



# An Improved Flexible Spatiotemporal DAta Fusion (IFSDAF) method for producing high spatiotemporal resolution normalized difference vegetation index time series

Meng Liu<sup>a,e</sup>, Wei Yang<sup>b</sup>, Xiaolin Zhu<sup>c</sup>, Jin Chen<sup>a,\*</sup>, Xuehong Chen<sup>a</sup>, Linqing Yang<sup>d,e</sup>, Eileen H. Helmer<sup>f</sup>

<sup>a</sup> State Key Laboratory of Earth Surface Processes and Resource Ecology, Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

<sup>b</sup> Center for Environmental Remote Sensing, Chiba University, Chiba 263-8522, Japan

<sup>c</sup> Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong

<sup>d</sup> State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

<sup>e</sup> Department of Ecosystem Science and Management, Texas A&M University, College Station, TX 77843, USA

<sup>f</sup> International Institute of Tropical Forestry, USDA Forest Service, Río Piedras, PR, 00926, USA

## ARTICLE INFO

### Keywords:

Normalized difference vegetation index (NDVI)  
Spatiotemporal data fusion  
High spatial and temporal resolution  
Constrained least squares (CLS) method  
Weighted integration  
Sentinel data

## ABSTRACT

The Normalized Difference Vegetation Index (NDVI) is one of the most commonly used vegetation indices for monitoring ecosystem dynamics and modeling biosphere processes. However, global NDVI products are usually provided with relatively coarse spatial resolutions that lack important spatial details. Producing NDVI time-series data with high spatiotemporal resolution is indispensable for monitoring land surfaces and ecosystem changes, especially in spatiotemporally heterogeneous areas. The Improved Flexible Spatiotemporal DAta Fusion (IFSDAF) method was developed in this study to fill this need. In accord with the distinctive characteristics of NDVIs with large data variance and high spatial autocorrelation compared with raw reflectance bands, the IFSDAF method first produces a time-dependent increment with linear unmixing and a space-dependent increment via thin plate spline interpolation. It then makes a final prediction by optimal integration of these two increments with the constrained least squares method. Moreover, the IFSDAF was developed with the capacity to use all available finer-scaled images, including those partly contaminated by clouds. NDVI images with coarse spatial resolution (MODIS) and fine spatial resolution (Landsat and Sentinel) in areas with great spatial heterogeneity and significant land cover changes were used to test the performance of the IFSDAF method. The root mean square error and relative root mean square error of predicted relative to observed results were 0.0884 and 22.12%, respectively, in heterogeneous areas, and 0.0546 and 25.77%, respectively, in areas of land-cover change. These promising results demonstrated the strength and robustness of the IFSDAF method in providing reliable NDVI datasets with high spatial and temporal resolution to support research on land surface processes. The efficiency of the proposed IFSDAF method can be greatly improved by using only the space-dependent increment. This simplification will make IFSDAF a feasible method for monitoring global vegetation.

## 1. Introduction

The Normalized Difference Vegetation Index (NDVI) enhances the absorptive and reflective features of vegetation and provides a way of estimating canopy greenness and vigor (Rouse et al., 1974; Huete et al., 2002). Accordingly, NDVI time-series data derived from spaceborne sensors have been widely used in monitoring ecosystem dynamics and modeling biosphere processes to help understand the responses of

ecosystems to climate change (Pettorelli et al., 2005). The coarse spatial resolutions of the available NDVI time-series products (e.g., GIMMS, MODIS, and SPOT VGT), which range from 250 m to 8 km, are the most significant constraints on these applications and prevent them from capturing the spatial details necessary for monitoring land surface and ecosystem changes, especially in geographically heterogeneous areas (Gao et al., 2006; Rao et al., 2015). The requirement for NDVI time-series data with both high spatial and high temporal resolution in such

\* Corresponding author.

E-mail address: [chenjin@bnu.edu.cn](mailto:chenjin@bnu.edu.cn) (J. Chen).

applications has created a need for the development of spatiotemporal fusion methods that blend high-frequency but low-spatial-resolution images (e.g., MODIS images, hereinafter referred to as coarse images) with high-spatial-resolution but low-frequency images (e.g., Landsat images, hereinafter referred to as fine images) (Zhu et al., 2018). Because of the recent emergence of constellations of CubeSats and new satellite systems (e.g., Sentinel-2 data with a five-day or better revisit cycle and 10-m spatial resolution), the latest satellite images no longer require a trade-off between spatial and temporal resolution. However, spatiotemporal fusion is still necessary for long-term studies that involve historical satellite images collected before 2015.

When using spatiotemporal fusion technology to produce NDVI data with high spatiotemporal resolution, users need to make two decisions: (I) selection of an appropriate blending strategy, either Blend-then-Index (BI) or Index-then-Blend (IB), and (II) selection of a suitable and accurate spatiotemporal fusion method. Recent studies relevant to the first decision (Chen et al., 2018; Jarihani et al., 2014; Tian et al., 2013) have demonstrated that the IB strategy consistently yields comparable or better results than the BI strategy, mainly because the IB method has the following advantages over the BI method: (i) there is less error propagation in the IB blending process; (ii) the IB process is less computationally expensive; and (iii) with advanced filters, it is easier to clean the noise (e.g., cloud effects) on the NDVI than on the raw reflectance bands (e.g., Chen et al., 2004). Consequently, the IB strategy is generally recommended and has become the dominant blending strategy for producing fused NDVI products.

A number of spatiotemporal fusion methods relevant to the second decision have been proposed and validated in recent years (Zhu et al., 2018). These methods require at least one pair of cloud-free fine and coarse NDVI images at a base date and a series of coarse NDVI images at the prediction dates as input. However, a consensus regarding the most suitable method for producing high spatiotemporal resolution NDVI data has not been reached. To enhance features, NDVI uses a band combination that enlarges the contrast between vegetated and non-vegetated pixels and therefore displays larger spatial and temporal differences than the raw reflectance bands in most satellite images. To maintain that greenness sensitivity, a suitable spatiotemporal method for fusing NDVI images should: (i) accurately predict greenness, even in areas with large spatial and temporal variance; (ii) require only one pair of clear fine and coarse NDVI images at a base date to ensure applicability in areas with frequent cloud contamination; and (iii) be able to handle land cover change, such as urbanization, deforestation/reforestation, wildfires, floods and other land cover transitions. Among the existing spatiotemporal fusion methods, only the Flexible Spatiotemporal DAta Fusion method (FSDAF) (Zhu et al., 2016) meets these criteria and can be considered a potential fusion method. Other existing methods fail to satisfy at least one criterion, especially the third. For example, none of the following methods can handle land cover changes that occur between the base date and prediction date: the spatial and temporal adaptive reflectance fusion model (STARFM, Gao et al., 2006), the enhanced STARFM (ESTARFM, Zhu et al., 2010), the spatial and temporal adaptive vegetation index fusion model (STAVFM, Meng et al., 2013), the unmixing-based spatiotemporal reflectance fusion model (U-STFM, Huang and Zhang, 2014), the NDVI linear mixing growth model (NDVI-LMGM, Rao et al., 2015), and the spatial and temporal reflectance unmixing model (STRUM, Gevaert and Garcia-Haro, 2015). Learning-based methods, such as the Sparse-representation-based spatiotemporal reflectance fusion model (SPSTFM, Huang and Song, 2012; Song and Huang, 2013), the error-bound-regularized semi-coupled dictionary learning model (EBSCDM, Wu et al., 2015), and the extreme learning machine-based fusion method of Liu et al. (2016) are better at capturing land cover change. However, their learning step is time consuming, and their accuracy decreases where spatial heterogeneity is high and scale differences between coarse and fine images are large (Zhu et al., 2016).

FSDAF is based on a spectral unmixing analysis and uses thin plate

spline (TPS) interpolation to capture land cover change, as long as the change is detectable in coarse images (Zhu et al., 2016). FSDAF requires the same input data as two widely used spatiotemporal fusion methods, including STAFRM (Gao et al., 2006) and the unmixing-based data fusion (UBDF) algorithm (Zurita-Milla et al., 2008), but its predictions are more accurate, especially in the NIR band of heterogeneous landscapes (Tables 3 and 4 in Zhu et al., 2016). Like the NDVI, the NIR band has larger spatial and temporal variances than the red band. NIR reflectance generally varies more among land cover types than the red band, and it changes more with vegetation growth cycles. Moreover, FSDAF can capture both gradual and abrupt land cover changes, something that is difficult for current spatiotemporal fusion methods to do. The many advantages of FSDAF suggest that it may be an appropriate method for producing high spatiotemporal resolution NDVI data. However, the FSDAF method can be improved. The FSDAF method relies entirely on the results of TPS interpolation to distribute residuals ( $\varepsilon$ ) between predicted and true values on the assumption that errors depend mainly on the landscape homogeneity. Such an assumption is very empirical and has no theoretical basis. It may not be an optimal way to distribute residuals for different scenarios. Furthermore, in practice, many available finer-scale images (hereafter, fine images) are partly contaminated by clouds. Clear pixels in these partly contaminated fine images can provide useful information on temporal changes, as demonstrated in the STAIR method for daily surface reflectance fusion (Luo et al., 2018). Using cloud-free fine images together with partly contaminated fine images will therefore facilitate spatiotemporal NDVI fusion and expand its applicability to cloudy regions. Unfortunately, the FSDAF method does not have a similar capability and is therefore not applicable in cloudy regions.

To address these limitations, we propose the Improved Flexible Spatiotemporal DAta Fusion (IFSDAF) method for producing high spatiotemporal resolution NDVI time series. The IFSDAF incorporates constrained least squares (CLS) theory into the FSDAF method. It uses CLS to combine a temporal prediction, derived from an unmixing procedure, and a spatial prediction, derived from TPS interpolation. This combination ensures that the final prediction is obtained from an optimal integration of temporal and spatial predictions. Moreover, the IFSDAF was developed with the capacity to employ all available fine images, including partly contaminated ones (e.g., cloud coverage < 70%). To validate the effectiveness of the proposed method, we compared the performances of the IFSDAF and three popular NDVI fusion methods (i.e., NDVI-LMGM, STARFM, and FSDAF) using the IB strategy in several experimental areas, including a site with a heterogeneous landscape, a site with abrupt land cover changes, and a site where satellite images contained numerous clouds.

## 2. Methodology

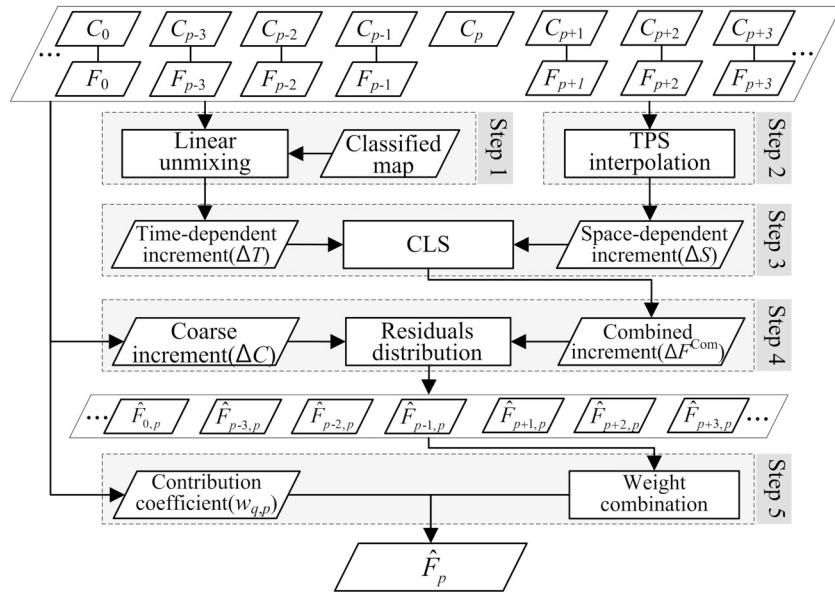
Although the principles of existing spatiotemporal fusion methods vary greatly, the main idea can be summarized by

$$F_p = F_0 + \Delta F + \varepsilon. \quad (1)$$

The fine increment of NDVI ( $\Delta F$ ) between the prediction date ( $t_p$ ) and the base date ( $t_0$ ) is first estimated, and then the predicted fine NDVI values ( $F_p$ ) at time  $t_p$  are equated to the sum of the base fine NDVI value ( $F_0$ ), the increment ( $\Delta F$ ), and the residuals  $\varepsilon$ .

