

A Spatial and Temporal Nonlocal Filter-Based Data Fusion Method

Qing Cheng, Huiqing Liu, Huanfeng Shen, *Senior Member, IEEE*, Penghai Wu, *Member, IEEE*,
and Liangpei Zhang, *Senior Member, IEEE*

Abstract—The tradeoff in remote sensing instruments that balances the spatial resolution and temporal frequency limits our capacity to monitor spatial and temporal dynamics effectively. The spatiotemporal data fusion technique is considered as a cost-effective way to obtain remote sensing data with both high spatial resolution and high temporal frequency, by blending observations from multiple sensors with different advantages or characteristics. In this paper, we develop the spatial and temporal nonlocal filter-based fusion model (STNLFFM) to enhance the prediction capacity and accuracy, especially for complex changed landscapes. The STNLFFM method provides a new transformation relationship between the fine-resolution reflectance images acquired from the same sensor at different dates with the help of coarse-resolution reflectance data, and makes full use of the high degree of spatiotemporal redundancy in the remote sensing image sequence to produce the final prediction. The proposed method was tested over both the Coleambally Irrigation Area study site and the Lower Gwydir Catchment study site. The results show that the proposed method can provide a more accurate and robust prediction, especially for heterogeneous landscapes and temporally dynamic areas.

Index Terms—Data fusion, nonlocal, reflectance prediction, similarity information, spatiotemporal.

I. INTRODUCTION

CAPTURING spatial and temporal dynamics is a significant issue for many remote sensing-based monitoring systems (e.g., the monitoring of land-cover change, intraseasonal ecosystem variations, and atmospheric environment dynamics). However, due to the technical limitations, remote sensor designs have a tradeoff between the spatial resolution and the revisit cycle [1], [2], which limits our capacity to

Manuscript received October 23, 2016; revised February 23, 2017; accepted March 28, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 41422108 and Grant 41601357 and in part by the Natural Science Foundation of Hubei Province under Grant 2016CFB333. (Corresponding author: Huanfeng Shen.)

Q. Cheng is with the School of Urban Design, Wuhan University, Wuhan 430072, China (e-mail: qingcheng@whu.edu.cn).

H. Liu is with the Surveying and Mapping Institute Lands and Resource Department of Guangdong Province, Guangzhou 510500, China (e-mail: liuhq0903@qq.com).

H. Shen is with the School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China, with the Key Laboratory of Geographic Information System, Ministry of Education, Wuhan University, Wuhan 430079, China, and also with the Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China (e-mail: shenhf@whu.edu.cn).

P. Wu is with the School of Resources and Environmental Engineering, Anhui University, Hefei 230039, China (e-mail: wuph@ahu.edu.cn).

L. Zhang is with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: zlp62@whu.edu.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2017.2692802

acquire remote sensing data with both high spatial resolution and high temporal resolution. For example, the data acquired from the Landsat Thematic Mapper (TM) or Enhanced TM Plus (ETM+) sensors and the SPOT High Resolution Visible (HRV) sensor with a 10–30 m spatial resolution are commonly applied for land-use mapping and biophysical parameter estimation [3]–[6]. However, such data cannot be used to capture rapid surface changes, such as crop growth and natural disasters due to their long revisit cycles (Landsat TM/ETM+: 16 days; SPOT HRV: 26 days) and frequent cloud contamination. In contrast, the Terra/Aqua Moderate Resolution Imaging Spectroradiometer (MODIS) and National Oceanic and Atmospheric Administration Advanced Very High Resolution Radiometer sensors can provide high temporal resolution (daily) observations, and are often applied for monitoring at global scales [7], [8]. However, data from these sensors cannot be used for research at heterogeneous local scales because of their coarse spatial resolution (250–1000 m). Therefore, combining the advantages of the different sensors by spatiotemporal data fusion methods is considered as a cost-effective way to solve the “spatial–temporal contradiction” problem [9]–[11], thereby enhancing the capability of remote sensing for monitoring land-surface dynamics, especially in rapidly changing areas.

Gao *et al.* [12] developed the spatial and temporal adaptive reflectance fusion model (STARFM) algorithm, which is a spatiotemporal filter-based algorithm that generates the fusion data by the use of a filtering model with the weighted sum of the spectrally similar neighboring information from the high spatial resolution images and the high temporal frequency images. The STARFM can be used to blend Landsat and MODIS data to produce daily surface reflectance at Landsat spatial resolution and MODIS temporal frequency. The STARFM algorithm has been shown to be a relatively reliable spatiotemporal data fusion approach, and has been widely applied to the investigation of vegetation dynamics [13], the generation of gross primary productivity [14], the analysis of dryland forest phenology [15], the estimation of daily evapotranspiration [16], the examination of virus dissemination [17], and the monitoring of urban heat islands [18].

Furthermore, some improved spatiotemporal filter-based algorithms have since been developed. Hilker *et al.* [19] proposed the spatial temporal adaptive algorithm for mapping reflectance change to identify highly detailed spatial and temporal patterns in land-cover changes. Zhu *et al.* [20] developed an enhanced STARFM (ESTARFM) model to enhance the prediction of the reflectance of heterogeneous landscapes by assigning different conversion coefficients for homogeneous

and heterogeneous pixels. Fu *et al.* [21] modified the procedure of similar pixel selection for the ESTARFM model with an optimal window size and additional ancillary data. Shen *et al.* [22] improved the step of weight calculation for the original STARFM model by considering the sensor observation differences for varied land-cover types. However, this approach requires a prior unsupervised classification for the fine spatial resolution data. Wu *et al.* [23] developed a spatiotemporal integrated temperature fusion model to expand the traditional two-sensor fusion approach into the fusion of data from an arbitrary number of sensors with a unified model.

In other frameworks, Hansen *et al.* [24] used regression trees to integrate Landsat and MODIS data on a 16-day repeat cycle to monitor forest cover change in the Congo Basin. This method demands a single “best” image to map forest cover status for a given year or decade. Zurita-Milla *et al.* [25] developed an unmixing-based fusion framework to produce Landsat-like images having the spectral and temporal resolution provided by the Medium Resolution Imaging Spectrometer. However, this unmixing-based fusion approach requires a prior unsupervised classification for the input fine spatial resolution images, or a high spatial resolution land-use database as auxiliary material for the pixel unmixing. Learning-based spatiotemporal fusion frameworks have been developed in recent years [26], [27], which are generally based on sparse representation and compressive sensing. This approach can predict both the phenology change and the land-cover type change during an observation period, in a unified way. Nonetheless, the practicability of the learning-based fusion methods needs to be further verified. Moreover, image super-resolution [28]–[30] can be considered as a different kind of technique to improve the spatial resolution for low spatial resolution but high temporal frequency images.

Generally speaking, the spatiotemporal filter-based fusion framework has been the most popular category of spatiotemporal fusion approach so far. Although it has been improved in many different ways, the spatiotemporal filter-based fusion framework still has some shortcomings that need to be improved, including the complex change prediction ability and the robustness of the prediction model. In this paper, a spatial and temporal nonlocal filter-based fusion model (STNLFFM) is presented. The proposed STNLFFM method provides a new transformation relationship between the multitemporal fine-resolution reflectance images with the help of multitemporal coarse-resolution reflectance data, and then makes full use of the high degree of spatiotemporal redundancy in the remote sensing image sequence to produce the final prediction. Moreover, some effective treatments are also proposed in the calculation procedure of the STNLFFM method.

II. METHOD

A. Theoretical Basis of the STNLFFM Method

For convenience, we refer to the image with low spatial resolution but high temporal frequency as the “coarse-resolution” image, while the image with high spatial resolution but low temporal frequency is called the “fine-resolution” image. For a homogeneous coarse-resolution pixel, neglecting

the differences in atmospheric correction, we suppose that the changes of reflectance from date t_k to t_0 are linear. This assumption is reasonable over a short time period [20]. Thus, the reflectance of the homogeneous coarse-resolution pixel at t_0 can be described by the reflectance at t_k as

$$C(x, y, B, t_0) = a(x, y, B, \Delta t) \times C(x, y, B, t_k) + b(x, y, B, \Delta t) \quad (1)$$

where C denotes the coarse-resolution reflectance, (x, y) is a given pixel location for both coarse-resolution images at two different dates, B is a given band, $\Delta t = t_0 - t_k$, and a and b are the coefficients of the linear regression model to describe the reflectance changes in the coarse-resolution images at dates t_0 and t_k . In this paper, we refer to coefficient a as the gain coefficient, and coefficient b as the bias coefficient.

We assume that the coarse-resolution sensor has similar spectral bands to the fine-resolution sensor. We also suppose that the coarse-resolution image has been upsampled to the same spatial resolution, size, and coverage as the fine-resolution image. Neglecting the geolocation errors and differences in atmospheric correction, the linear relationship between the coarse-resolution reflectance images, as in (1), can also apply to the fine-resolution reflectance images at dates t_0 and t_k . The fine-resolution reflectance at date t_0 is then calculated as

$$F(x, y, B, t_0) = a(x, y, B, \Delta t) \times F(x, y, B, t_k) + b(x, y, B, \Delta t) \quad (2)$$

where F denotes the fine-resolution reflectance. Here, the coefficients a and b are determined by the reflectance changes from date t_k to date t_0 . It is notable that a and b may vary with location due to the complexity of the land cover, and thus they are derived locally rather than using global coefficients.

