 km × 2 km)	NIR	0.058393	0.056115	3.9
	Red	0.014994	0.014835	1.1
	Green	0.009942	0.009687	2.6
MOD09A1 (0–5°)	NIR	0.035140	0.032278	8.1
	Red	0.011431	0.010916	4.5
	Green	0.007187	0.006760	5.9

Table 3. Evaluation of the sensitivity of the proposed algorithm to class number and classification algorithms.

Algorithm		IsoData				ML	SA
Class number		3	5	7	10	7	7
AAD	NIR	0.032114	0.031242	0.031218	0.031238	0.031192	0.031189
	Red	0.010805	0.010683	0.010674	0.010698	0.010671	0.010663
	Green	0.006813	0.006648	0.006628	0.006637	0.006650	0.006635

than 5. When the class number is equal to or greater than 5, however, AAD values change very slightly. Even when some ground classes were divided into several sub-classes, fusion accuracy was not markedly affected. This implies that we can set a relatively large class number when implementing the proposed method.

The supervised classification algorithms of maximum likelihood (ML) and spectral angle (SA) were also implemented for comparison with the IsoData algorithm for the fusion problem. The results show that the proposed method is not sensitive to the choice of classification algorithm. In the experiments, we found that although slight classification errors may exist, fusion accuracy can still be ensured because of usage of the M-estimation method, which is very robust with regard to outliers.

To further test the performance of the proposed method, it was implemented on a relatively homogeneous area (90 km × 90 km). The temporal interval (32 days) was also shorter. Figures 5(a)–(c) show the ETM+ image acquired on 13 December 2001, the MOD09GA sub-image (near-nadir) acquired on 13 December 2001, and the MOD09GA sub-image (near-nadir) of 11 November 2001, respectively. Figure 5(d) shows the classified image of Figure 5(a). The synthetic ETM+ images predicted by STARFM and the proposed method are shown in Figures 5(e) and (f), and the actual ETM+ image acquired on 11 November 2001 is shown in Figure 5(g). To enhance visual judgement, some detailed regions cropped from Figures 5(e)–(g) are shown in Figures 6(a)–(c). It is clear that the image predicted by the proposed method is similar to the actual image, while the image predicted by STARFM seems incoherent and has some outliers. The evaluation results