Given that  $F_0$  is known, IFSDAF also follows Eq. (9), but it estimates the increment in two ways: (i) time-dependent increments using unmixing analysis and (ii) space-dependent increments using the Thin Plate Spline (TPS) interpolation method. The IFSDAF then combines the two increments to obtain a final  $\Delta F$  by the constrained least squares (CLS) method. The CLS method adopted here improved upon the original FSDAF by adaptively combining the two increments.

Fig. 1 shows the flowchart of the proposed IFSDAF. The input data for IFSDAF include coarse NDVI time-series images and all available fine NDVI images within the same time interval. In these images, coarse



**Fig. 1.** Flowchart of the Improved Flexible Spatiotemporal DAta Fusion method (IFSDAF).

NDVI and fine NDVI images acquired on the same date were designated as one pair. The pair with the least cloud contamination was selected as the base pair of images ( $C_0$  and  $F_0$ ), and its acquisition date was the base date  $t_0$ . The dates of other pairs are denoted as ...,  $p - 3, p - 2, p - 1, p + 1, p + 2, p + 3, \dots$ . The coarse and fine NDVI images of these pairs were denoted as ...,  $C_{p-3}, C_{p-2}, C_{p-1}, C_p, C_{p+1}, C_{p+2}, C_{p+3}, \dots$  and ...,  $F_{p-3}, F_{p-2}, F_{p-1}, F_p, F_{p+1}, F_{p+2}, F_{p+3}, \dots$ , respectively. The task of IFSDAF was to predict fine NDVI images on any date for which a coarse NDVI image was available, e.g., the date  $t_p$ . In IFSDAF, the input fine NDVI images were not required to be cloud-free, with the exception of  $F_0$ . As is the case with other spatiotemporal fusion methods, all the coarse and fine NDVI images were georegistered and cropped because they had to be the same size. In addition, coarse NDVI time-series were smoothed by an algorithm based on the Savitzky-Golay filter (Cao et al., 2018; Chen et al., 2004), which was designed to reconstruct high-quality NDVI time-series data by retaining clear-sky values and interpolating cloudy values. Cloud pixels in partly cloud-contaminated fine NDVI images were masked by the Fmask algorithm (Zhu and Woodcock, 2012). A land cover classification map at a fine resolution, which could be derived from either existing land cover products such as Globeland30 (Chen et al., 2015) or the classification result of the input clear fine images, was needed to provide fractional cover for the unmixing process. The output of IFSDAF was synthetic fine NDVI images ( $\hat{F}_p$ ) on the prediction date  $t_p$  ( $p = 1, 2, 3, \dots$ ). A more detailed description for each implementation step of IFSDAF is given below, and a list of notations and explanation is given in the Appendix.

### 2.1. Generation of time-dependent increments with the unmixing method

According to linear spectral mixing theory, the time-dependent NDVI change (increment) of a coarse pixel can be considered to be a linear combination of NDVI increments of all fine pixels within that coarse pixel during a short period of time (Rao et al., 2015). A linear mixture model was therefore used to unmix the increment of coarse pixels from the base date  $t_0$  to the prediction date  $t_p$  on the assumption that fine pixels belonging to the same class of land cover had a similar increment within the local region (Busetto et al., 2008; Rao et al., 2015). Neighboring coarse pixels within a moving window centered on a coarse pixel ( $x, y$ ) were used to establish a linear equation system, as shown by

$$\begin{bmatrix} \Delta C(1,1) \\ \vdots \\ \Delta C(x,y) \\ \vdots \\ \Delta C(n,n) \end{bmatrix} = \begin{bmatrix} f_1(1,1) & f_2(1,1) & \cdots & f_l(1,1) \\ \vdots & \vdots & & \vdots \\ f_1(x,y) & f_2(x,y) & \cdots & f_l(x,y) \\ \vdots & \vdots & & \vdots \\ f_1(n,n) & f_2(n,n) & \cdots & f_l(n,n) \end{bmatrix} \begin{bmatrix} \Delta F_1 \\ \vdots \\ \Delta F_c \\ \vdots \\ \Delta F_l \end{bmatrix}, \quad (2)$$

$$\text{with s. t. } \min(\Delta C_{\text{window}}) - \text{std}(\Delta C_{\text{window}}) \leq \Delta F_c \leq \max(\Delta C_{\text{window}}) + \text{std}(\Delta C_{\text{window}})$$

where  $n$  is the number of coarse pixels and  $l$  is the number of land cover classes within the moving window.  $\Delta C(x, y)$  is the NDVI increment of the coarse pixel ( $x, y$ ) that can be obtained directly from coarse NDVI time series images.  $\Delta F_c$  is the fine NDVI increment of class  $c$  within the window.  $f_l(x, y)$  is the fraction of class  $l$  within the coarse pixel ( $x, y$ ), which can be obtained from the land cover map at a fine resolution.  $\Delta C_{\text{window}}$  is the set of all coarse NDVI increments in the window. The variables  $\min(\Delta C_{\text{window}})$ ,  $\max(\Delta C_{\text{window}})$ , and  $\text{std}(\Delta C_{\text{window}})$  are the minimum value, maximum value, and standard deviation of  $\Delta C_{\text{window}}$ , respectively. A moving window sized at  $7 \times 7$  coarse pixels is recommended because the number of coarse pixels in the window, 49, is commonly much larger than the number of land cover classes. This choice of window size ensures that the abovementioned over-determined linear equations are minimally influenced by collinearity and land cover changes. By solving the linear equations, the time-dependent NDVI increment of each class ( $\Delta F_c$ ) in the moving window can be acquired. The fine time-dependent increment  $\Delta T(x_j, y_j)$ , where  $(x_j, y_j)$  devotes the  $j$ th fine pixel in the coarse pixel ( $x, y$ ), is then defined as follows:

$$\Delta T(x_j, y_j) = \Delta F_c \text{ if fine pixel } (x_j, y_j) \text{ belongs to class } c. \quad (3)$$

The fine-resolution land cover map used to compute the class fractions can be an available land cover product or classification of a cloud-free fine image. In practice, to make the fusion process automatic, existing fusion methods often use unsupervised classifiers (e.g., K-means and ISODATA) to obtain spectral classes rather than real land cover classes (Rao et al., 2015; Zhu et al., 2016). Users need to set the number of classes in unsupervised classification. Results of previous studies have indicated that 3 to 6 classes can produce satisfactory results in most situations (Rao et al., 2015; Zhu et al., 2016). Assessment of the accuracy of the classification map is not included in the fusion process because (1) aggregation of the fine-scale class and the coarse-scale fraction will average out some errors in classification and may therefore

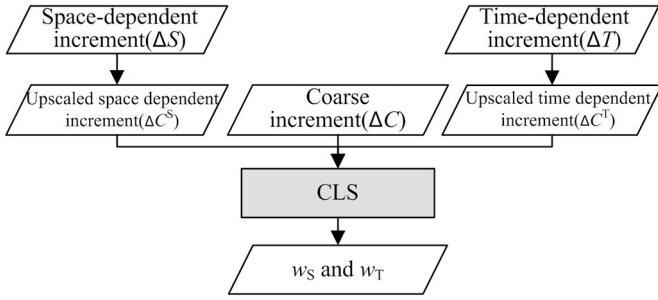


Fig. 2. Illustration of weighted calibration based on the CLS method.

not cause a large problem in solving Eq. (2), (2) temporal change assigned to a pixel with incorrect class labels using Eq. (3) will be compensated by the space-dependent increment introduced in Section 2.2, and (3) selection of reference samples for accuracy assessment will introduce more human-computer interactions. Although the proposed method is not sensitive to classification accuracy, inclusion of more accurate and robust classification methods in the FSDAF could further improve its performance.

## 2.2. Generation of space-dependent increments by TPS interpolation

A coarse NDVI image at time  $t_p$  contains signals of land cover changes when changes are significant enough to appear in coarse pixels. Spatial interpolation of coarse NDVI to fine resolution will therefore retain useful information on land cover changes. Accordingly, coarse-spatial-resolution NDVI images at times  $t_p$  and  $t_0$  are interpolated to fine spatial resolution with the Thin Plate Spline (TPS) interpolation method (Chen et al., 2014; Zhu et al., 2016). TPS is a spatial interpolation technique for point data based on spatial dependence (Dubrule, 1984) and has been employed to obtain interpolation results because of its high accuracy. Another increment from the difference between interpolation results at times  $t_p$  and  $t_0$  can then be acquired. Because this increment uses only the spatial dependence among coarse pixels, it can be referred to as the space-dependent increment  $\Delta S(x_j, y_j)$ , as shown in the following equation, where  $F_p^{\text{TPS}}(x_j, y_j)$  and  $F_0^{\text{TPS}}(x_j, y_j)$  are TPS-interpolated values at times  $t_p$  and  $t_0$ , respectively, and  $(x_j, y_j)$  is the  $j$ th fine pixel within the coarse pixel  $(x, y)$ :

$$\Delta S(x_j, y_j) = F_p^{\text{TPS}}(x_j, y_j) - F_0^{\text{TPS}}(x_j, y_j). \quad (4)$$

A space-dependent increment has two advantages over a time-dependent increment. First, a coarse NDVI image on date  $t_p$  contains signals of land cover changes if the changes are significant enough to be recorded. Such land cover change information can be directly captured at a fine resolution by TPS interpolation. Second, because a space-dependent increment is independent of the classification map and unmixing procedure, it has the potential to justify errors in the time-dependent increment that resulted from classification or unmixing. In this study, TPS was used to estimate the space-dependent increment rather than to estimate the NDVI value at time  $t_p$ . This strategy has been used in FSDAF, because the space-dependent increment reveals the changes of NDVI directly. Zhang et al. (2015) have also suggested that using an increment yields higher accuracy than predicting the value directly at time  $t_p$ . The use of this space-dependent increment is further discussed in Section 5.

## 2.3. Combination of the two increments by CLS

The time-dependent increment and space-dependent increment can be regarded as two independent predictions by two different models. The former uses information of the temporal changes of NDVI, and the latter uses mainly the spatial dependence of the NDVI. Their prediction accuracies should vary under different scenarios and spatial

dependencies. It is therefore likely that a reasonable combination of the two increments can improve the performance and robustness of the fusion method.

The simplest and most effective way of combining the time-dependent increment ( $\Delta T$ ) and space-dependent increment ( $\Delta S$ ) would be summing them with reasonable weights. Moreover, an ideal combination should be as close to the true fine NDVI increment ( $\Delta F$ ) as possible. Thus, an objective function of weighted increments can be written as follows:

$$(\hat{w}_S, \hat{w}_T) = \arg \min_{(w_S, w_T) \in (0, 1)} \sum_k (w_S \Delta S_k + w_T \Delta T_k - \Delta F_k)^2, \quad (5)$$

where  $\Delta S_k$ ,  $\Delta T_k$ , and  $\Delta F_k$  are the space-dependent increment, the time-dependent increment, and the true increment of the  $k$ th fine pixel, respectively. The  $w_S$  and  $w_T$  are the weights of the space-dependent increment and the time-dependent increment, respectively. Eq. (5) can be solved by the CLS method. The values of  $w_S$  and  $w_T$  must be non-negative and sum to 1.0.