In reality, the land cover may undergo significant and complex changes during the prediction period, and the coarse-resolution observation may not be a homogeneous pixel but includes mixed land-cover types. In order to make the prediction more accurate, some additional information needs to be introduced. We take advantage of the same-class pixels (similar pixels) within the image by considering the same-class pixels with similar reflectance changes over time. As we all know, remote sensing images are often used for generating a wide range of surface information, so there will always be a lot of similar information within an image. This similar information not only includes the local neighborhood similarity, but also includes the nonlocal similarity [31]–[35], such as some repeated ground information, or long edge structures. Moreover, the temporal correlation between remote sensing image sequences makes the similar information (redundancy) even greater, as shown in Fig. 1. These observations prompted us to introduce the idea of nonlocal filtering [33], [34], which attempts to make full use of the high degree of redundant information in the image restoration.

B. Nonlocal Filter

The nonlocal filter is an effective image denoising algorithm [33]–[35]. Its basic idea is to estimate an unknown

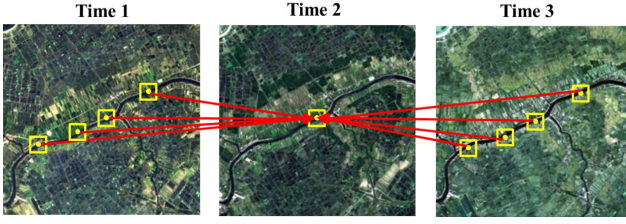


Fig. 1. Spatial and temporal nonlocal similarity in remote sensing imagery.

pixel with image redundancy. More precisely, given a noisy image f , Ω is its pixel domain. The restored value \hat{f} for pixel (x, y) in image f is

$$\hat{f}(x, y) = \sum_{(x_i, y_j) \in \Omega} w_f(x_i, y_j) \times f(x_i, y_j) \quad (3)$$

where

$$w_f(x_i, y_j) = \frac{1}{C(x, y)} \exp \left\{ -\frac{G * \|f(P_{(x_i, y_j)}) - f(P_{(x, y)})\|^2}{h^2} \right\} \quad (4)$$

where $C(x, y)$ is a normalizing factor, G is a Gaussian kernel, h is a parameter, which is positively related to the noise intensity of the image f , and $P_{(x_i, y_j)}$ is a patch centered at point (x_i, y_j) . Following (3), the current pixel is restored by averaging the other similar pixels in the image.

C. Prediction Model of STNLFFM

As noted earlier, the similar information in the image can be used to enhance the reflectance prediction, so we employ the idea of the nonlocal filter. Moreover, in order to take advantage of the redundant information that exists in both the spatial and temporal directions, we propose the STNLFFM (5), which integrates the image's spatial and temporal redundancy into the fine-resolution reflectance calculation

$$F(x, y, B, t_p) = \sum_{k=1}^M \sum_{i=1}^N W(x_i, y_i, B, t_k) \times [a(x_i, y_i, B, \Delta t_k) \times F(x_i, y_i, B, t_k) + b(x_i, y_i, B, \Delta t_k)] \quad (5)$$

where $F(x, y, B, t_p)$ is the fine-resolution reflectance of the target (predicted) pixel (x, y) at prediction date t_p ; M is the number of base (reference) dates; and N is the number of similar pixels (with the same land-cover type as the target pixel) within the image, including the target pixel itself. (x_i, y_i) is the location of the i th similar pixel, and $W(x_i, y_i, B, t_k)$ is the weight of the i th similar pixel of the fine-resolution reflectance image at base date t_k . A way to more effectively exploit the image redundancy is by searching for similar pixels [35], so we select similar pixels to estimate the target pixel (5), instead of taking all the image pixels to estimate the target pixel as in the original nonlocal filtering model (3).

1) *Similar Pixel Selection*: As mentioned before, the pixels with the same land-cover type as the target pixel are the similar pixels. Selecting the similar pixels ensures that the appropriate reflectance information is used for the production of the target fine-resolution pixel, which can avoid trivial calculation and improve the prediction accuracy. Since the

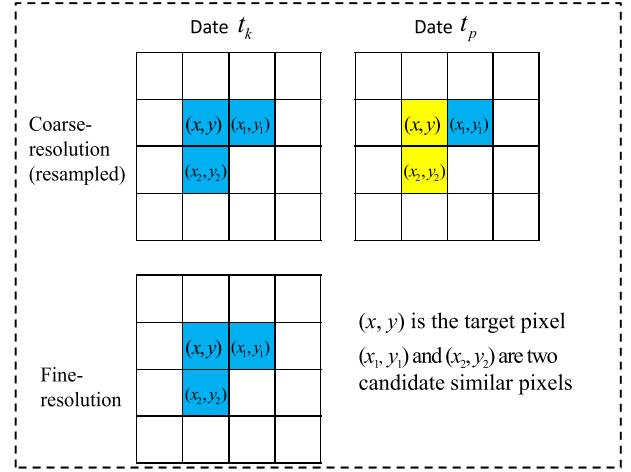


Fig. 2. Similar pixel selection.

similar pixels have close reflectances and changes, we use two constraints (spectral consistency and change consistency) to pick them out. A good candidate pixel (x_i, y_j) should satisfy the following conditions for the reflectance of all the bands:

$$\begin{aligned} |F(x_i, y_j, B, t_k) - F(x, y, B, t_k)| &\leq d \times 2^{F(x, y, B, t_k)} \quad (6) \\ ||C(x_i, y_j, B, t_k) - C(x_i, y_j, B, t_p)| \\ - |C(x, y, B, t_k) - C(x, y, B, t_p)|| &< \sigma_{CC} \quad (7) \end{aligned}$$

where d is a free parameter [22]. Parameter d may be slightly different for different sensors, and in this paper, it is set as 0.01 for the ETM+ case. σ_{CC} is the uncertainty of the temporal difference between the two coarse-resolution reflectance images, which is mainly caused by the bias in the atmospheric correction. In our experiments, the uncertainty of the MODIS surface reflectance was set to 0.005 for the visible band and the near infrared band, as in the research of Gao *et al.* [12]. Equation (6) is the spectral consistency condition to ensure that the reflectance between a similar pixel and the target pixel is close. It is noteworthy that (7) is the temporal change consistency condition, which means that a pixel whose temporal change between date t_k and prediction date t_p is close to that of the target pixel is more likely to be selected as a similar pixel. As for the STARFM method [12], it regards a pixel whose temporal change between dates t_k and t_p is smaller than that of the target pixel as a candidate similar pixel. However, this results in significant error when the predicted target pixel shows land-cover type change, because, in this case, a similar pixel would also show considerable reflectance change. As shown in Fig. 2, the target pixel (x, y) shows significant change between reference date t_k and prediction date t_p . We assume that pixels (x_1, y_1) and (x_2, y_2) are two candidate similar pixels that satisfy (6). However, we can see that pixel (x_1, y_1) shows a small temporal change between dates t_k and t_p , but this change is not consistent with target pixel (x, y) , so pixel (x_1, y_1) should not be chosen as a similar pixel. The temporal change of pixel (x_2, y_2) is close to that of the target pixel, and thus pixel (x_2, y_2) is more likely to be a similar pixel. There are also some other papers [13], [22] in which the temporal difference measurement is excluded in the

procedure of selecting similar pixels, for the sake of avoiding the error mentioned above. Here, we use (7) to effectively solve the problem of temporal difference measurement and improve the accuracy of selecting similar pixels, especially for the target pixels with land-cover type changes. In practice, we use a search window of size 51×51 centered at the target pixel, as in [34], since searching for similar pixels over the whole image domain is very expensive.

2) *Weight Calculation*: The weight W decides the contribution of each similar pixel to predict the fine-resolution reflectance of the target pixel. According to the nonlocal filtering framework, the weight W is determined by the reflectance similarity between the similar pixel and the target pixel. Since the fine-resolution reflectance of the target pixel at prediction date t_p is unknown, we use the coarse-resolution reflectance difference between the similar pixel and the target pixel to measure the reflectance similarity, and we then propose a nonlocal filter-based individual weight $W_{\text{individual}}$ as follows:

$$W_{\text{individual}}(x_i, y_j, B, t_k) = \exp\left(-\frac{G * \|C(P(x_i, y_j, B, t_k)) - C(P(x, y, B, t_p))\|}{h^2}\right) \quad (8)$$

where h is a parameter positively related to the noise level of the coarse-resolution images input. Since the coarse-resolution (MODIS) reflectance images used in our experiments contained a low level of noise, parameter h was set as 0.15 in the experiments. G is a Gaussian kernel, and $C(P(x_i, y_j, B, t_k))$ is the coarse-resolution reflectance of patch P centered at pixel (x_i, y_j) . The size of patch P is related to the spatial resolution difference between the coarse- and fine-resolution data input. If the spatial resolution difference is large, it is better to set a small size for patch P , since the ground structures in the upsampled coarse-resolution data may not be that clear and would not benefit the similarity identification. In contrast, if the spatial resolution difference is small, we can set a relatively large size for patch P as the structure information in the upsampled coarse-resolution data would benefit the similarity identification. The nonlocal weight $W_{\text{individual}}$ determines the weight of the individual pixel. Furthermore, considering that fine-resolution data closer in date to the prediction date should provide more reliable reflectance information, it is reasonable, in this case, to set a larger weight for the fine-resolution data input. Thus, we introduce the whole weight W_{whole} , which is calculated according to the change magnitude detected by the coarse-resolution reflectance between reference date t_k and prediction date t_p , and this weight W_{whole} is applied within each local window $w \times w$ (9), to decide which fine-resolution image input provides more reliable information in the local window

$$W_{\text{whole}} = \frac{1 / \sum_{i=1}^w \sum_{j=1}^w (|C(x_i, y_j, B, t_k) - C(x_i, y_j, B, t_p)|)}{\sum_k (1 / \sum_{i=1}^w \sum_{j=1}^w (|C(x_i, y_j, B, t_k) - C(x_i, y_j, B, t_p)|))}. \quad (9)$$

A larger value of W_{whole} means that the fine-resolution reflectance at date t_k should be given a higher weight.