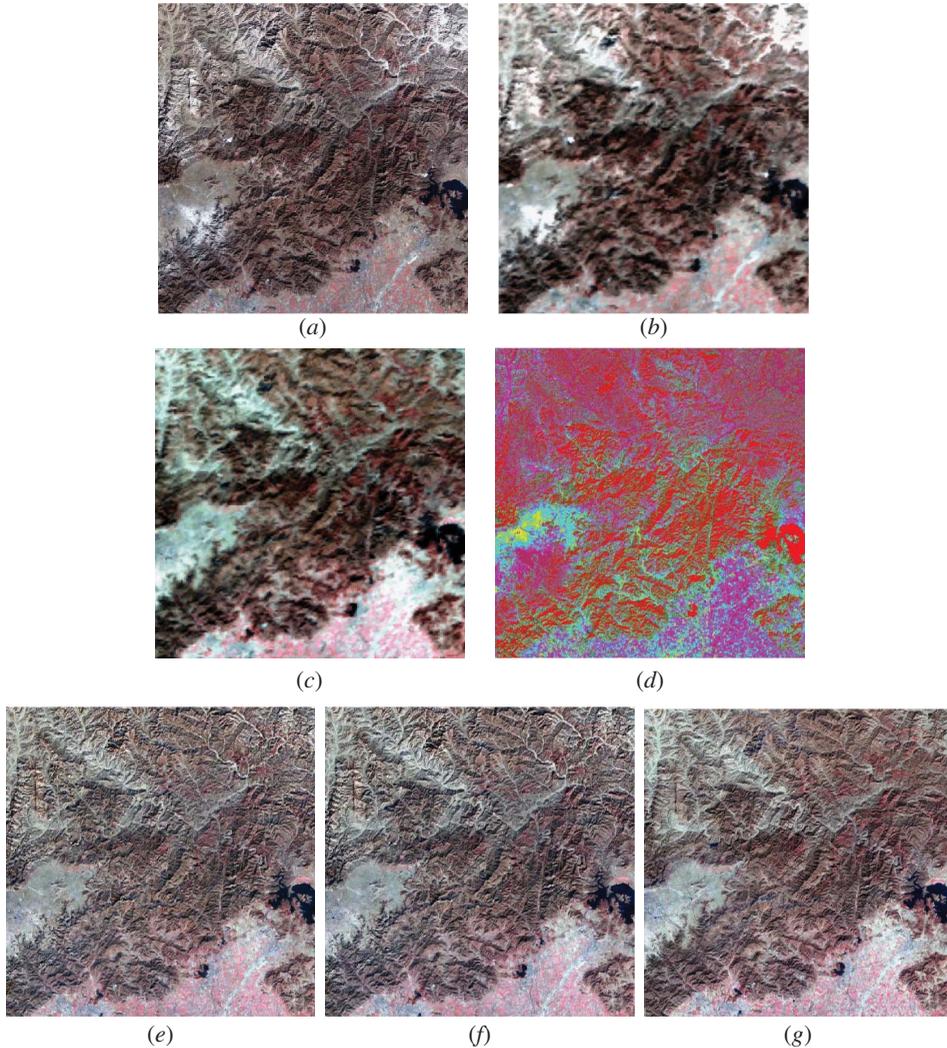


Figure 5. Prediction results of the second experiment (band combination: NIR, red, and green). (a) ETM+ image acquired on 13 December 2001, (b) MOD09GA image (0–8° near-nadir view) acquired on 13 December 2001, (c) MOD09GA image (0–8° near-nadir view) acquired on 11 November 2001, (d) classified image of (a), (e) predicted image by STARFM, (f) predicted image by the proposed method, and (g) observed ETM+ image on 11 November 2001.

shown in Table 4 agree with the visual judgement in that the proposed method yields higher evaluation scores. For the different MOD09 data sets, the same conclusions as the first experiment can be drawn from the results shown in Table 4. In addition, since spatial distribution was relatively homogeneous and the time interval was quite short in this experiment, the improvement ratios are all much higher than those from previous experiments. To be specific, the improvement ratios from the MOD09A1 data sets reached 8.5%, 9.8%, and 10.6% for the NIR, red, and green bands, respectively.

The final experiment was performed to confirm that it is reasonable to remove the temporal difference T_{ij} from the weight function (6). The original STARFM and T_{ij} -removed

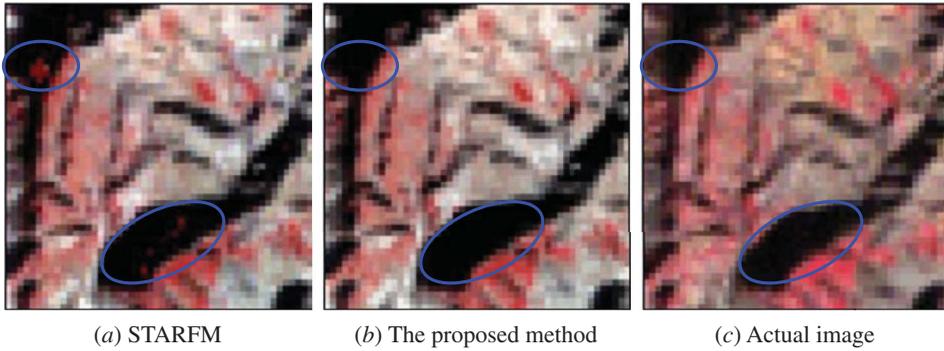


Figure 6. Detailed regions cropped from Figures 5(e)–(g): (a) and (b) are the prediction images by STARFM and the proposed method, respectively, and (c) is the actual image observed on 11 November 2001.