However, because the fine NDVI values at time  $t_p$  are unknown, it is impossible to obtain the true fine increment ( $\Delta F$ ). Fortunately, a real NDVI increment of a coarse pixel ( $\Delta C$ ) from  $t_0$  to  $t_p$  is available because coarse observations are available on multiple dates. Therefore, both the time-dependent increment and the space-dependent increment were scaled up to the resolution of a coarse pixel ( $\Delta C^T$  and  $\Delta C^S$ ), as shown in Fig. 2. Then,  $w_S$  and  $w_T$  in Eq. (5) could be obtained by solving

$$(\hat{w}_S, \hat{w}_T) = \arg \min_{(w_S, w_T) \in (0, 1)} \sum_k (w_S \Delta C_k^S + w_T \Delta C_k^T - \Delta C_k)^2, \quad (6)$$

where  $\Delta C_k^S$ ,  $\Delta C_k^T$  and  $\Delta C_k$  are the scaled-up space-dependent increment, scaled-up time-dependent increment, and true increment of the  $k$ th coarse pixel, respectively. Here, the average value of all fine NDVI pixels within a coarse pixel is used to produce a scaled-up space-dependent increment ( $\Delta C^S$ ) and a scaled-up time-dependent increment ( $\Delta C^T$ ), and  $\Delta C_k$  is the difference between the coarse NDVI values at times  $t_p$  and  $t_0$ . Because the weights  $w_S$  and  $w_T$  are spatially dependent, Eq. (6) is solved in a  $7 \times 7$  pixel moving window at a coarse resolution corresponding to the window size of the unmixing process. The final fine increment can then be calculated with the estimated  $w_S$  and  $w_T$  as follows:

$$\Delta F^{\text{Com}}(x_j, y_j) = w_S \times \Delta S(x_j, y_j) + w_T \times \Delta T(x_j, y_j), \quad (7)$$

where  $\Delta F^{\text{Com}}(x_j, y_j)$  is the combined increment of fine pixel  $(x_j, y_j)$ . The  $w_S$  and  $w_T$  are assumed to be scale-invariant. The rationale for this assumption is discussed in Section 5.

## 2.4. Distribution of residuals

After CLS optimization, the combined increment could capture most of the fine NDVI increment. Although the residuals were minimized, they were not zero. The residuals are expressed mathematically by

$$R(x, y) = \Delta C(x, y) - \frac{1}{m} \sum_{j=1}^m \Delta F^{\text{Com}}(x_j, y_j), \quad (8)$$

where  $R(x, y)$  is the residual within a coarse pixel  $(x, y)$  and  $m$  is the number of fine pixels within the coarse pixel. To further improve the accuracy of the combined increment, it was necessary to allocate this residual to each fine pixel  $(x_j, y_j)$  within the coarse pixel  $(x, y)$ . Because the residuals were minimized by the CLS method, they could be distributed equally to all pixels (Chen et al., 2014), as indicated by

$$\hat{F}_{0,p}(x_j, y_j) = F_0(x_j, y_j) + \Delta F^{\text{Com}}(x_j, y_j) + R(x, y), \quad (9)$$

where  $F_0(x_j, y_j)$  is the fine NDVI of pixel  $(x_j, y_j)$  on date  $t_0$  and  $\hat{F}_{0,p}(x_j, y_j)$  is the predicted fine NDVI on date  $t_p$ . After the distribution of residuals, a smoothing process based on the similarity of pixels (Zhu et al., 2016) was applied to remove block effects in the fused image.

## 2.5. Combination of multi-time predictions

Eq. (9) makes it possible to predict a fine NDVI,  $\hat{F}_{0,p}$ , on date  $t_p$  based on the fine NDVI on date  $t_0$ . In the same way, there will be several NDVI predictions, such as  $\hat{F}_{p-3,p}$ ,  $\hat{F}_{p-2,p}$ ,  $\hat{F}_{p-1,p}$ ,  $\hat{F}_{p+1,p}$ ,  $\hat{F}_{p+2,p}$ ,  $\hat{F}_{p+3,p}$ , ... for date  $t_p$  based on clear observations at times  $p + i$  ( $i = \dots, -3, -2, -1, 1, 2, 3, \dots$ ) in other partly cloudy fine NDVI images. Recognition of a pixel as either clear or cloudy can be performed based on the Fmask algorithm (Zhu and Woodcock, 2012). Generally, predictions with a base date too far from  $t_p$  were excluded on the assumption that the base NDVI images were weakly related to the NDVI image on date  $t_p$ . Operationally, the maximum interval between the base date and the prediction date was set to two months. The NDVI difference of coarse pixels between the base date and the prediction date was then used to calculate the contribution of each prediction, as shown in the following equation:

$$w_{q,p}(x,y) = \frac{1}{\sum_{i=1}^9 |C_q^i(x,y) - C_p^i(x,y)|}, \quad (10)$$

where  $C_q^i(x,y)$  and  $C_p^i(x,y)$  are the coarse NDVI values of the  $i$ th pixel on base date  $q$  and the prediction date  $t_p$  in the  $3 \times 3$  moving window centered on coarse pixel  $(x, y)$ . The  $w_{q,p}(x, y)$  is the contribution coefficient of the predicted fine NDVI value  $\hat{F}_{q,p}(x_j, y_j)$  within the center coarse pixel  $(x, y)$ . Based on this contribution coefficient, the combined prediction of a fine pixel  $(x_j, y_j)$  on date  $t_p$  is given by

$$\hat{F}_p(x_j, y_j) = \sum_q [w_{q,p}(x, y) \times \hat{F}_{q,p}(x_j, y_j)] / \sum_q w_{q,p}(x, y). \quad (11)$$

If  $C_q^i(x,y)$  equals  $C_p^i(x,y)$ ,  $\hat{F}_p(x_j, y_j)$  will be equated to  $\hat{F}_{q,p}(x_j, y_j)$  because  $w_{q,p}(x, y)$  is infinite in this case. Finally, for each prediction date in the time series, a final prediction via Eq. (11) can be obtained by using the routine described in Sections 2.1–2.5.

To assess the performance of the new method, four accuracy indices, the root mean square error (RMSE), relative RMSE (rRMSE, defined as RMSE divided by averaged observation value and multiplied by 100%), Pearson correlation coefficient ( $r$ ) and average difference (AD) were used. These indices have been widely used to assess the accuracy of fused images in previous studies (e.g. Gao et al., 2006; Rao et al., 2015; Zhu et al., 2016). Moreover, to test the performance improvement compared with FSDAF, a  $t$ -test was employed to check whether the prediction accuracy between IFSDAF and FSDAF is significantly different.

## 3. Data

### 3.1. Data for experiments using single cloud-free fine images

We used Landsat images without clouded pixels to evaluate the performance of the proposed IFSDAF model at two sites with different land-cover characteristics (Table 1). Because the existing spatio-temporal fusion methods have been shown to generally perform well in homogeneous areas (Zhu et al., 2018), this study tested the performance of the new method only in relatively complex cases (i.e., a heterogeneous site or a site with significant changes of land cover). The Landsat images covering the two sites have also been used by Emelyanova et al. (2013), and they were used to test the NDVI-LMGM and FSDAF algorithms by Rao et al. (2015) and Zhu et al. (2016).

**Table 1**

Summary of test sites and data sources.

Experiments	Location	MODIS	Landsat	Sentinel-2A	Prediction mode
Test 1	Coleambally irrigated area, Australia	Yes	Yes	No	Single date
Test 2	Gwydir area, Australia	Yes	Yes	No	Single date
Test 3	Shennongjia, central China	Yes	Yes	No	Time series
Test 4	Southeast Asia	Yes	No	Yes	Time series

This first site was located in the Coleambally irrigated area, Australia ( $34^{\circ}54'S, 145^{\circ}57'E$ ), which is characterized by a very heterogeneous landscape with many small patches of farmland and rapid phenological changes (Fig. 3). Two Landsat ETM+ images ( $800 \times 800$  pixels) acquired on 25 November 2001 ( $t_0$ ) and 12 January 2002 ( $t_p$ ) during the growing season were scaled up by a factor of 8 to synthesize MODIS images. In this test, the synthesized MODIS images instead of the real MODIS images were used, because synthesized MODIS images can exclude co-registration errors (Gevaert and Garcia-Haro, 2015; Wang and Atkinson, 2018; Zhu et al., 2016). This exclusion ensured a fair comparison of different algorithms, because different algorithms have different sensitivities to co-registration errors. The NDVI data were then derived from the corresponding reflectance images. A land cover classification map was then obtained by the Iterative Self-Organizing Data Analysis Technique (ISODATA) method based on the Landsat image acquired on 25 November 2001 ( $t_0$ ).

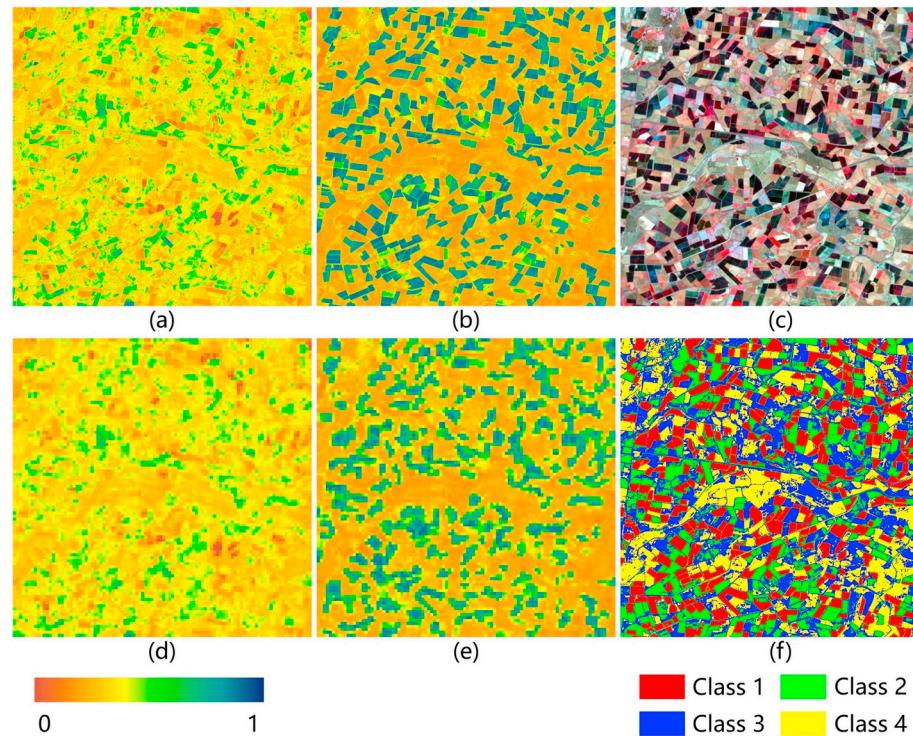
The second site was located in the Gwydir area, Australia ( $29^{\circ}07'S, 149^{\circ}04'E$ ), where there was flooding in December 2004. Two Landsat TM images ( $800 \times 800$  pixels) on 26 November 2004 ( $t_0$ ) and 12 December 2004 ( $t_p$ ) were used at this site (Fig. 4). Abrupt land cover changes are apparent in these two images because of the flood (Emelyanova et al., 2013). These two Landsat images were also scaled up by a factor of 8 to synthesize MODIS images. All the NDVI data were derived from the original images. The ISODATA method was used to obtain a land cover classification map based on the Landsat image on 26 November 2004 ( $t_0$ ).