Then, synthesizing the two weights $W_{\text{individual}}$ and W_{whole} , the final weight of the similar pixel (x_i, y_j) in the fine-resolution reflectance image at reference date t_k is

$$W(x_i, y_j, B, t_k) = W_{\text{individual}}(x_i, y_j, B, t_k) \times W_{\text{whole}}. \quad (10)$$

3) *Regression Coefficients Calculation*: The regression coefficients a and b for each similar pixel in the fine-resolution reflectance images can be calculated from the available coarse-resolution reflectance images. Since the similar pixels have the same reflectance change as the target pixel, they should have the same regression coefficients. Thus, it is feasible to make use of the information from similar pixels to compute the regression coefficients. In the case of the coarse-resolution reflectance at dates t_k and t_p being perfectly correlated, i.e., the gain coefficient a is equal to 1, we have

$$C(x, y, B, t_p) = C(x, y, B, t_k) + b(x, y, B, \Delta t). \quad (11)$$

The predicted fine-resolution reflectance is then

$$F(x, y, B, t_p) = F(x, y, B, t_k) + C(x, y, B, t_p) - C(x, y, B, t_k). \quad (12)$$

Therefore, we can see that the STARFM model is a special case of the STNLFFM model. In reality, due to the complexity of the land cover, the gain coefficient a may not be equal to 1, but varies in the vicinity of 1. Therefore, we apply a restricted least-squares model to the coarse-resolution reflectance of the similar pixels to obtain the regression coefficients a and b for the target pixel

$$\begin{aligned} \text{set } A &= \begin{pmatrix} a \\ b \end{pmatrix}, \quad I = (1, 0) \\ \arg \min f(A) &= \frac{1}{2} \left[\begin{pmatrix} C_{p1} \\ C_{p2} \\ \vdots \\ C_{pn} \end{pmatrix} - \begin{pmatrix} C_{k1} & 1 \\ C_{k2} & 1 \\ \vdots & \vdots \\ C_{kn} & 1 \end{pmatrix} \times A \right]^2 \\ &\quad + \frac{1}{2} \gamma [I \times A - 1]^2 \end{aligned} \quad (13)$$

where C_{pn} and C_{kn} are the coarse-resolution reflectance of the n th similar pixel at dates t_p and t_k , respectively. The first item on the right-hand side of (13) is the least-squares term, and the second item is the regularization term, which gives a prior constraint of the regression coefficients. γ is a regularization parameter, which controls the tradeoff between the least-squares term and the regularization term. The value of γ was set as 1 in the experiments.

D. Implementation Process of STNLFFM

In practice, we use at least two pairs of fine- and coarse-resolution images acquired at the reference dates (t_m and t_n) and a coarse-resolution image acquired at the prediction date (t_p) to predict the desired fine-resolution image. The STNLFFM algorithm implementation includes three main steps: 1) for each pixel in the reference fine-resolution images, we search for pixels similar to it in the image; 2) the weights W of all the similar pixels are calculated; and 3) the regression coefficients a and b for each similar pixel are calculated from the available coarse-resolution reflectance images. Finally, as in (5), the weights and regression coefficients are applied

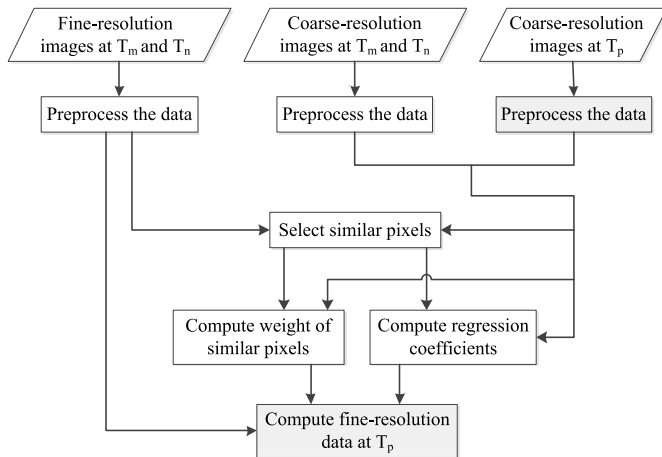


Fig. 3. Flowchart of the STNLFFM algorithm.

to the two fine-resolution reflectance images acquired at the reference dates to produce the fine-resolution reflectance at the desired prediction date. All of the steps are discussed in detail in the following. The flowchart of the STNLFFM method is shown in Fig. 3.

E. Differences Between the STNLFFM Method and the STARFM/ESTARFM Methods

Although they all belong to the filter-based fusion framework, the proposed STNLFFM method and the conventional STARFM or ESTARFM methods have big differences, including the fundamental assumptions, the form of the prediction models, the method of weight calculation, and the procedure of searching for similar pixels.

1) *Differences in the Starting Points and Fundamental Assumptions:* Notably, the starting points and fundamental assumptions of the proposed STNLFFM method and the conventional STARFM/ESTARFM methods are different. For the STARFM and ESTARFM methods, their starting point is the transformation relationship between the fine-resolution and coarse-resolution reflectance images, and their fundamental assumption is that the transformation relationship between the fine-resolution and coarse-resolution reflectance is constant at different dates. However, for the proposed STNLFFM method, the starting point is the transformation relationship between the multitemporal reflectance images, and the fundamental assumption is that the transformation relationship between the multitemporal reflectance is constant for different sensors. These differences can be seen in the initial prediction models of the three methods.

2) *Differences in the Prediction Models:* Due to the differences in the starting points and fundamental assumptions, the prediction models are also different. In the prediction model of STNLFFM, the regression coefficients a and b are introduced to describe the reflectance changes between the predicted and reference high-resolution reflectance images. For the STARFM and ESTARFM models, they both take the reflectance changes between the predicted and reference high-resolution reflectance images as being perfectly linear with the gain coefficient a equal to 1. As noted before, the STARFM model is a special case of the proposed STNLFFM model with

the gain coefficient a equal to 1. However, this assumption may not be appropriate for complex changed landscapes. Moreover, the proposed STNLFFM model calculates the regression coefficients a and b from the available multitemporal coarse-resolution reflectance images. As for the ESTARFM method, it calculates a conversion coefficient v from the available fine-resolution and coarse-resolution image pairs. That is to say, the method of calculation of the prediction models is also different. Overall, the proposed STNLFFM model pays more attention to the reflectance changes in the multitemporal images, and it employs the gain coefficient a and bias coefficient b to more accurately describe the reflectance changes over time, thereby improving the prediction accuracy, especially for complex changed landscapes.

Moreover, in the prediction model, STNLFFM employs the idea of nonlocal filtering to take advantage of the redundant information in the remote sensing image sequence. Both the local neighborhood similarity information and the nonlocal similarity information are used to jointly produce the unknown pixels. As for some ground objects with small areas or edge structures, the number of neighborhood similar pixels may be small and not enough to provide an accurate prediction. Thus, the nonlocal similarity can be used to solve this problem, and provide a more robust prediction.

3) *Differences in the Weight Calculation:* The weight calculation is also an important step in the fusion methods. STNLFFM separates the weight calculation into individual weight and whole weight calculations. The individual weight is used to measure which similar pixels within a fine-resolution image provide more reliable information; the whole weight is used to measure which fine-resolution image input provides more reliable information. For the individual weight calculation, the original STARFM algorithm takes the spectral difference, the temporal difference, and the location distance into consideration. However, the many terms make the weight calculation complex. According to the nonlocal filtering method, the STNLFFM algorithm only uses the reflectance similarity of patches in the multitemporal images to calculate the individual weight, which avoids introducing extra error, improves the calculation efficiency, and can reduce the interference of image noise.

Furthermore, in the procedure of selecting similar pixels, the STNLFFM algorithm resolves the problem of temporal difference measurement. Compared with the STARFM algorithm, which takes a pixel whose temporal change between the reference date t_k and the prediction date t_p is smaller than that of the target pixel as a candidate similar pixel, the proposed STNLFFM algorithm selects a pixel whose temporal change between dates t_k and t_p is closer to that of the target pixel as a candidate similar pixel, which is more reasonable. This approach can significantly improve the accuracy of selecting similar pixels, especially for the target pixels with land-cover type changes.

III. EXPERIMENTS

A. Study Sites and Data

The study sites and data tested in this paper are the same as those used in the research of Emelyanova *et al.* [36].

The first study site is the Coleambally Irrigation Area (CIA) located in southern New South Wales, for which 17 cloud-free Landsat-MODIS pairs are available for the 2001–2002 austral summer growing season. The other study site is the Lower Gwydir Catchment (LGC) located in northern New South Wales, for which 14 cloud-free Landsat-MODIS pairs are available from April 2004 to April 2005. All the Landsat images were atmospherically corrected in the same way as the research of Emelyanova *et al.* [36]: the CIA images were atmospherically corrected using MODTRAN4 [37], and the LGC images were atmospherically corrected using Li *et al.*'s algorithm [38]. For both study sites, the latest MODIS Terra MOD09GA Collection 6 data were used. These data were upsampled to the same spatial resolution (25 m) as the Landsat data using a cubic convolution algorithm.

The CIA has an overall area of 2193 km² (1720 × 2040 pixels in Landsat images). The irrigation area, which is scattered over the whole study site, exhibits temporal dynamics associated with crop phenology over a single growing season. However, the surrounding agricultural and woodland areas vary less over time. Due to the small field sizes and sporadic distribution of the irrigation area, the CIA can be considered a spatially heterogeneous site. The LGC site has an overall area of 5440 km² (3200 × 2720 pixels in Landsat images). The temporal extent of the LGC data is approximately one year. A large flood occurred in mid-December 2004, which caused inundation over large areas (about 44%). Due to the flooding event leading to different spatial and temporal variations, the LGC is considered a temporally dynamic site. Some of the temporal data from the two sites are shown in Fig. 5.