Table 4. Quantitative evaluation results (AAD) from the second experiment.

		STARFM	Proposed method	IR (%)
MOD09GA (0–8°)	NIR	0.020321	0.018834	7.2
	Red	0.012968	0.011862	8.4
	Green	0.009886	0.008945	9.3
MOD09GA (26–34°) (0.5 km × 0.5 km)	NIR	0.023358	0.021995	5.8
	Red	0.017884	0.017232	3.6
	Green	0.013761	0.013182	4.2
MOD09GA (26–34°) (1 km × 1 km)	NIR	0.024028	0.023042	4.1
	Red	0.018229	0.017771	2.5
	Green	0.013937	0.013512	3.0
MOD09A1 (0–8°)	NIR	0.021322	0.019501	8.5
	Red	0.013696	0.012354	9.8
	Green	0.010461	0.009353	10.6

STARFM were, respectively, implemented, and the input and resulting images are shown in Figure 7. Although there are no distinct visual differences between the two resulting images, the quantitative evaluation results of Figure 7(e) (without T_{ij}) are better than those of Figure 7(d) (with T_{ij}) for all three bands, as shown in Table 5. Furthermore, similar experiments were implemented using several different data sets. In most cases, the removal of temporal difference yielded improved results, the exception being when multitemporal images contained few changes.

5. Conclusion

This article presents a method of fusing ETM+ and MODIS reflectance data in order to enhance the temporal resolution of fine spatial resolution data. Compared with the original STARFM, the proposed method pays attention to the issue of sensor observation differences and the removability of temporal differences when calculating pixel weights. The proposed method was tested under varying landscape situations, MOD09 scenes, and time intervals. The experimental results show that the issues addressed by the proposed method

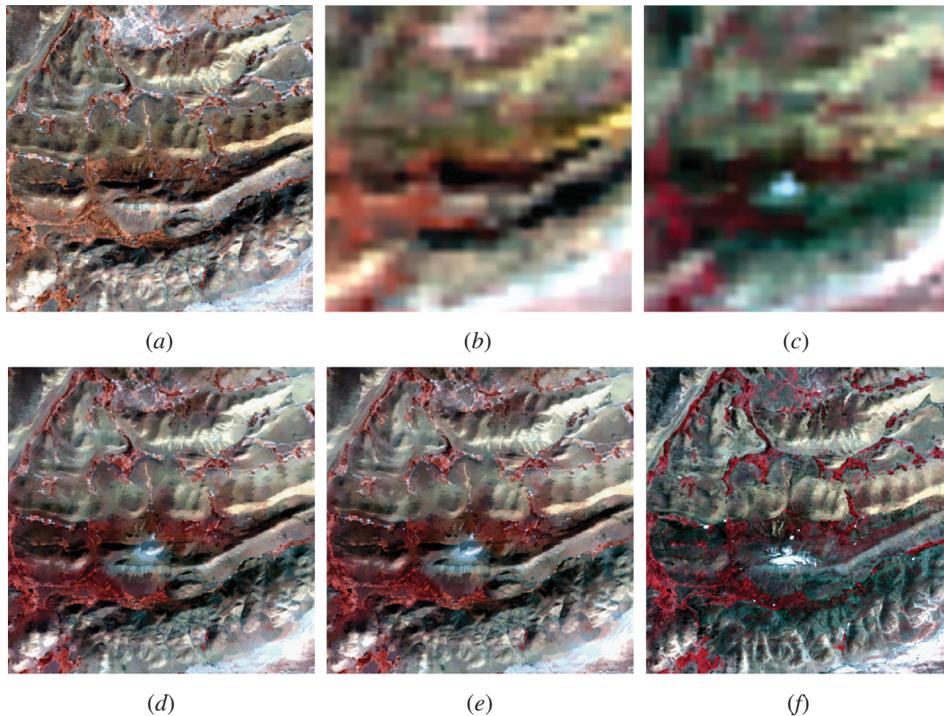


Figure 7. Prediction results of the third experiment (band combination: NIR, red, and green). (a) ETM+ image acquired on 4 November 2001, (b) MOD09GA image (0–3° near-nadir view) acquired on 4 November 2001, (c) MOD09GA image (0–3° near-nadir view) acquired on 13 June 2001, (d) predicted synthetic image by STARFM, (e) predicted synthetic image by the revised STARFM, which removes the factor of temporal difference, and (f) observed ETM+ image on 13 June 2001.