For these two sites, NDVI-LMGM (Rao et al., 2015), STARFM (Gao et al., 2006), and FSDAF (Zhu et al., 2016) were also applied to the same datasets for comparison.

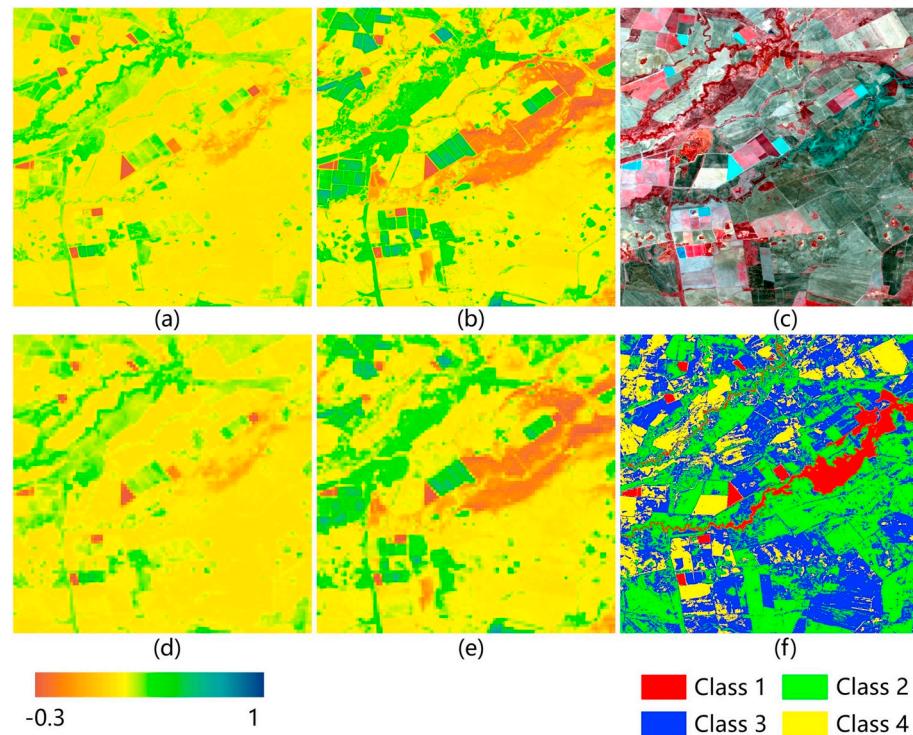
### 3.2. Data for experiments using multiple cloudy fine images

Experiments using multiple fine images contaminated by clouds were implemented to assess the performance of the proposed IFSDAF method for predicting the NDVI time series when the input fine images were partly contaminated by clouds. To test the applicability of IFSDAF for fusing images from diverse sensors, we fused both Landsat and Sentinel-2 images with MODIS at two cloudy sites (Table 1). The first site is mainly covered by natural vegetation and the change could be slow, while the second site is covered by many crops thus the change is fast.

The first site was the Shennongjia Forestry District ( $109^{\circ}59'–110^{\circ}58'E, 31^{\circ}15'–31^{\circ}57'N$ ), which is located in the western part of Hubei Province, central China (Fig. 5). This area is characterized by a subtropical monsoon climate and its elevation ranges from 398 to 3105 m. The distribution of vegetation in this area is very heterogeneous and includes evergreen broadleaf forests, deciduous broadleaf forests, and evergreen coniferous forests. There are also farmlands and artificial surfaces in this area (Wang et al., 2018; Zhao et al., 2005). Considering that the area is mainly covered by subtropical forest with relatively slow ecosystem dynamics, the 16-day composite MODIS NDVI dataset (250-m-resolution, MOD13Q1 in 2015), instead of daily data, was used (<https://ladsweb.nascom.nasa.gov/search/>), because it can adequately capture ecosystem dynamics and 16-day composite data contain high-quality VI values selected from the daily observations. Landsat 8 level 2A surface reflectance products in 2015 and their cloud masks were



**Fig. 3.** Test data from the heterogeneous site in the Coleambally irrigation area: Landsat NDVI on (a) 25 November 2001 and (b) 12 January 2002, (c) false-color-composite Landsat image on 25 November 2001, (d) MODIS-scale NDVI on 25 November 2001 and (e) 12 January 2002, and (f) land cover map on 25 November 2001 by ISODATA.

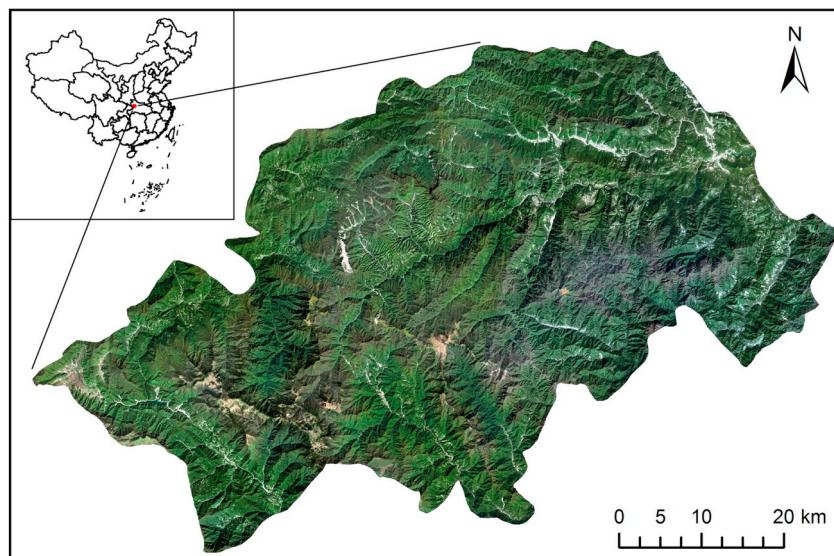


**Fig. 4.** Test data for a site that experienced land cover change in the Gwydir area: Landsat NDVI on (a) 26 November 2004 and (b) 12 December 2004, (c) false-color-composite Landsat image on 26 November 2004, MODIS-scale NDVI on (d) 26 November 2004 and (e) 12 December 2004, and (f) classification map on 26 November 2004 obtained with ISODATA.

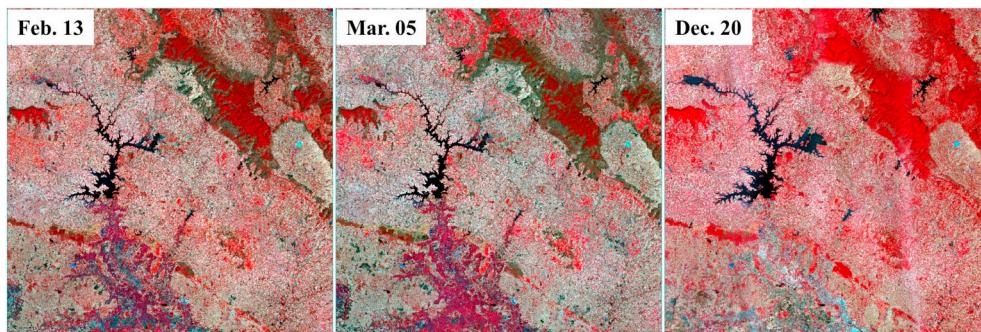
downloaded from the USGS (<https://espa.cr.usgs.gov/ordering/new/>). All Landsat images were co-registered to MODIS images. A mosaic of two adjacent Landsat-8 scenes covered the entire area of this site. When mosaicking two Landsat 8 images with close acquisition dates, pixels in the overlapped region had two NDVI values. The higher value was retained because a high NDVI is less likely to have been affected by poor atmospheric conditions (Van Leeuwen et al., 1999). Those Landsat images with clouds, shadows, and snow that covered > 70% of the image were discarded. Finally, one clear Landsat image ( $t_0$ , on 14

October 2015) and nine partly contaminated Landsat images (Fig. 1 in Supplementary data) were selected as the fine-resolution NDVI images for data fusion.

The second site was in Southeast Asia and was characterized by a complex landscape that included croplands, water, forests, and urban areas. The site was covered by one Sentinel-2A scene (size  $10,980 \times 10,980$  10-m-resolution pixels) with a tile number of T48QUD. We acquired Sentinel-2A satellite level 1C products from EarthExplorer (<https://earthexplorer.usgs.gov/>). Considering the



**Fig. 5.** Shennongjia Forestry District in Hubei Province, central China. The image is a true-color-composite Landsat 8 OLI image acquired on 14 October 2015.



**Fig. 6.** False-color-composites of clear Sentinel 2A images on 13 February, 5 March, and 20 November 2017 with tile number T48QUD.

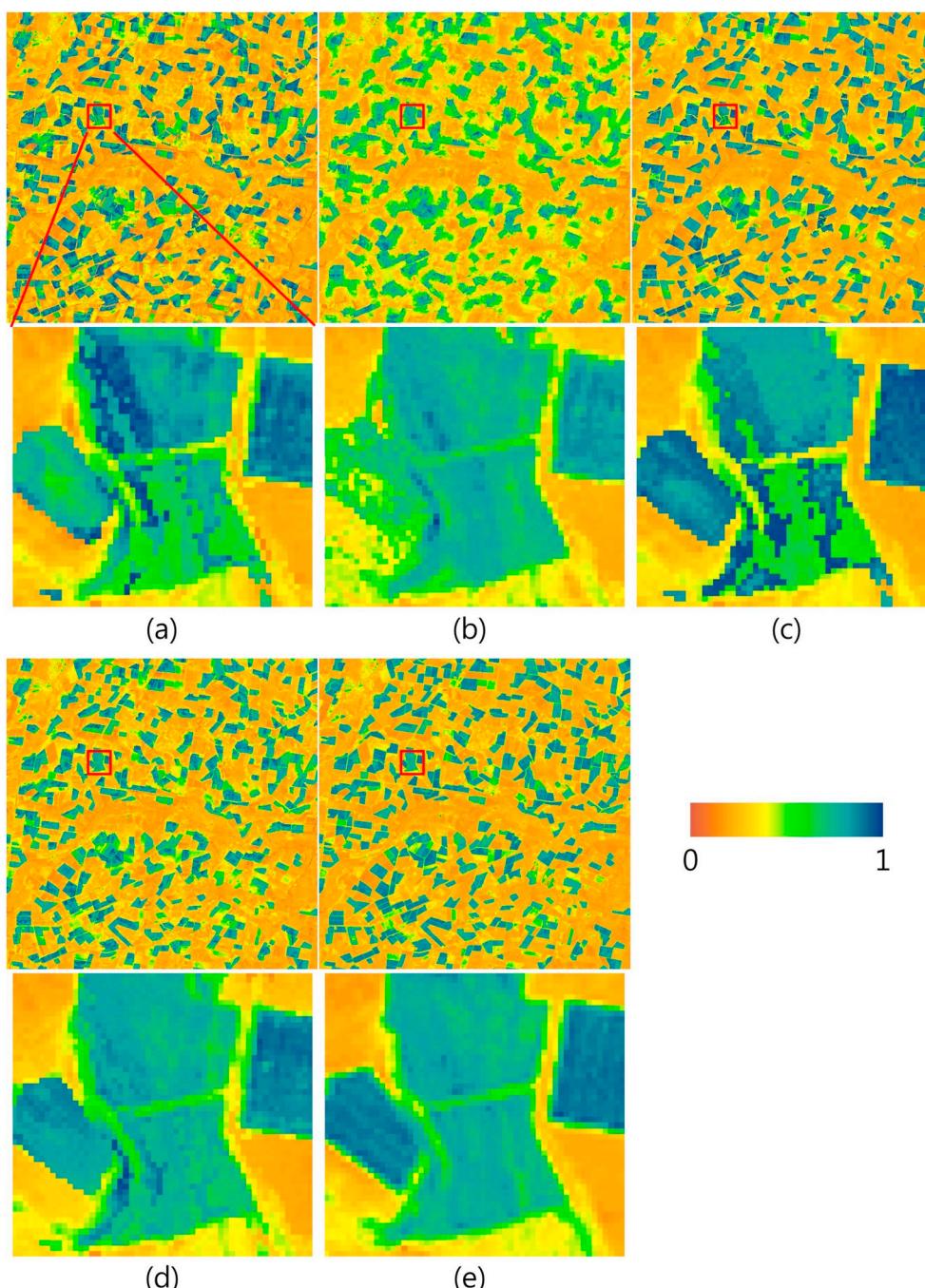
tropical climate with relatively fast ecosystem dynamics, the 8-day composite MODIS NDVI dataset (calculated from 8-day composite MODIS surface reflectance products (MOD09Q1)), instead of daily data, was used to capture the ecosystem dynamics and meanwhile alleviate cloud impacts. Both images were acquired in 2017. Atmospheric correction of Sentinel-2A images was done with a tool provided by the European Space Agency, Sen2Cor (<http://step.esa.int/main/third-party-plugins-2/sen2cor/>). Cloud masks of the sentinel-2A images were produced with Fmask software (<https://github.com/gersl/fmask>), and images with > 70% cloud cover were discarded. We obtained three clear Sentinel-2A images on 13 February, 5 March, and 20 December 2017 (Fig. 6). Finally, we also obtained 21 partly cloud-contaminated images (Fig. 2 in Supplementary data). Sentinel-2A NDVI was calculated based on band 4 and band 8, which have some overlap in wavelength and bandwidth with the Landsat and MODIS red and NIR bands. Although Sentinel-2A has another narrow NIR band (band 8a) which is more consistent with MODIS NIR band, it was not used in this study because of its coarser spatial resolution 20 m. In addition, simple linear regression was applied to normalize MODIS NDVI to Sentinel-2A NDVI to compensate for inconsistencies caused by differences in atmospheric correction between the two image sources. There is high consistency between Landsat data and MODIS data (Zhu et al., 2016), so simple linear regression is only used to normalize NDVI between Sentinel-2A and MODIS.