B. Quantitative Evaluation Indices

A Landsat-like image on a certain date is predicted using two Landsat-MODIS pairs on other dates and an MODIS image on the prediction date, as shown in detail in the following. The predicted Landsat image is compared with the real observed Landsat image acquired at the prediction date. (The real observed Landsat images at the prediction date are not used as input, and are only used for the validation.) Several statistical indices are used to give a quantitative assessment. The first index is the root-mean-square error (RMSE), which is used to give a global depiction of the radiometric difference between the predicted image and the real observed image. It is defined as

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (P_i - O_i)^2}{N}} \quad (14)$$

where N is the total number of pixels in the predicted image, and P_i and O_i are the values of the i th pixel in the predicted image and the observed image, respectively. The smaller the RMSE, the better the prediction.

The second index is the coefficient of determination (R^2). This metric can be used to show the degree of consistency between the predicted and observed images. It is defined as

$$R^2 = \left(\frac{\sum_{i=1}^N (P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^N (P_i - \bar{P})^2} \sqrt{\sum_{i=1}^N (O_i - \bar{O})^2}} \right)^2 \quad (15)$$

where \bar{P} and \bar{O} represent the mean gray values of the predicted image and the observed image, respectively. A larger R^2 indicates a closer consistency between the two groups of pixels.

The third index is the universal image quality index (UIQI) [39]. This indicator provides a measure of the closeness between the predicted and observed images by utilizing the differences in the statistical distributions

$$\text{UIQI} = \frac{|\sigma_{PO}|}{\sigma_{PO}} \cdot \frac{2\sigma_P \cdot \sigma_O}{\sigma_O^2 + \sigma_P^2} \cdot \frac{2|\bar{P}| \cdot |\bar{O}|}{|\bar{P}|^2 + |\bar{O}|^2} \quad (16)$$

where σ_P and σ_O represent the variances of the predicted and observed images, respectively, and σ_{PO} represents the covariance between the predicted and observed images.

The fourth index is the structural similarity (SSIM) index [40], which is used to show the similarity of the image structures between the predicted and observed data. A larger SSIM means a better result. This metric is defined as

$$\text{SSIM} = \frac{(2\mu_P \mu_O + C_1)(2\sigma_{PO} + C_2)}{(\mu_P^2 + \mu_O^2 + C_1)(\sigma_P^2 + \sigma_O^2 + C_2)} \quad (17)$$

where C_1 and C_2 are two constants.

The fifth index is the spectral angle mapper (SAM) [41]. This metric reflects the spectral distortion introduced by the fusion process, which gives a quantitative assessment from the spectral fidelity aspect. A smaller SAM means a better result. It is defined as

$$\text{SAM} = \frac{1}{N} \sum_{i=1}^N \arccos \left(\frac{\sum_{j=1}^M (P_j^i O_j^i)}{\sqrt{\sum_{j=1}^M (P_j^i)^2} \sqrt{\sum_{j=1}^M (O_j^i)^2}} \right) \quad (18)$$

where M is the total number of bands.

The sixth index is the $Q4$ index [42], which is an extension of the UIQI index. This metric is a product of the correlation, mean bias, and contrast variation, and is suitable for assessing the prediction images from both spatial and spectral fidelity aspects. The ideal value for the $Q4$ index is 1, and it is defined as

$$Q4 = \frac{|\sigma_{Z_1 Z_2}|}{\sigma_{Z_1 Z_2}} \cdot \frac{2\sigma_{Z_1} \cdot \sigma_{Z_2}}{\sigma_{Z_1}^2 + \sigma_{Z_2}^2} \cdot \frac{2|\bar{Z}_1| \cdot |\bar{Z}_2|}{|\bar{Z}_1|^2 + |\bar{Z}_2|^2} \quad (19)$$

where Z_1 and Z_2 are defined as $Z_1 = P_1 + iP_2 + jP_3 + kP_4$ and $Z_2 = O_1 + iO_2 + jO_3 + kO_4$, respectively. $\sigma_{Z_1 Z_2}$ is the covariance between Z_1 and Z_2 . σ_{z_1} and σ_{z_2} are the variances of Z_1 and Z_2 , respectively. \bar{Z}_1 and \bar{Z}_2 are the mean values of Z_1 and Z_2 , respectively.

C. Experimental Results

To verify the efficacy of the STNLFFM algorithm, we conducted two groups of experiments: 1) short time series fusion experiments and 2) long time series fusion experiments.

1) *Short Time Series Fusion Experiments*: For both study sites, all the Landsat-MODIS pairs were arranged in chronological order. The STNLFFM algorithm was tested by predicting a Landsat-like image on a certain date using two Landsat-MODIS pairs that were the nearest temporal neighbors to the predicted date, one before and one

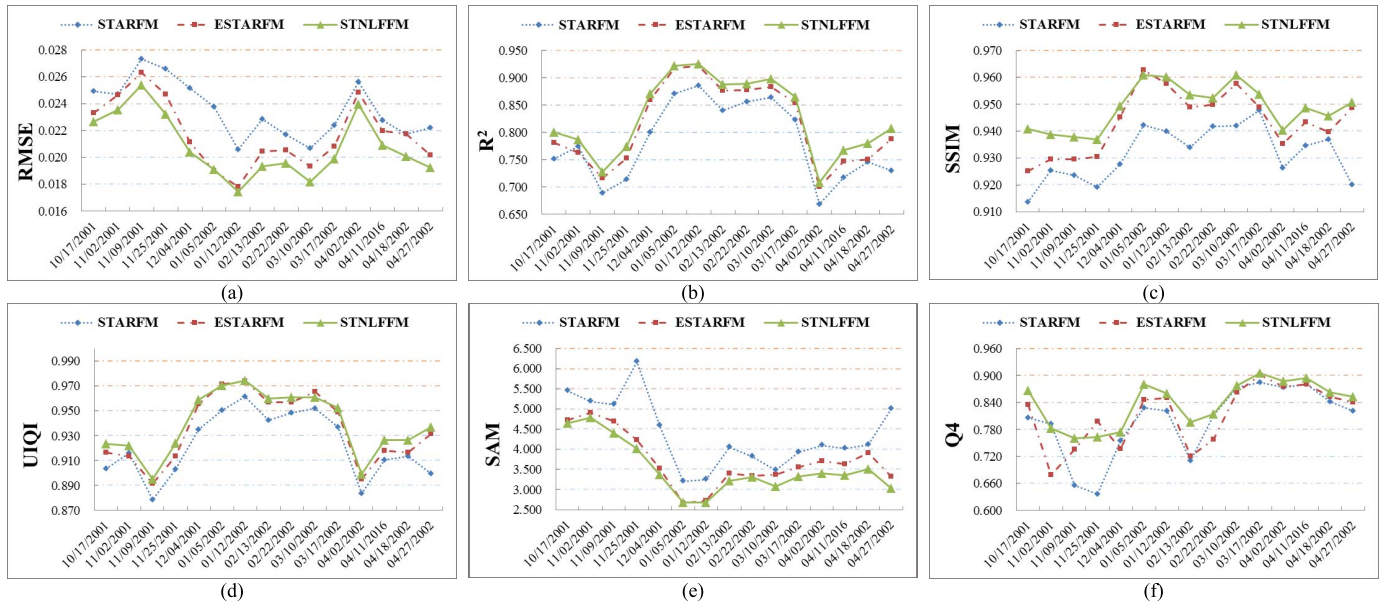


Fig. 4. Quantitative assessment results for all the prediction dates at the CIA site. (a) RMSE. (b) R^2 . (c) SSIM. (d) UIQI. (e) SAM. (f) $Q4$.

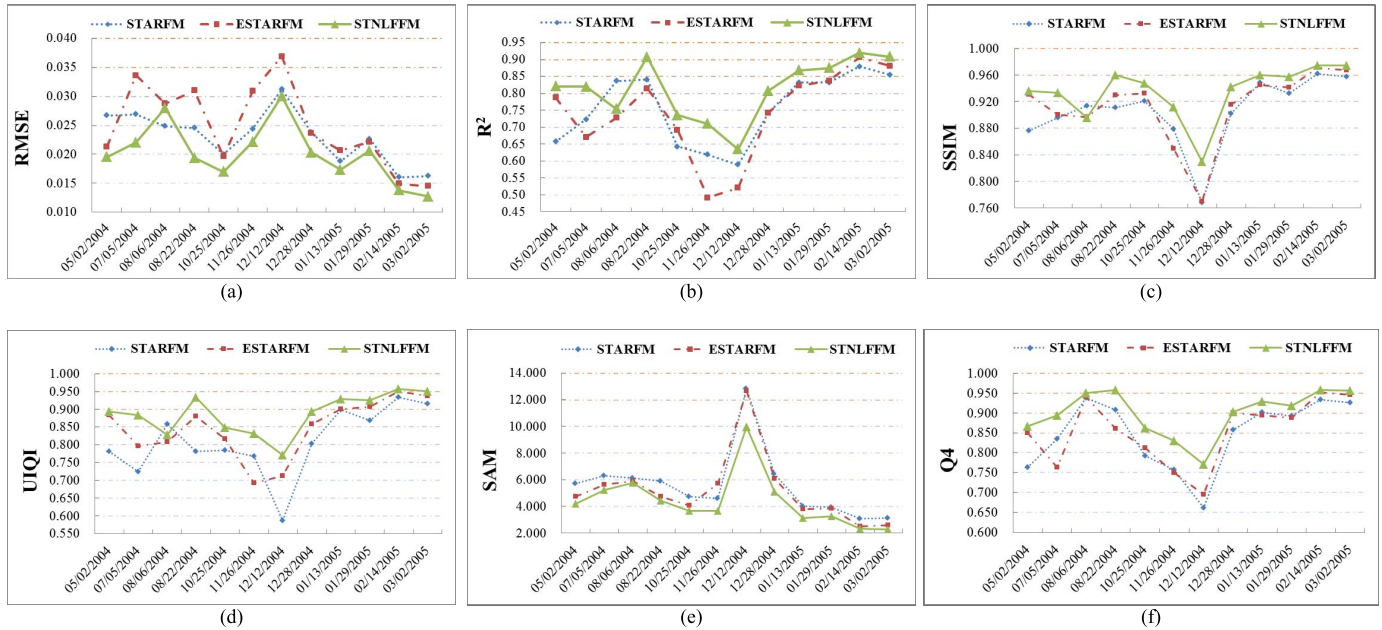


Fig. 5. Quantitative assessment results for all the prediction dates at the LGC site. (a) RMSE. (b) R^2 . (c) SSIM. (d) UIQI. (e) SAM. (f) $Q4$.

after, and the MODIS image on the predicted date was also used as an input. All possible combinations of predictions were processed at both study sites. For the CIA site, Landsat-like images at 15 middle dates were predicted, and for the LGC site, Landsat-like images at 12 middle dates were predicted, except for the first and last dates.