Table 5. Validation of removal of temporal difference in the STARFM method.

		STARFM	STARFM (without T_{ij})	IR (%)
MOD09GA (0–3°)	NIR	0.015000	0.014846	1.1
	Red	0.013001	0.012782	1.7
	Green	0.008622	0.008518	1.2

can improve the accuracy of synthetic images. As with other methods, nevertheless, the predicted image appears smoothed when the temporal interval is long and/or the spatial distribution is inhomogeneous. These issues will be further addressed in our future work.

Acknowledgement

The authors gratefully acknowledge the research support from the Major State Basic Research Development Programme (973 Programme) of China under Grant No. 2011CB707103, the National High Technology Research and Development Programme (863 Programme) under Grant 2013AA12A301, National Natural Science Foundation of China under Grant 41271376, and Fundamental Research Funds for the Central Universities under Grant 2012205020205. We are grateful to the reviewers for their helpful comments and suggestions.

References

- Anderson, M., M. P. Kustas, J. M. Norman, C. R. Hain, J. R. Mecikalski, L. Schultz, M. P. Gonzalez-Dugo, C. Cammalleri, G. d'Urso, A. Pimstein, and F. Gao. 2011. "Mapping Daily Evapotranspiration at Field to Continental Scales Using Geostationary and Polar Orbiting Satellite Imagery." *Hydrology and Earth System Sciences* 15: 223–39.
- Ehlers, M. 1991. "Multisensor Image Fusion Techniques in Remote Sensing." *ISPRS Journal of Photogrammetry and Remote Sensing* 46: 19–30.
- Eismann, M., and R. Hardie. 2005. "Hyperspectral Resolution Enhancement Using High-Resolution Multispectral Imagery with Arbitrary Response Functions." *IEEE Transactions on Geoscience and Remote Sensing* 43: 455–65.
- Fox, J., and S. Weisberg. 2011. *An R Companion to Applied Regression*. Thousand Oaks, CA: SAGE Publications.
- Gao, F., J. Masek, M. Schwaller, and F. Hall. 2006. "On the Blending of the Landsat and MODIS Surface Reflectance: Predicting Daily Landsat Surface Reflectance." *IEEE Transactions on Geoscience and Remote Sensing* 44: 2207–18.
- Gao, F., J. G. Masek, R. E. Wolfe, and C. Huang. 2010. "Building a Consistent Medium Resolution Satellite Data Set Using Moderate Resolution Imaging Spectroradiometer Products as Reference." *Journal of Applied Remote Sensing* 4: 043526.
- Gutman, G., and J. G. Masek. 2012. "Long-Term Time Series of the Earth's Land-Surface Observations From Space." *International Journal of Remote Sensing* 33: 4700–19.
- Hall, D. L., and J. Llinas. 1997. "An Introduction to Multisensor Data Fusion." *Proceedings of the IEEE* 85: 6–23.
- Hansen, M., D. Roy, E. Lindquist, B. Adusei, C. Justice, and A. Altstatt. 2008. "A Method for Integrating Modis and Landsat Data for Systematic Monitoring of Forest Cover and Change in the Congo Basin." *Remote Sensing of Environment* 112: 2495–513.
- Hilker, T., M. Wulder, N. Coops, J. Linke, G. Mcdermid, J. Masek, F. Gao, and J. White. 2009a. "A New Data Fusion Model for High Spatial-and Temporal-Resolution Mapping of Forest Disturbance Based on Landsat and MODIS." *Remote Sensing of Environment* 113: 1613–27.