## 4. Results

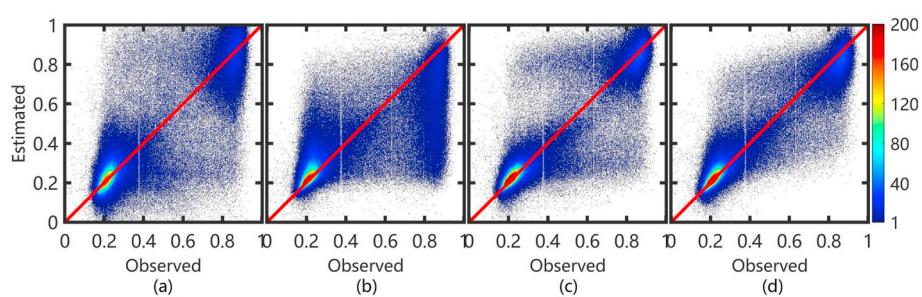
### 4.1. Fusion using a single cloud-free fine image

Fig. 7 provides a visual comparison of a Landsat NDVI image predicted by IFSDAF and the three existing methods with the observed Landsat NDVI on 12 January 2002 ( $t_p$ ) for the farmland site with great heterogeneity and rapid phenological changes. Compared with the other three methods, the image fused by IFSDAF (Fig. 7d) more closely resembled the actual NDVI image (Fig. 7e) (e.g., the zoomed-in sub-region). In contrast, the NDVI-LMGM (Fig. 7a) and FSDAF (Fig. 7c) methods produced large errors in some pixels that led to discontinuities in the fused images, and STARFM (Fig. 7b) led to an unsatisfactory blurring effect for small objects. Scatter plots (Fig. 8) and quantitative assessment also confirmed that the proposed method achieved the highest accuracy (Table 2). The IFSDAF had the lowest RMSE (0.084), lowest rRMSE (22.1%), and highest  $r$  (0.938) for the whole image. Furthermore, the fact that the AD ( $-0.0001$ ) of the newly proposed method was closer to zero than the AD of the other methods indicated that it was the least biased of the methods. In addition, the metrics of accuracy of the NDVI-LMGM (RMSE = 0.130 and rRMSE = 32.5%) and STARFM (RMSE = 0.165 and rRMSE = 41.2%) were much lower for the whole image than the corresponding metrics of the FSDAF (RMSE = 0.100 and rRMSE = 25.1%).

We also tested the hypothesis that the variance of the NDVI is normally larger than that of the raw reflectance bands. Fig. 9a shows frequency distributions of the Red and NIR bands and the corresponding NDVI from the Landsat image on 12 January 2002 ( $t_p$ ). The



**Fig. 7.** Landsat NDVI on 12 January 2002: predictions by (a) NDVI-LMGM, (b) STARFM, (c) FSDAF, (d) IFSDAF, and (e) the actual NDVI. For better visualization, all images were enlarged in the supplementary data.



**Fig. 8.** Scatter plots of estimated results compared with observed values of Landsat NDVI on 12 January 2002: (a) NDVI-LMGM, (b) STARFM, (c) FSDAF, and (d) IFSDAF.

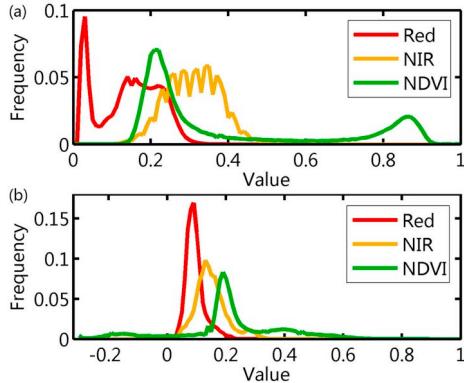
**Table 2**

Comparison of RMSE, rRMSE,  $r$ , and AD between predicted NDVI and observed NDVI for NDVI-LMGM, STARFM, FSDAF, and the IFSDAF method in the Coleambally irrigation area. The results of  $t$ -test between predictions accuracy of IFSDAF and FSDAF are marked in RMSE column under IFSDAF.

Method	NDVI-LMGM				STARFM				FSDAF				IFSDAF			
	RMSE	rRMSE	$r$	AD	RMSE	rRMSE	$r$	AD	RMSE	rRMSE	$r$	AD	RMSE	rRMSE	$r$	AD
Low NDVI	0.0916	38.85%	0.4476	<b>0.0148***</b>	0.1068	45.29%	0.5214	0.0564	0.0669	28.36%	0.5576	0.0160	<b>0.0664*</b>	<b>28.16%</b>	<b>0.6334</b>	0.0170
Medium NDVI	0.2415	45.11%	0.2917	−0.0476	0.1482	27.69%	0.2805	−0.0175	0.1962	36.64%	0.3740	−0.0205	<b>0.1473*</b>	<b>27.51%</b>	<b>0.4328</b>	−0.0060
High NDVI	0.1589	19.03%	0.3171	−0.0477	0.2130	25.52%	0.2983	−0.1656	0.1221	14.62%	0.3926	−0.0426	<b>0.1116*</b>	<b>13.36%</b>	<b>0.4493</b>	−0.0408
Whole image	0.1300	32.54%	0.8744	−0.0053	0.1646	41.19%	0.7778	−0.0295	0.1002	25.06%	0.9238	−0.0012	<b>0.0884*</b>	<b>22.12%</b>	<b>0.9376</b>	−0.0001

\* For  $t$ -test, means  $p < 0.05$ .

\*\* For  $t$ -test, means  $p < 0.01$ .



**Fig. 9.** Frequency distributions of the Red and NIR bands and the NDVI for (a) the Coleambally irrigation area on 12 January 2002 and (b) the Gwydir area on 12 December 2004. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

NDVI displayed two peaks with significantly greater variance than those of the Red and NIR bands because of the amplification of vegetation signals and the suppression of non-vegetation signals. To further investigate the performance of the proposed method in sub-regions with different NDVI signatures, we divided the frequency distribution of the NDVI (Fig. 9a) into three parts: low NDVI ( $< 0.4$ ), medium NDVI (0.4–0.7), and high NDVI ( $> 0.7$ ). It is apparent from Table 1 that the IFSDAF and FSDAF were more accurate than the NDVI-LMGM and STARFM. Moreover, the IFSDAF performed better than the FSDAF in the medium NDVI and high NDVI sections. The  $t$ -test also suggests that the IFSDAF is statistically significant in performance improvement compared with FSDAF.

For the Gwydir site, where there was flooding, the IFSDAF fusion result (RMSE = 0.0546) captured the change (Fig. 10d) and produced results that more closely resembled the actual NDVI pattern than the other three methods. The NDVI-LMGM produced a fused image (RMSE = 0.0794) with significant block effects (Fig. 10a). The image produced by STARFM (RMSE = 0.0686) was generally similar to the actual NDVI image (Fig. 10b). FSDAF also produced an accurate fused image (RMSE = 0.0617), but it made erroneous predictions for some pixels (Fig. 10c). The blue arrow in Fig. 10c indicates the error edges produced by the FSDAF. In fact, before the flood, there was a small river in the enlarged area that resulted in the edge (marked by the blue arrow) between water and barren land. However, the river overflowed as a result of the flood and covered nearby farmland. The original edge of the river therefore disappeared, as shown in the actual Landsat NDVI image (Fig. 10e). IFSDAF was the only one of the four methods that captured this phenomenon. The scatter plots in Fig. 11a-d show no obvious bias among the four methods, but the fact that the FSDAF and IFSDAF points are closer to the 1:1 line than the NDVI-LMGM and STARFM points indicates that IFSDAF and FSDAF have comparable capacity to capture land cover changes.