To allow a comparative analysis, the STNLFFM method was compared with two popular methods: STARFM and ESTARFM. The quantitative assessment results for all the prediction dates for the CIA site and the LGC site are shown in Figs. 4 and 5, respectively. We can see from Figs. 4 and 5 that, in the vast majority of the 15 predicted results of the CIA site and the 12 predicted results of the LGC site, the RMSE

and SAM values obtained using the STNLFFM method are the lowest, and the R^2 , SSIM, UIQI, and $Q4$ values obtained using the STNLFFM method are the highest. That is to say, the proposed method is able to provide prediction data with higher fidelity from the radiometric (gray level), spatial structure, and spectral aspects. Furthermore, the proposed STNLFFM method is also robust.

We also present some details of the test data and the above-mentioned results. Fig. 6 shows the observed Landsat-MODIS pairs on a key date and the two nearest dates at the CIA and LGC sites, respectively. For the CIA site, the images are presented as Landsat bands 4, 3, and 2 (MODIS 2, 1, and 4), displayed as RGB, as shown in Fig. 6(a). We can see from

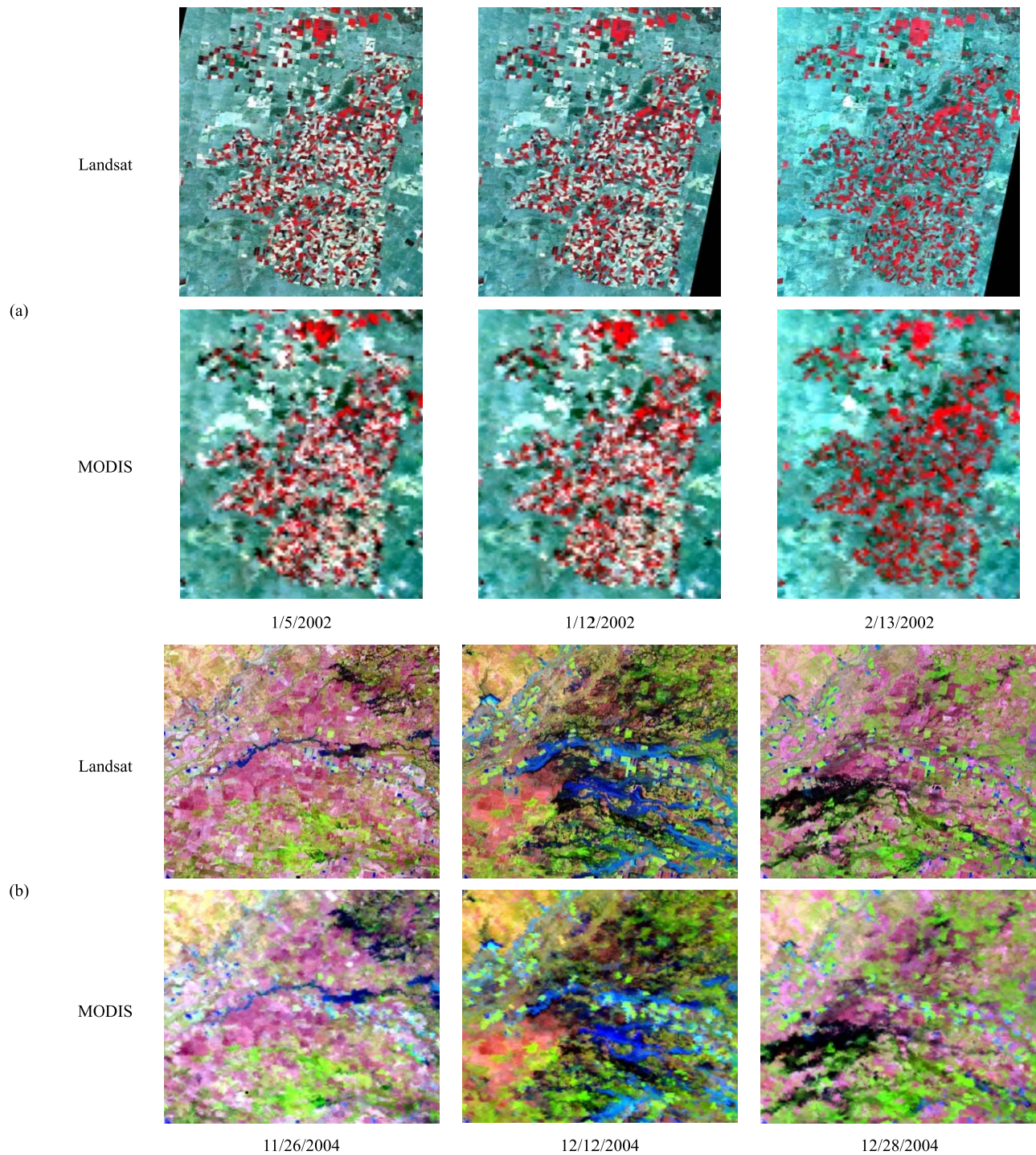


Fig. 6. Observed Landsat-MODIS pairs on a key date and the two nearest dates at the CIA and LGC sites. (a) Observed Landsat-MODIS pairs at the CIA site on January 5, 2002, January 12, 2002, and February 13, 2002, respectively. (b) Observed Landsat-MODIS pairs at the LGC site on November 26, 2004, December 12, 2004, and December 28, 2004, respectively.

Fig. 6(a) that the crop in the sporadic irrigation fields begins to turn green through January to February, but the surrounding agricultural and woodland areas show less change during this time. This makes the CIA area spatially heterogeneous. For the LGC site, the large flood occurred in mid-December 2004, causing temporal dynamics and abnormal change of the ground surface. In order to show the ground features of the LGC site more clearly, the LGC images are presented as

Landsat bands 5, 4, and 3 (MODIS 6, 2, and 1), displayed as RGB, as shown in Fig. 6(b). We use the Landsat-MODIS pair dates before and after to predict a Landsat-like image at the middle date. The prediction results for these two groups of data are shown in Figs. 7 and 8.

For the results of the CIA study site (Fig. 7), it can be seen that all the three methods are generally able to predict the crop phenological changes, and the results are satisfactory in most

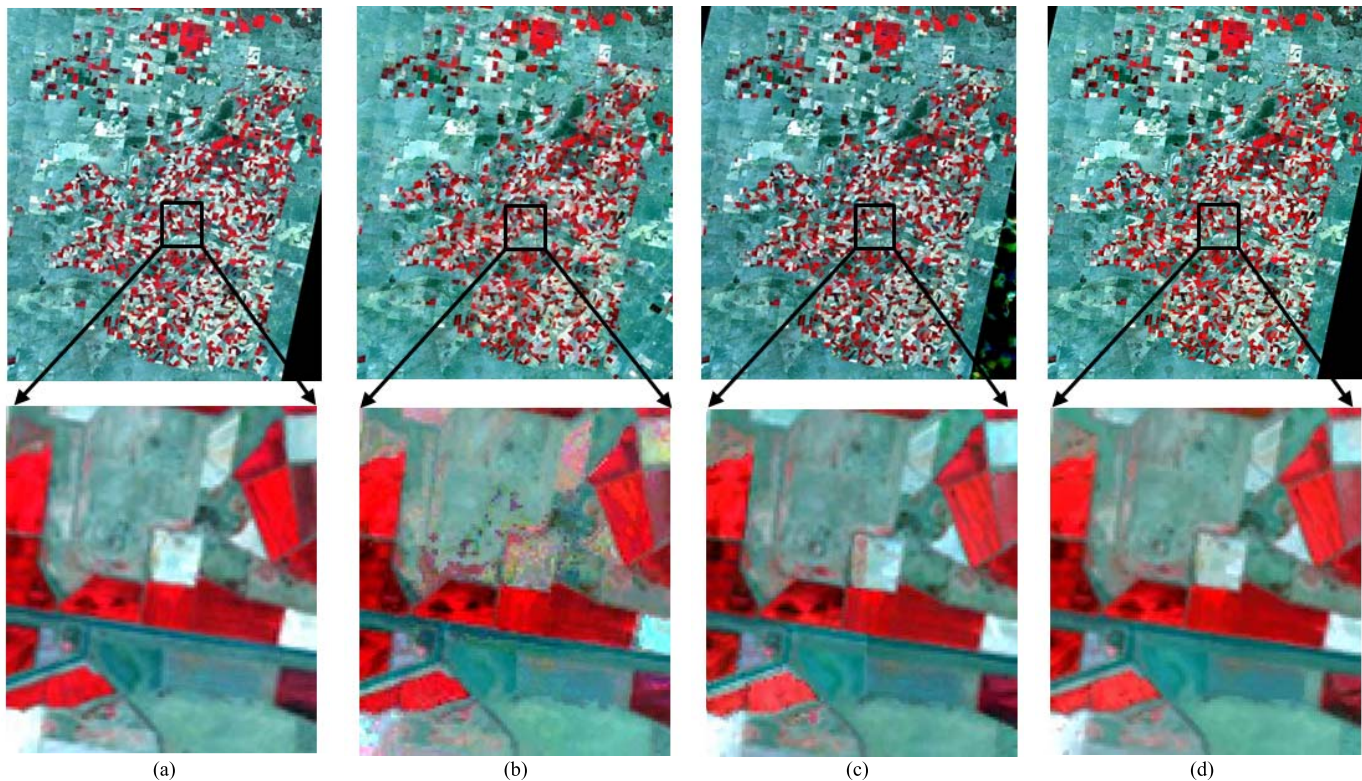


Fig. 7. Prediction results for the CIA site on January 12, 2002. (a) Observed Landsat image. (b)–(d) Landsat-like images predicted by STARFM, ESTARFM, and the proposed STNLFFM method, respectively.