- Hilker, T., M. Wulder, N. Coops, N. Seitz, J. White, F. Gao, J. Masek, and G. Stenhouse. 2009b. "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: 1988–99.
- Huber, P. J. 1964. "Robust Estimation of a Location Parameter." *The Annals of Mathematical Statistics* 35: 73–101.
- Liu, H., and Q. Weng. 2011. "Enhancing Temporal Resolution of Satellite Imagery for Public Health Studies: A Case Study of West Nile Virus Outbreak in Los Angeles in 2007." *Remote Sensing of Environment* 117: 57–71.
- Luo, R. C., Y. Chih-Chen, and S. Kuo Lan. 2002. "Multisensor Fusion and Integration: Approaches, Applications, and Future Research Directions." *IEEE Sensors Journal* 2: 107–19.
- Masek, J., E. Vermote, N. Saleous, R. Wolfe, F. Hall, K. Huemmrich, F. Gao, J. Kutler, and T. Lim. 2006. "A Landsat Surface Reflectance Dataset for North America, 1990–2000." *IEEE Geoscience and Remote Sensing Letters* 3: 68–72.
- Pohl, C., and J. Van Genderen. 1998. "Review Article Multisensor Image Fusion in Remote Sensing: Concepts, Methods and Applications." *International Journal of Remote Sensing* 19: 823–54.
- Roy, D., J. Ju, P. Lewis, C. Schaaf, F. Gao, M. Hansen, and E. Lindquist. 2008. "Multi-temporal Modis-landsat Data Fusion for Relative Radiometric Normalization, Gap Filling, and Prediction of Landsat Data." *Remote Sensing of Environment* 112: 3112–30.
- Shen, H., M. Ng, P. Li, and L. Zhang. 2009. "Super-Resolution Reconstruction Algorithm to MODIS Remote Sensing Images." *The Computer Journal* 52: 90–100.
- Singh, D. 2010. "Generation and Evaluation of Gross Primary Productivity Using Landsat Data through Blending with MODIS Data." *International Journal of Applied Earth Observation and Geoinformation* 13: 59–69.
- Thomas, C., T. Ranchin, L. Wald, and J. Chanussot. 2008. "Synthesis of Multispectral Images to High Spatial Resolution: A Critical Review of Fusion Methods Based on Remote Sensing Physics." *IEEE Transactions on Geoscience and Remote Sensing* 46: 1301–12.
- Walker, J., K. de Beurs, R. Wynne, and F. Gao. 2011. "Evaluation of Landsat and MODIS Data Fusion Products for Analysis of Dryland Forest Phenology." *Remote Sensing of Environment* 117: 381–93.

- Wang, Z., D. Ziou, C. Armenakis, D. Li, and Q. Li. 2005. "A Comparative Analysis of Image Fusion Methods." *IEEE Transactions on Geoscience and Remote Sensing* 43: 1391–402.
- Watts, J. D., S. L. Powell, R. L. Lawrence, and T. Hilker. 2011. "Improved Classification of Conservation Tillage Adoption Using High Temporal and Synthetic Satellite Imagery." *Remote Sensing of Environment* 115: 66–75.
- Wu, Y., and K. W. Tam. 2001. "M-Estimation in Exponential Signal Models." *IEEE Transactions on Signal Processing* 49: 373–80.
- Zhu, X., J. Chen, F. Gao, X. Chen, and J. G. Masek. 2010. "An Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model for Complex Heterogeneous Regions." *Remote Sensing of Environment* 114: 2610–23.
- Zurita-Milla, R., G. Kaiser, J. Clevers, W. Schneider, and M. Schaepman. 2009. "Downscaling Time Series of MERIS Full Resolution Data to Monitor Vegetation Seasonal Dynamics." *Remote Sensing of Environment* 113: 1874–85.