Fig. 9b shows frequency distributions of the Red and NIR bands and

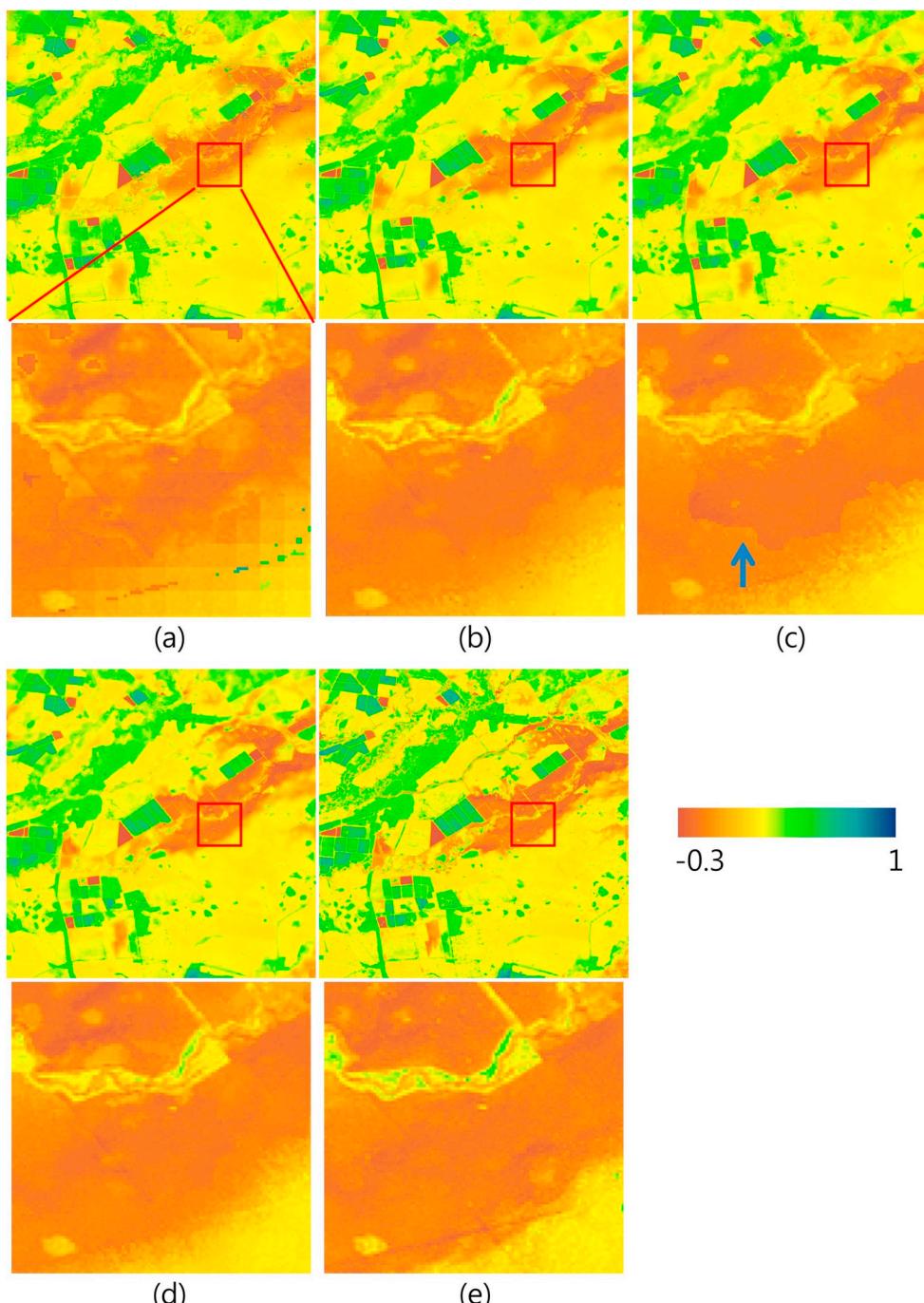
the NDVI on 12 December 2004 ( $t_p$ ). The variance of the NDVI was significantly higher than that of the Red or NIR bands. Moreover, because of the flood, there were many negative NDVI values that caused three peaks around NDVI = −0.2, NDVI = 0.2, and NDVI = 0.4 in the frequency distribution of NDVI. We divided the whole image into three parts—low NDVI ( $< 0$ ), medium NDVI (0–0.3), and high NDVI ( $> 0.3$ )—to quantitatively assess the accuracy. It is clear (Table 3) that the IFSDAF was highly accurate. The RMSE of 0.0546 was lower, and the  $r$  of 0.953 was higher, than the corresponding statistics for the other three methods. The IFSDAF was also more accurate in each of the low, medium, and high NDVI regions, where the RMSE statistics for the IFSDAF were 0.0798, 0.0467, and 0.0584, respectively. The  $t$ -test also indicates that the IFSDAF is statistically significant in performance improvement compared with FSDAF in each of the low, medium, and high NDVI regions.

#### 4.2. Fusion using multiple fine images partly covered by clouds

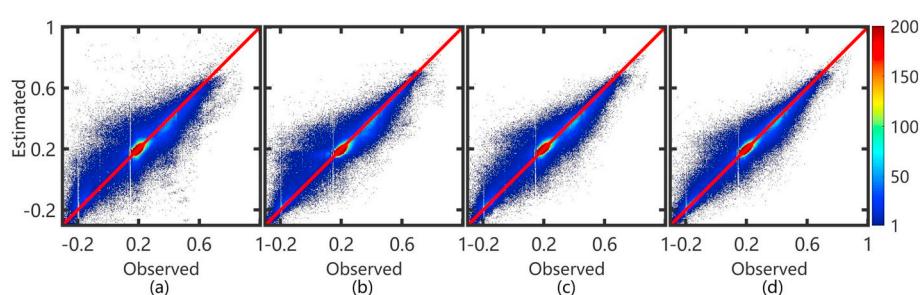
Twenty-three Landsat-like NDVI images were produced by IFSDAF using the 16-day MODIS NDVI time series and available Landsat images in 2015 (see all 23 fused NDVI images and the fused NDVI profiles of sample pixels in the Supplementary data). Here, four Landsat NDVI images of the Shennongjia site captured on 14 April, 3 July, 5 September, and 10 December 2015 served as reference data for the independent validation. In an example analysis, the 14 April image was predicted by IFSDAF using all other eight partly contaminated fine NDVI images as input, and then clear pixels of the true 14 April image were used to assess the accuracy of the predicted 14 April image. For comparison, FSDAF used only one clear Landsat NDVI image on 14 October 2015 to predict the above four Landsat NDVI images. Table 4 summarizes the accuracies of the fusion results of the four images, and Fig. 12 shows the predictions. To simplify the comparison, the results of the NDVI-LMGM and STARFM fusions are not shown because they yielded less accurate results than FSDAF.

It is evident from Fig. 12 that IFSDAF can produce fused images that more closely resemble the real Landsat NDVI than FSDAF. The RMSE values on all dates were lower for IFSDAF than for FSDAF (Table 4). These improvements of accuracy are attributable mainly to the extra information provided by the partly contaminated Landsat images, which can be put to good use in IFSDAF but not in FSDAF. In contrast, FSDAF used only one fine image on 14 October 2015, which was separated by as much as 3–6 months from several of the prediction dates. This long time interval led to low accuracy on these prediction dates. More important, the fact that improvement of the IFSDAF on 3 July and 5 September during the peak stage of vegetation growth was more statistically significant than the improvement on the other two dates indicates that IFSDAF may be more effective for fusing images with medium to high NDVI values. This result is similar to the result of the Coleambally irrigation area experiment.

At the South Asia site, the clear Sentinel-2A NDVI image on 5 March 2017 was selected as the base image. The other two clear Sentinel NDVI images (12 February and 20 December) were used as reference data to



**Fig. 10.** Landsat NDVI for the Gwydir site on 12 December 2004: predictions by (a) NDVI-LMGM, (b) STARFM, (c) FSDAF, (d) IFSDAF, and (e) the actual NDVI. For better visualization, all images were enlarged in the supplementary data (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.).



**Fig. 11.** Scatter plots of estimated results compared with observed value of Landsat NDVI on 12 December 2004: (a) NDVI-LMGM, (b) STARFM, (c) FSDAF, and (d) IFSDAF.

**Table 3**  
Comparison of RMSE, rRMSE,  $r$ , and AD between predicted NDVI and observed NDVI of the NDVI-LMGM, STARFM, FSDAF, and IFSDAF methods in the Gwydir area. The results of  $t$ -test between prediction accuracy of IFSDAF and FSDAF are marked in RMSE column under IFSDAF.

Methods	NDVI-LMGM			STARFM			FSDAF			IFSDAF		
	RMSE	rRMSE	$r$	AD	RMSE	rRMSE	$r$	AD	RMSE	rRMSE	$r$	AD
Low NDVI	0.1267	–89.59%	0.6406	0.0708	0.0964	–68.16%	0.7364	0.0497	0.0906	–64.03%	0.7521	0.0431
Medium NDVI	0.0636	33.46%	0.4872	0.0061	0.0526	27.65%	0.5794	<b>0.0042***</b>	0.0516	27.11%	0.6040	0.0066
High NDVI	0.0858	19.67%	0.7425	–0.0442	0.0654	15.00%	0.8410	–0.0341	0.0679	15.55%	0.8302	–0.0368
Whole image	0.0794	37.45%	0.8970	0.0013	0.0686	29.53%	0.9250	0.0028	0.0617	29.09%	0.9395	0.0002

\* For  $t$ -test, means  $p < 0.05$ .  
\*\* For  $t$ -test, means  $p < 0.01$ .

**Table 4**

Comparison of RMSE, rRMSE,  $r$ , and AD between the predicted NDVI and observed partly contaminated fine NDVI on 14 April, 3 July, 5 September, and 10 December 2015 in the Shennongjia Forestry District.

Date	Methods	RMSE	rRMSE	$r$	AD
Apr. 14	FSDAF	0.0873	13.33%	0.6319	–0.0481
	IFSDAF	<b>0.0819</b>	<b>12.51%</b>	<b>0.6620***</b>	<b>–0.0475</b>
Jul. 3	FSDAF	0.0578	6.44%	0.6504	–0.0138
	IFSDAF	<b>0.0368</b>	<b>4.09%</b>	<b>0.8508</b>	<b>–0.0137</b>
Sep. 5	FSDAF	0.0671	7.86%	0.7279	–0.0306
	IFSDAF	<b>0.0393</b>	<b>4.61%</b>	<b>0.8615</b>	<b>–0.0173</b>
Dec. 10	FSDAF	0.1246	21.24%	0.6516	–0.0729
	IFSDAF	<b>0.0913</b>	<b>15.57%</b>	<b>0.7768</b>	<b>–0.0366</b>

\* For  $t$ -test, means  $p < 0.05$ .

\*\* For  $t$ -test, means  $p < 0.01$ .

assess the accuracy of the IFSDAF and FSDAF. The base fine spatial resolution NDVI image and the 21 partly cloud-contaminated fine NDVI images were used as input for IFSDAF, whereas only the 5 March 2017 image was input to FSDAF. The results (Fig. 13 and Table 5) showed that the IFSDAF produced more accurate predictions. The RMSE values were lower on both dates for IFSDAF (0.0863 and 0.0740) compared with FSDAF (0.0999 and 0.1469).

## 5. Discussion

The proposed spatiotemporal data fusion model IFSDAF, involves several assumptions. The superiority over the original FSDAF model implies the proposed method has some strengths. This section will discuss the rationale behind key steps and the major improvements compared with FSDAF.

### 5.1. Derivation of space-dependent increments

In this study, a space-dependent increment ( $\Delta S$ ) was derived based on the difference between interpolation results of coarse NDVI on dates  $t_p$  and  $t_0$ , as shown in Eq. (4). However, it is also possible to obtain the  $\Delta S$  by using the fine NDVI values on date  $t_0$   $F_0(x_j, y_j)$  as following:

$$\Delta S(x_j, y_j) = F_p^{\text{TPS}}(x_j, y_j) - F_0(x_j, y_j), \quad (12)$$

We argue that the  $\Delta S$  derived from Eq. (4) is a better indicator than from Eq. (12), based on a theoretical comparison of these two  $\Delta S$  values explained below. The predicted fine NDVI on date  $t_p$  (i.e., Eq. (9)) can be simplified as following:

$$\hat{F}_{0,p} = F_0 + \Delta F^{\text{Com}} = F_0 + w_S \Delta S + w_T \Delta T, \quad (13)$$

where the residuals  $R$  had been discarded because they are negligible, and the notation  $(x_i, y_i)$  was removed for simplified expression. Because  $w_S + w_T = 1$ ,  $F_0$  can be replaced by  $w_S F_0 + w_T F_0$ . Then

$$\hat{F}_{0,p} = w_S (F_0 + \Delta S) + w_T (F_0 + \Delta T) \quad (14)$$

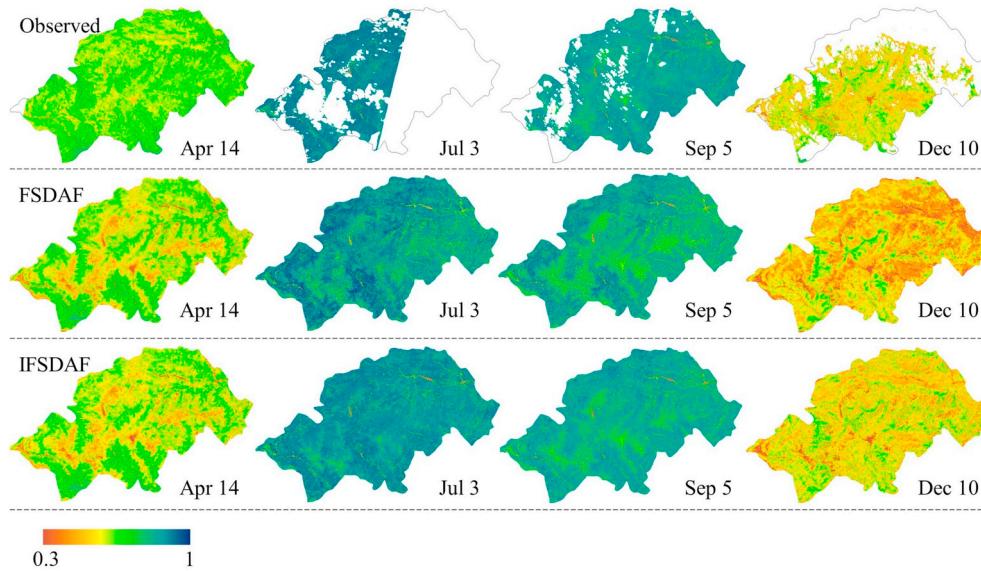
If the right-hand side of Eq. (4) is substituted for  $\Delta S$ , Eq. (14) can be rewritten as follows:

$$\hat{F}_{0,p} = w_S (F_0 + F_p^{\text{TPS}} - F_0^{\text{TPS}}) + w_T (F_0 + \Delta T). \quad (15)$$