regions. However, for some special heterogeneous regions, such as the zoomed-in-view detailed regions in Fig. 7, we can see that, for the STARFM and ESTARFM methods, they produce spectral distortion in their results [Fig. 7(b) and (c)], especially for the STARFM method, which is poor at handling spatially heterogeneous areas. For the proposed STNLFFM method, it obtains a visually convincing result, which is closest to the observed Landsat data, as shown in Fig. 7(d). For the test results of the LGC study site (Fig. 8), we can see from the zoomed-in-view detailed regions that there is some noise arising in the edge regions for the STARFM method [Fig. 8(b)]. For the result of the ESTARFM method [Fig. 8(c)], obvious spectral distortion occurs in the flooded area, since ESTARFM is less good at handling temporally dynamic areas [36]. The proposed STNLFFM method obtains a result [Fig. 8(d)] that is visually similar to the observed Landsat data. The quantitative assessment results for Figs. 7 and 8 are shown in Tables I and II, respectively. It can be seen from Table I that, for a few bands of the CIA data, such as band 2 (green) and band 3 (red) (Fig. 7), the ESTARFM method obtains the best quantitative values. Meanwhile, most of the results obtained using STNLFFM have the best quantitative values, and the improvement is remarkable, especially for the LGC data in Fig. 8 (Table II).

2) *Long Time Series Fusion Experiments*: In this part, we analyze the influence of the time interval length on the prediction results. The dates of the LGC image series are irregular and have quite different time intervals, while the dates of the CIA image series are regular and have

TABLE I
QUANTITATIVE VALUES OF THE PREDICTION RESULTS IN FIG. 7

		STARFM	ESTARFM	STNLFFM
RMSE	B1	0.0102	0.0098	0.0093
	B2	0.0131	0.0104	0.0104
	B3	0.0213	0.0162	0.0166
	B4	0.0283	0.0227	0.0218
	B5	0.0270	0.0249	0.0249
	B6	0.0236	0.0229	0.0216
	B7	0.0236	0.0229	0.0216
R ²	B1	0.8770	0.9209	0.9259
	B2	0.8846	0.9307	0.9325
	B3	0.8937	0.9391	0.9380
	B4	0.8094	0.8827	0.8895
	B5	0.9229	0.9330	0.9325
	B6	0.9233	0.9277	0.9345
	B7	0.9233	0.9277	0.9345
SSIM	B1	0.9551	0.9680	0.9701
	B2	0.9451	0.9679	0.9675
	B3	0.9296	0.9625	0.9584
	B4	0.8780	0.9288	0.9299
	B5	0.9432	0.9517	0.9497
	B6	0.9431	0.9471	0.9515
	B7	0.9431	0.9471	0.9515
UIQI	B1	0.9518	0.9661	0.9684
	B2	0.9587	0.9762	0.9758
	B3	0.9568	0.9767	0.9742
	B4	0.9534	0.9718	0.9733
	B5	0.9763	0.9783	0.9790
	B6	0.9722	0.9737	0.9762
	B7	0.9722	0.9737	0.9762
SAM	3.2569	2.7311	2.6838	
Q4	0.8210	0.8527	0.8600	

similar time intervals. In order to make a convenient and effective analysis, we only used the CIA image series in the experiments. The Landsat-MODIS pairs of the CIA site were

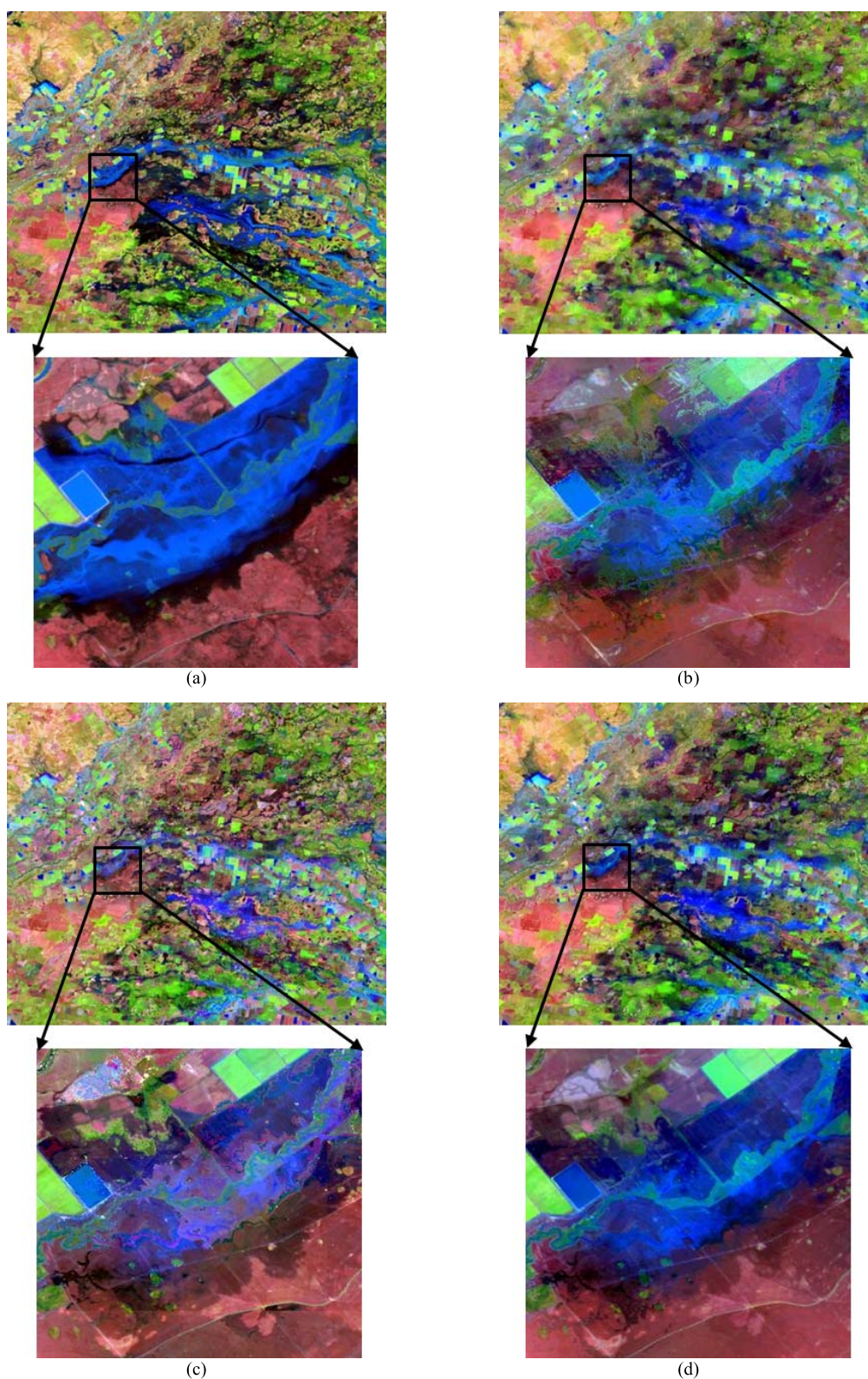


Fig. 8. Prediction results for the LGC site on December 12, 2004. (a) Observed Landsat image. (b)–(d) Landsat-like images predicted by STARFM, ESTARFM, and the proposed STNLFFM method, respectively.

arranged in chronological order, as before. The middlemost date, February 13, 2002, was treated as the prediction date. The proposed STNLFFM algorithm was tested by predicting a Landsat-like image on the prediction date (February 13, 2002)

using two Landsat-MODIS pairs on reference dates that were symmetrically distributed with the prediction date, one before and one after, as shown in Fig. 9. As the time intervals between the prediction date and the reference dates were increased,

TABLE II
QUANTITATIVE VALUES OF THE PREDICTION RESULTS IN FIG. 8

		STARFM	ESTARFM	STNLFFM
RMSE	B1	0.0138	0.0148	0.0135
	B2	0.0198	0.0203	0.0189
	B3	0.0246	0.0255	0.0236
	B4	0.0346	0.0408	0.0317
	B5	0.0535	0.0631	0.0516
	B7	0.0409	0.0567	0.0407
	R ²	B1	0.5360	0.5482
B2		0.5060	0.5180	0.5563
B3		0.5298	0.5343	0.5721
B4		0.7004	0.5906	0.7672
B5		0.6394	0.5315	0.6805
B7		0.6230	0.4015	0.6438
SSIM		B1	0.8898	0.8863
	B2	0.8324	0.8337	0.8497
	B3	0.8044	0.8052	0.8224
	B4	0.8052	0.7518	0.8491
	B5	0.6645	0.6834	0.7705
	B7	0.5719	0.6136	0.7551
	UIQI	B1	0.7235	0.7540
B2		0.6941	0.7377	0.7386
B3		0.7131	0.7487	0.7494
B4		0.7591	0.8435	0.8541
B5		0.4136	0.7369	0.7815
B7		0.2149	0.6841	0.7509
SAM		12.8043	11.3013	9.9631
Q4	0.6618	0.6949	0.7713	

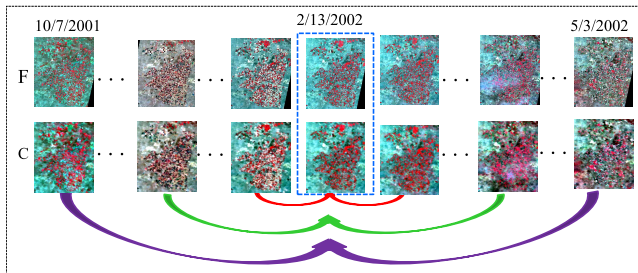


Fig. 9. Predicting a Landsat-like image on the middlemost date (February 13, 2002) using two Landsat-MODIS pairs on reference dates that are symmetrically distributed, one before and one after.

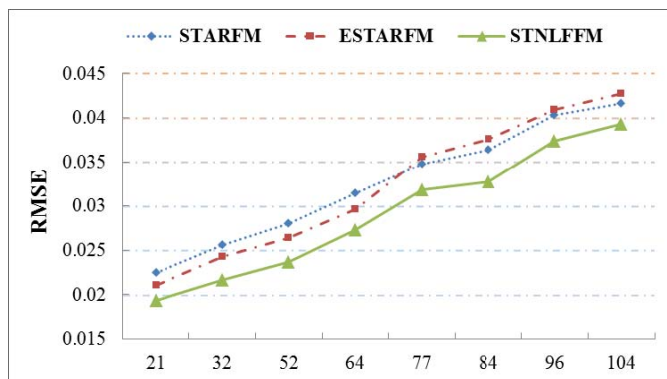


Fig. 10. RMSE values of the predicted results. The horizontal axis represents the average time interval between the prediction date and the two reference dates.

all possible combinations of predictions were processed. The RMSE values of the predicted results were calculated, as shown in Fig. 10. The horizontal axis in Fig. 10 represents the average time interval between the prediction date and

TABLE III
MEAN COMPUTATION TIMES OF THE STARFM, ESTARFM, AND STNLFFM METHODS

Method	Computation time (s)
STARFM	1296
ESTARFM	3866
STNLFFM	901

the two reference dates. From Fig. 10, we can observe the following phenomena.

- 1) For all three methods, the shorter the time interval, the better the prediction, and vice versa. When the average time interval is longer than about 90 days, the prediction accuracy fluctuates less.
- 2) From the two curves of STARFM and ESTARFM, we can see that when the average time interval is short (less than about 65 days), the prediction of ESTARFM is better than that of STARFM. This is because the spatial variability is the dominant factor affecting the prediction when the time interval is short, and ESTARFM is better able to deal with spatial variability than STARFM [20]; however, when the average time interval is long (more than about 65 days), the prediction of STARFM is better than that of ESTARFM. This is because the temporal variability becomes the dominant factor affecting the prediction when the time interval is sufficiently long, and STARFM is better able to deal with temporal variability than ESTARFM [36]. Moreover, we can infer that the intersection of the two curves (STARFM and ESTARFM) is the point where the levels of spatial and temporal variability are equal, and this point is located at approximately 65 days (the average time interval between the prediction date and the two reference dates) for the CIA site.
- 3) For every prediction result in Fig. 10, the RMSE value of the STNLFFM method is lower than those of the other two methods, which suggests that, with the interference from both the spatial variability and temporal variability, the STNLFFM method can still obtain better prediction results than the other two methods.

Finally, we compared the computation times of the proposed STNLFFM method and the STARFM and ESTARFM methods on a PC with an Intel Core i3 3.4-GHz CPU and 8 GB of RAM. We took nine groups of image blocks with the size of 1000×1000 pixels from the CIA and LGC data, and then recorded the mean time consumption of the nine groups of fusion experiments, as listed in Table III. It can be seen in Table III that the STNLFFM method is slightly faster than the STARFM method, and it is about four times faster than the ESTARFM method. It is suggested that, although STNLFFM expands the search space in a nonlocal manner to select similar pixels, the effective ways of identifying similar pixels and calculating the weights improve the computational efficiency. However, we can see that the STNLFFM method is still time-consuming, and should be further speeded up before being used in large-scale applications.

IV. CONCLUSION

In this paper, in order to obtain remote sensing data with both high spatial resolution and high temporal frequency, we have proposed the STNLFFM. The STNLFFM algorithm was tested over the CIA study site and the LGC study site, and we conducted two groups of experiments to verify the efficacy of the STNLFFM algorithm. The experimental results show that the STNLFFM algorithm can predict the fine-resolution reflectance accurately and robustly, for both heterogeneous landscapes and temporally dynamic areas.

The proposed STNLFFM algorithm makes several improvements to the STARFM and ESTARFM algorithms. First, in the prediction model, the STNLFFM algorithm uses two regression coefficients to more accurately describe the land-cover change information, thereby enhancing the prediction capability for complex changed landscapes. Second, STNLFFM introduces the idea of nonlocal filtering, which takes advantage of the high degree of redundancy in the image sequence to produce a more accurate and robust prediction. Third, STNLFFM uses a simple method of weight calculation, which can improve the computational efficiency and reduce the interference of image noise. Finally, STNLFFM solves the problem of temporal difference measurement in the procedure of searching for similar pixels, and improves the accuracy of similar pixel selection.

There are, however, some limitations to the STNLFFM method. STNLFFM is based on the assumption that the reflectance change rate is linear. However, this assumption might not be appropriate in some situations, especially over a long time period. In addition, although the current computation speed of the STNLFFM algorithm is faster than that of the STARFM and ESTARFM algorithms, the STNLFFM algorithm is still time-consuming, and the calculation efficiency needs to be further improved.