Alternatively, if the right-hand side of Eq. (12) is substituted for  $\Delta S$ , Eq. (14) can also be rewritten as follows:

$$\begin{aligned} \hat{F}_{0,p} &= w_S (F_0 + F_p^{\text{TPS}} - F_0) + w_T (F_0 + \Delta T) \\ &= w_S (F_p^{\text{TPS}}) + w_T (F_0 + \Delta T) \end{aligned} \quad (16)$$

The difference between Eqs. (15) and (16) is the term  $F_0 - F_0^{\text{TPS}}$  in Eq. (15). The TPS prediction is a spatially smoothed prediction that loses spatial details to some degree. As a result,  $F_0 - F_0^{\text{TPS}}$  functions similarly to a high-pass filter that modulates spatial contrast at time  $t_0$ .



**Fig. 12.** Landsat 8 NDVI in the Shennongjia Forestry District on 14 April, 3 July, 5 September, and 10 December 2015 predicted by FSDAF and IFSDAF.

Several fusion models (Song and Huang, 2013; Luo et al., 2018) have assumed that this spatial contrast is relatively stable from  $t_0$  to  $t_p$ . Therefore, the term  $F_0 - F_0^{\text{TPS}}$  in Eq. (15) should better capture spatial details in the fused image (see an experiment comparing the two space-dependent increments in supplementary data S5).

### 5.2. Rationale for assuming weights to be scale-invariant

The optimized weights used for combining the space-dependent increments and time-dependent increments were assumed to be scale-invariant in the proposed IFSDAF method (Section 2.3). This assumption has also been verified by an experiment conducted in the Co-leambally irrigation area. A moving window of  $7 \times 7$  coarse-resolution pixels was used to calculate the weights ( $w_S$  and  $w_T$ ) for the space- and time-dependent increments of the center coarse pixel. Meanwhile, the fine resolution weights were derived based on the fine increment ( $\Delta F = F_p - F_0$ ) by using a fine NDVI image  $F_p$  through the CLS method. The similarity of the scatter plots of the space-dependent increment weights derived from the two approaches (Fig. 14a) suggested that the weights derived from both the coarse and fine images were interchangeable. The combined increments calculated using the two types of weights agreed well (Fig. 14b). The RMSE values of the fusion results using these two types of weights (0.0941 vs. 0.0934 to weights based on  $\Delta C$  and  $\Delta F$ , respectively) were not significantly different ( $t$ -test,  $p = 0.15$ ). The scale effect on the derived weights was therefore negligible and would not cause significant differences in the combined increment.

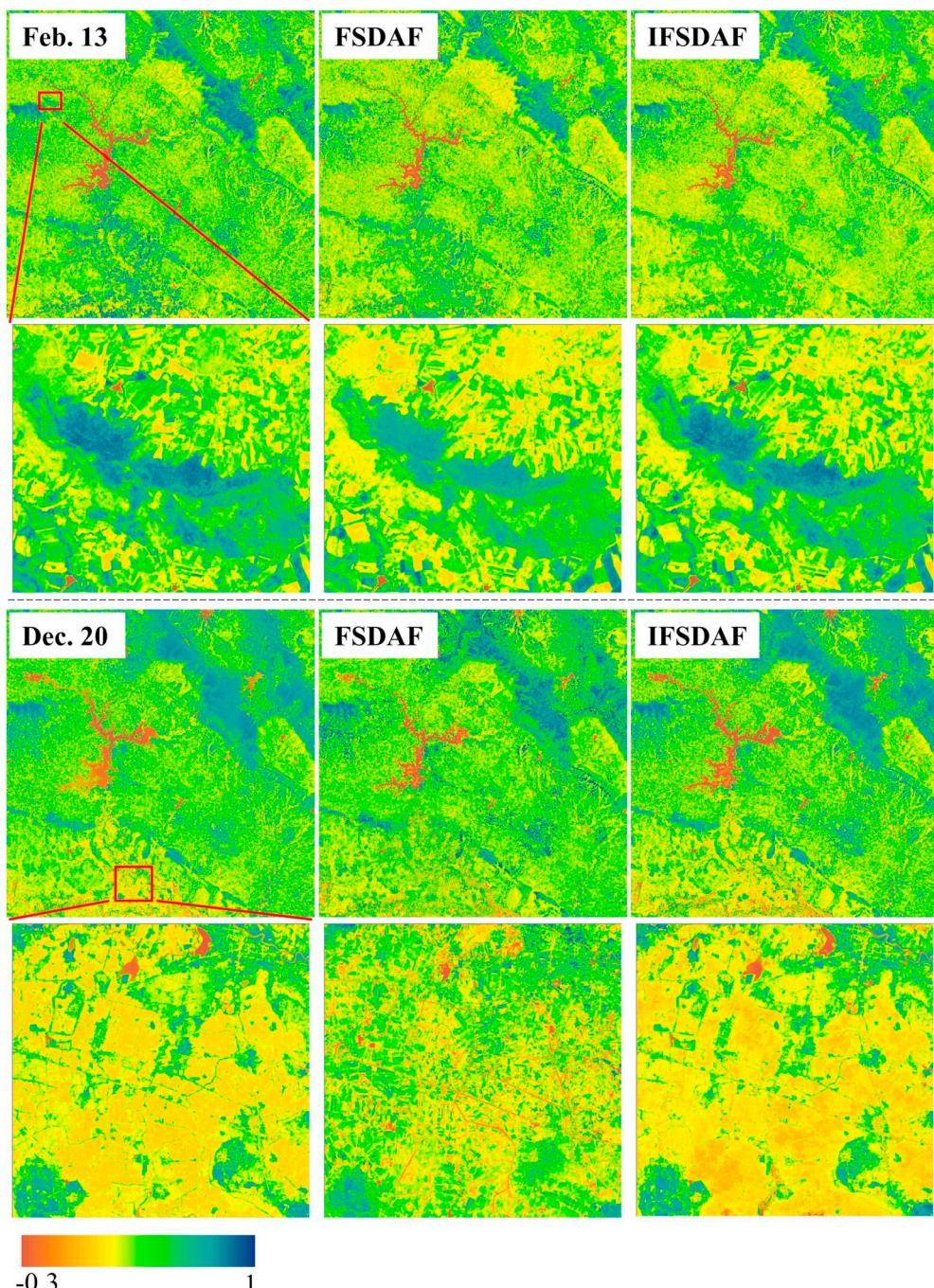
### 5.3. Necessity for combining time-dependent and space-dependent increments

In IFSDAF, the time-dependent and space-dependent increments were combined in moving windows with the CLS method. Such a combination was based on the assumption that the accuracies of the two increment estimations were different under different scenarios. The weighted combination was therefore able to improve the accuracy of the NDVI prediction by balancing the biases in the estimates of the two increments. We verified this argument by comparing the time-dependent increment, the space-dependent increment, and the combined increment with the real increment (Fig. 15). The performance of the CLS-based combination agreed with our expectation. The small RMSEs at both study sites demonstrated the necessity of combining the two

increments. Moreover, the residual of the space-dependent increment ( $\Delta S - \Delta F$ ) was much closer to the residual of the combined increment ( $\Delta F^{\text{Com}} - \Delta F$ ) than the residual of the time-dependent increment ( $\Delta T - \Delta F$ ). This suggested that the space-dependent increment contributed more to the combined increment than the time-dependent increment.

### 5.4. Improvements of IFSDAF compared with FSDAF

IFSDAF was an improvement over FSDAF in the following ways. First, IFSDAF outperformed the FSDAF for fusing remote sensing images with high spatial autocorrelation, i.e., the relationship between pixel values their neighboring pixels. With FSDAF, the increment estimation was produced mainly by the unmixing process, and the TPS interpolation result was used only to guide the distribution of residuals rather than to produce space-dependent increments. However, as shown in Fig. 15, the space-dependent increment estimated by the TPS interpolation may have been more accurate than the time-dependent increment made with the unmixing process. The FSDAF underestimated the contribution of the TPS interpolation to some extent. The reason why the space-dependent increment was superior to the time-dependent increment is apparent in Table 6, where we calculated the global Moran's I index (Paradis, 2011) of the coarse images for the Red and NIR bands, as well as the NDVI, on the base date  $t_0$  and prediction date  $t_p$ . The global Moran's I index measures the spatial autocorrelation of an image. The larger the Moran's I index, the higher the spatial autocorrelation. Table 6 shows that the spatial autocorrelation of the NDVI was greater than that of both Red and NIR reflectance. This is because the NDVI is a feature-enhancing index, increases the data variance compared with reflectance in either the Red or NIR bands (Fig. 9a–b). It is well known that greater spatial autocorrelation can yield more accurate results in spatial interpolation (Gozdowski et al., 2015). Accordingly, the space-dependent increment estimated by TPS interpolation should be more accurate for the NDVI than for the Red and NIR reflectance. The fact that the space-dependent increment was more important for fusing the NDVI than for the raw bands individually greatly benefitted the NDVI fusion in IFSDAF. This study implies that use of the space-dependent increment alone may produce acceptable fusion results, without the need to combine the time-dependent and space-dependent increments. This simplification of the fusion process would greatly reduce the computing cost. As a result, this simplified fusion model would be useful for applications in relatively large areas



**Fig. 13.** Sentinel-2A NDVI images in the Shennongjia Forestry District on 13 February and 20 December (left) and the results predicted by FSDAF (middle) and IFSDAF (right).

**Table 5**

Comparison of RMSE, rRMSE,  $r$ , and AD between the predicted NDVI and observed fine NDVI on 12 February and 20 December 2017 with Sentinel-2A data.

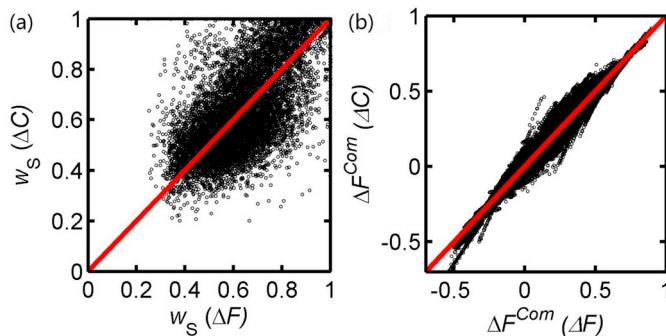
Date	Methods	RMSE	rRMSE	$r$	AD
Feb. 13	FSDAF	0.0999	24.04%	0.8885	-0.0463
	IFSDAF	0.0863*	20.76%	0.9305***	-0.0427
Dec. 20	FSDAF	0.1469	33.69%	0.7401	-0.0082
	IFSDAF	0.0740*	17.69%	0.9584	-0.0141

\* For  $t$ -test results, means  $p < 0.05$ .