REFERENCES

- [1] B. Chen, B. Huang, and B. Xu, "Comparison of spatiotemporal fusion models: A review," *Remote Sens.*, vol. 7, no. 2, pp. 1798–1835, 2015.
- [2] H. K. Zhang, B. Huang, M. Zhang, K. Cao, and L. Yu, "A generalization of spatial and temporal fusion methods for remotely sensed surface parameters," *Int. J. Remote Sens.*, vol. 36, no. 17, pp. 4411–4445, 2015.
- [3] W. G. Rees, M. Williams, and P. Vitebsky, "Mapping land cover change in a reindeer herding area of the Russian Arctic using Landsat TM and ETM+ imagery and indigenous knowledge," *Remote Sens. Environ.*, vol. 85, no. 4, pp. 441–452, 2003.
- [4] O. R. van Lier, J. E. Luther, D. G. Leckie, and W. W. Bowers, "Development of large-area land cover and forest change indicators using multi-sensor Landsat imagery: Application to the Humber River Basin, Canada," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 13, no. 5, pp. 819–829, 2011.
- [5] W. B. Cohen and S. N. Goward, "Landsat's role in ecological applications of remote sensing," *Bioscience*, vol. 54, no. 6, pp. 535–545, 2004.
- [6] J. G. Masek *et al.*, "North American forest disturbance mapped from a decadal Landsat record," *Remote Sens. Environ.*, vol. 112, no. 6, pp. 2914–2926, 2008.
- [7] M. A. Friedl *et al.*, "Global land cover mapping from MODIS: Algorithms and early results," *Remote Sens. Environ.*, vol. 83, nos. 1–2, pp. 287–302, 2002.
- [8] H. Zhou, E. Aizen, and V. Aizen, "Deriving long term snow cover extent dataset from AVHRR and MODIS data: Central Asia case study," *Remote Sens. Environ.*, vol. 136, pp. 146–162, Sep. 2013.
- [9] F. Gao, "Integrating Landsat with MODIS products for vegetation monitoring," in *Satellite-Based Applications on Climate Change*. Dordrecht, The Netherlands: Springer, 2013, pp. 247–261.
- [10] E. Arai, Y. E. Shimabukuro, G. Pereira, and N. L. Vijaykumar, "A multi-resolution multi-temporal technique for detecting and mapping deforestation in the Brazilian Amazon rainforest," *Remote Sens.*, vol. 3, no. 9, pp. 1943–1956, 2011.
- [11] W. Zhang *et al.*, "An enhanced spatial and temporal data fusion model for fusing Landsat and MODIS surface reflectance to generate high temporal Landsat-like data," *Remote Sens.*, vol. 5, no. 10, pp. 5346–5368, 2013.
- [12] F. Gao, J. Masek, M. Schwaller, and F. Hall, "On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2207–2218, Aug. 2006.
- [13] T. Hilker *et al.*, "Generation of dense time series synthetic Landsat data through data blending with MODIS using a spatial and temporal adaptive reflectance fusion model," *Remote Sens. Environ.*, vol. 113, no. 9, pp. 1988–1999, 2009.
- [14] D. Singh, "Generation and evaluation of gross primary productivity using Landsat data through blending with MODIS data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 13, no. 1, pp. 59–69, 2011.
- [15] J. J. Walker, K. M. de Beurs, R. H. Wynne, and F. Gao, "Evaluation of Landsat and MODIS data fusion products for analysis of dryland forest phenology," *Remote Sens. Environ.*, vol. 117, pp. 381–393, Feb. 2012.
- [16] M. C. Anderson *et al.*, "Mapping daily evapotranspiration at field to continental scales using geostationary and polar orbiting satellite imagery," *Hydrol. Earth Syst. Sci.*, vol. 15, pp. 223–239, Jan. 2011.
- [17] H. Liu and Q. Weng, "Enhancing temporal resolution of satellite imagery for public health studies: A case study of West Nile Virus outbreak in Los Angeles in 2007," *Remote Sens. Environ.*, vol. 117, pp. 57–71, Feb. 2012.
- [18] H. Shen, L. Huang, L. Zhang, P. Wu, and C. Zeng, "Long-term and fine-scale satellite monitoring of the urban heat island effect by the fusion of multi-temporal and multi-sensor remote sensed data: A 26-year case study of the city of Wuhan in China," *Remote Sens. Environ.*, vol. 172, pp. 109–125, Jan. 2016.
- [19] T. Hilker *et al.*, "A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on Landsat and MODIS," *Remote Sens. Environ.*, vol. 113, no. 8, pp. 1613–1627, 2009.
- [20] X. Zhu, J. Chen, F. Gao, X. Chen, and J. G. Masek, "An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions," *Remote Sens. Environ.*, vol. 114, no. 11, pp. 2610–2623, 2010.
- [21] D. Fu, B. Chen, J. Wang, X. Zhu, and T. Hilker, "An improved image fusion approach based on enhanced spatial and temporal the adaptive reflectance fusion model," *Remote Sens.*, vol. 5, no. 12, pp. 6346–6360, 2013.
- [22] H. Shen, P. Wu, Y. Liu, T. Ai, Y. Wang, and X. Liu, "A spatial and temporal reflectance fusion model considering sensor observation differences," *Int. J. Remote Sens.*, vol. 34, no. 12, pp. 4367–4383, 2013.
- [23] P. Wu, H. Shen, L. Zhang, and F.-M. Göttsche, "Integrated fusion of multi-scale polar-orbiting and geostationary satellite observations for the mapping of high spatial and temporal resolution land surface temperature," *Remote Sens. Environ.*, vol. 156, pp. 169–181, Jan. 2015.
- [24] M. C. Hansen, D. P. Roy, E. Lindquist, B. Adusei, C. O. Justice, and A. Altstatt, "A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in the Congo Basin," *Remote Sens. Environ.*, vol. 112, no. 5, pp. 2495–2513, 2008.
- [25] R. Zurita-Milla, G. Kaiser, J. G. P. W. Clevers, W. Schneider, and M. E. Schaepman, "Downscaling time series of MERIS full resolution data to monitor vegetation seasonal dynamics," *Remote Sens. Environ.*, vol. 113, no. 9, pp. 1874–1885, 2009.
- [26] B. Huang and H. Song, "Spatiotemporal reflectance fusion via sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3707–3716, Oct. 2012.
- [27] H. Song and B. Huang, "Spatiotemporal satellite image fusion through one-pair image learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 1883–1896, Apr. 2013.
- [28] H. Shen, M. K. Ng, P. Li, and L. Zhang, "Super-resolution reconstruction algorithm to MODIS remote sensing images," *Comput. J.*, vol. 52, no. 1, pp. 90–100, 2009.
- [29] H. Shen, L. Peng, L. Yue, Q. Yuan, and L. Zhang, "Adaptive norm selection for regularized image restoration and super-resolution," *IEEE Trans. Cybern.*, vol. 46, no. 6, pp. 1388–1399, Jun. 2016.
- [30] H. Shen, L. Zhang, B. Huang, and P. Li, "A MAP approach for joint motion estimation, segmentation, and super resolution," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 479–490, Feb. 2007.

- [31] G. Gilboa and S. Osher, "Nonlocal operators with applications to image processing," *Multiscale Model. Simul.*, vol. 7, no. 3, pp. 1005–1028, 2008.
- [32] Q. Cheng, H. Shen, L. Zhang, and P. Li, "Inpainting for remotely sensed images with a multichannel nonlocal total variation model," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 175–187, Jan. 2014.
- [33] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 60–65.
- [34] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, 2005.
- [35] X. Su, C.-A. Deledalle, F. Tupin, and H. Sun, "Two steps multi-temporal non-local means for SAR images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2012, pp. 2008–2011.
- [36] I. V. Emelyanova, T. R. McVicar, T. G. Van Niel, L. T. Li, and A. I. J. M. van Dijk, "Assessing the accuracy of blending Landsat–MODIS surface reflectances in two landscapes with contrasting spatial and temporal dynamics: A framework for algorithm selection," *Remote Sens. Environ.*, vol. 133, pp. 193–209, Jun. 2013.
- [37] A. Berk *et al.*, "MODTRAN4 radiative transfer modeling for atmospheric correction," *Proc. SPIE*, vol. 3756, pp. 348–353, Oct. 1999.
- [38] F. Li *et al.*, "An evaluation of the use of atmospheric and BRDF correction to standardize Landsat data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 3, no. 3, pp. 257–270, Sep. 2010.
- [39] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.
- [40] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [41] P. E. Dennison, K. Q. Halligan, and D. A. Roberts, "A comparison of error metrics and constraints for multiple endmember spectral mixture analysis and spectral angle mapper," *Remote Sens. Environ.*, vol. 93, no. 3, pp. 359–367, 2004.
- [42] L. Alparone, S. Baronti, A. Garzelli, and F. Nencini, "A global quality measurement of pan-sharpened multispectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 4, pp. 313–317, Oct. 2004.



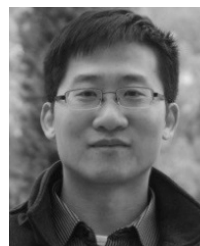
Qing Cheng received the B.S. degree in geographic information system and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2010 and 2015, respectively.

She is currently a Lecturer with the School of Urban Design, Wuhan University. Her research interests include remote sensing data reconstruction, data fusion, and urban remote sensing.



Huiqing Liu received the B.S. degree in geographic information system from Huazhong Agricultural University, Wuhan, China, in 2013, and the M.S. degree in cartography and geographic information system from Wuhan University, Wuhan, in 2016.

She is currently with the Surveying and Mapping Institute Lands and Resource Department of Guangdong Province, Guangzhou, China.

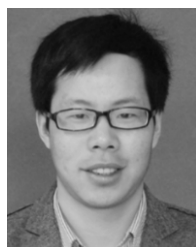


Huanfeng Shen (M'10–SM'13) received the B.S. degree in surveying and mapping engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2002 and 2007, respectively.

In 2007, he joined the School of Resource and Environmental Sciences, Wuhan University, where he is currently a Luojia Distinguished Professor. He has authored over 100 research papers. He has been supported by several talent programs, such as The Youth Talent Support Program of China in 2015,

the China National Science Fund for Excellent Young Scholars in 2014, and the New Century Excellent Talents by the Ministry of Education of China in 2011. His research interests include image quality improvement, remote sensing mapping and application, data fusion and assimilation, and regional and global environmental change.

Dr. Shen is currently a member of the Editorial Board of the *Journal of Applied Remote Sensing*.



Penghai Wu (M'16) received the B.S. degree in environmental science from Anqing Normal College, Anqing, China, in 2009, and the M.S. degree in surveying and mapping engineering and the Ph.D. degree in cartography and geographical information engineering from Wuhan University, Wuhan, China, in 2011 and 2014, respectively.

He is currently a Lecturer with the School of Resources and Environmental Engineering, Anhui University, Hefei, China. His research interests include spatiotemporal fusion, the disaggregation of land surface temperature, the reconstruction of land surface temperature, and regional eco-environmental change.



Liangpei Zhang (M'06–SM'08) received the B.S. degree in physics from Hunan Normal University, Changsha, China, in 1982, the M.S. degree in optics from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China, in 1988, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998.

He is currently the Head of the Remote Sensing Division, State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University. He is also a Chang-Jiang Scholar Chair Professor appointed by the Ministry of Education of China. He is currently a Principal Scientist for the China State Key Basic Research Project from 2011 to 2016 appointed by the Ministry of National Science and Technology of China to lead the Remote Sensing Program in China. He has authored over 500 research papers and five books. He holds 15 patents. He edits several conference proceedings, issues, and geoinformatics symposiums. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

Dr. Zhang is a Fellow of the Institution of Engineering and Technology and an Executive Member (Board of Governor) of the China National Committee of International Geosphere–Biosphere Programme and the China Society of Image and Graphics. He was a recipient of the 2010 Best Paper Boeing Award and the 2013 Best Paper ERDAS Award from the American Society of Photogrammetry and Remote Sensing. He received the best reviewer awards from the IEEE Geoscience and Remote Sensing Society (GRSS) for his service to the IEEE JOURNAL OF SELECTED TOPICS IN EARTH OBSERVATIONS AND APPLIED REMOTE SENSING (JSTARS) in 2012 and the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS in 2014. His research teams received the top three prizes of the IEEE GRSS 2014 Data Fusion Contest, and his students have been selected as the winners or finalists of the IEEE International Geoscience and Remote Sensing Symposium Student Paper Contest in recent years. He was the General Chair of the 4th IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing. He regularly serves as a Co-Chair of the Series SPIE Conference on Multispectral Image Processing and Pattern Recognition, the Conference on Asia Remote Sensing, and many other conferences. He is the Founding Chair of the IEEE GRSS Wuhan Chapter. He was a Guest Editor of JSTARS. He also serves as an Associate Editor of the *International Journal of Ambient Computing and Intelligence*, the *International Journal of Image and Graphics*, the *International Journal of Digital Multimedia Broadcasting*, the *Journal of Geo-Spatial Information Science*, and the *Journal of Remote Sensing*, and the Guest Editor of the *Journal of Applied Remote Sensing* and the *Journal of Sensors*. He is currently serving as an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.