\*\* For  $t$ -test results, means  $p < 0.01$ .

and when the scale difference between the coarse and fine images is relatively small. The latter condition ensures that the number of sample points (the centers of the coarse pixels) is adequate to obtain an accurate prediction of the fine image via TPS interpolation.

Second, IFSDAF is more flexible than FSDAF in dealing with the estimation errors of time-dependent increments. The collinearity effect impacts the accuracy of unmixing for time-dependent increment estimation. Moreover, errors in the classification map and changes in land cover also cause uncertainties in the time-dependent increment. To correct the potential errors in the time-dependent increment, FSDAF introduces a homogeneity index  $HI(x_j, y_j)$ , which is derived from the classification map at  $t_0$  and ranges in value from 0 to 1. This index helps allocate the residuals  $R(x, y)$  within the coarse pixel. However, when



**Fig. 14.** Scatter plot of weights of space-dependent increments based on (a) coarse resolution increment ( $\Delta C$ ) and fine resolution increment ( $\Delta F$ ); (b) comparison of combined increments using weights derived from coarse and fine images.

there are land cover changes and misclassifications, the value of  $HI(x_j, y_j)$  calculated from the classification map at  $t_0$  will not be suitable for allocating residuals on date  $t_p$ . In this case, the effectiveness of residual distribution in FSDAF is restricted. Unlike FSDAF, the IFSDAF employs the CLS method to moving windows and avoids the use of  $HI(x_j, y_j)$ ; moreover, it allows the final increment estimation with local and adaptive capacity to better combine the time-dependent and space-dependent increments.

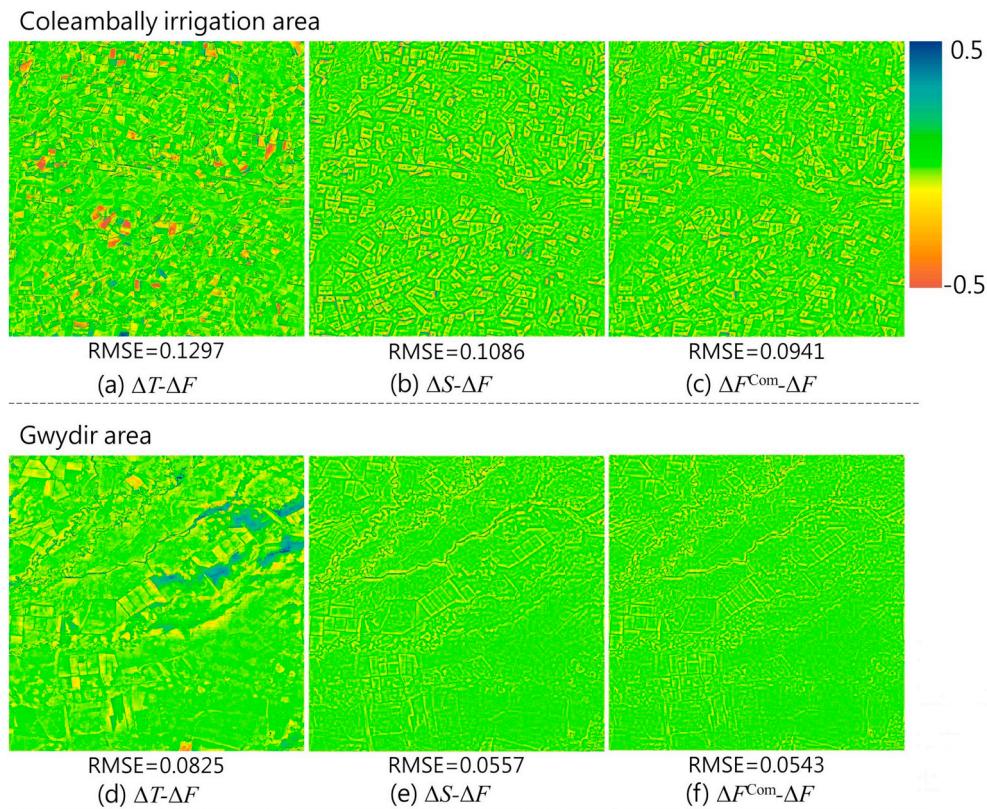
The third advantage of IFSDAF is that it employs fine NDVI images partly contaminated by clouds, in which clear pixels also provide valuable information. In IFSDAF, the clear pixels in those fine images are also used as base data to estimate the fine NDVI values on the prediction date, and all predictions on date  $t_p$  are finally integrated using weights based on the magnitude of the temporal change in the NDVI between base and prediction dates. This weighted prediction can reduce

**Table 6**

Moran's I of Red and NIR bands as well as the NDVI on the base date  $t_0$  and prediction date  $t_p$  in both the Coleambally irrigation area and Gwydir area at coarse resolution.

Band	Base date $t_0$			Prediction date $t_p$		
	Red	NIR	NDVI	Red	NIR	NDVI
Coleambally irrigation area	0.5048	0.5225	0.6439	0.5677	0.4764	0.6840
Gwydir area	0.5867	0.7069	0.7881	0.6401	0.7568	0.8584

the critical dependency on the clear fine NDVI image and alleviate the prediction uncertainties if the date of the clear fine NDVI image is much earlier than the prediction date. Of course, better use of partly contaminated images requires an accurate cloud labeling method (e.g., the Fmask method). If there are mistakes in cloud labels, the IFSDAF estimates will be impacted. For instance, a land surface with high reflectance (e.g., sand or snow) may be misidentified as clouds (Chen et al., 2016). Moreover, the fact that Fmask sometimes omits thin clouds can result in cloudy pixels being used in the process of data fusion. Fortunately, the effect of errors in the cloud mask is minimized in IFSDAF because it uses a weighted combination of predictions from multiple dates. Moreover, with the advance of cloud-screening methods (Zhu and Helmer, 2018), this problem can be greatly alleviated. In addition, like other existing spatiotemporal fusion methods, a completely clear image guarantees that all pixels are predicted by IFSDAF. However, in areas such as tropical regions (e.g., Amazonia), completely clear fine images are rarely available throughout the year. Under such conditions, using all partly contaminated fine images instead of a completely clear image is a practical possibility in IFSDAF, although such use may omit predictions for some pixels in the fused images if those pixels do not have even one cloud-free observation in the time



**Fig. 15.** Errors of different increments: (a) time-dependent increment  $\Delta T$ , (b) space-dependent increment  $\Delta S$ , and (c) combined increment  $\Delta F^{\text{Com}}$  in the Coleambally irrigation area; Errors of (d) time-dependent increment  $\Delta T$ , (e) space-dependent increment  $\Delta S$ , and (f) combined increment  $\Delta F^{\text{Com}}$  in the Gwydir area.

**Table 7**

RMSE (rRMSE in parenthesis) values of predictions by IFSDAF and FSDAF of EVI, Red and NIR bands.

Index/surface reflectance	EVI	Red	NIR
FSDAF	0.0755 (29.11%)	0.0271 (19.03%)	0.0341 (11.02%)
IFSDAF	0.0650 (25.09%)	0.0245 (17.21%)	0.0337 (10.91%)

series. A by-product of using all partly cloudy images with IFSDAF is reconstruction of cloudy pixels in these input fine images, i.e., replacing the clouded pixels in these fine images by the IFSDAF predictions, as shown in Fig. 12.

### 5.5. Applications to other remote sensing products

Although IFSDAF is designed for the spatiotemporal fusion of NDVI time series, it also has the potential to fuse other vegetation indices, such as the Enhanced Vegetation Index (EVI; Huete et al., 2002), and other products (e.g., surface reflectance). To test the applicability of IFSDAF to other products, we assessed its performance in fusing EVI, Red and NIR reflectance in the Coleambally Irrigation area, within which there is great heterogeneity of vegetation. RMSEs of the fused images on 12 January 2002 (Table 7) suggested that IFSDAF produced more accurate results than FSDAF when fusing EVI. However, the accuracies of surface reflectance predictions (Red and NIR bands) did not differ remarkably between IFSDAF and FSDAF. Still, these results confirm that IFSDAF is more suitable than the original FSDAF model for fusing remote sensing products with high spatial autocorrelation.

## 6. Conclusions

In this study, we proposed an improved FSDAF method specifically for producing NDVI time series with a high spatiotemporal resolution. Coarse NDVI (MODIS) and fine NDVI images (Landsat and Sentinel) were used to test the performance of the new method for the two different sensors. Experiments showed that the NDVI images were more accurate when they were fused by IFSDAF than by FSDAF as well as by

two other existing methods (NDVI-LMGM and STARFM) in areas with a great degree of spatial heterogeneity and with significant land cover changes. The better performance of IFSDAF can be attributed to its production of space-dependent increments by TPS interpolation, use of the CLS method in moving windows to adaptively combine the time-dependent and space-dependent increments, and better use of partly contaminated fine images. Such significant improvements were consistent with the characteristics of NDVI. It is more variable and larger spatial autocorrelation compared with the raw reflectance bands. By taking into consideration the significant contribution of the space-dependent increment via TPS interpolation when the scale difference between coarse and fine images is not very large, the proposed IFSDAF method can be further simplified by using only the space-dependent increment to improve efficiency. This result is also consistent with the IFSDAF as being a feasible method for applications in large areas and with different sensors. Moreover, IFSDAF is also applicable to other vegetation index data. We call for more testing of the new method by using other satellite data (e.g., Sentinel and VIIRS data) and in other places.

## Acknowledgement

This study was supported by National Key Research and Development Program of China (No. 2017YFD0300201), the Research Grants Council of Hong Kong (project no. 25222717), the National Natural Science Foundation of China (project no. 41701378), the Japan Society for the Promotion of Science KAKENHI under Grant 16H02948, and CEReS Oversea Joint Research Program, Chiba University (No. CI17-103).

## Appendix A

### Useful notations.

$t_0$	Base date	$f_l(x, y)$	Fraction of class $l$ within coarse pixel $(x, y)$
$t_p$	Prediction date	$\Delta F_c$	Fine spatial resolution increment of class $c$ within the moving window
$(x, y)$	Location of coarse spatial resolution pixel $(x, y)$	$F_0^{\text{TPS}}$	Result of TPS interpolation based on coarse NDVI on $t_0$
$(x_j, y_j)$	Location of $j$ th fine spatial resolution pixel within coarse pixel $(x, y)$	$F_p^{\text{TPS}}$	Result of TPS interpolation based on coarse NDVI on $t_p$
$F_0$	Fine spatial resolution NDVI on $t_0$	$w_s$	Weight of spatial-dependent increment
$F_p$	Fine spatial resolution NDVI on $t_p$	$w_t$	Weight of temporal increment
$C_0$	Coarse spatial resolution NDVI on $t_0$	$\hat{F}_p$	Fine spatial resolution prediction on date $t_p$
$C_p$	Coarse spatial resolution NDVI on $t_p$	$\hat{F}_{0,p}$	Fine spatial resolution prediction on date $t_p$ based on fine NDVI on date $t_0$
$\Delta F$	Fine spatial resolution NDVI increment	$\hat{F}_{p+1,p}$	Fine spatial resolution prediction on date $t_p$ based on fine NDVI on date $p + 1$
$\Delta C$	Coarse spatial resolution NDVI increment	$\Delta F^{\text{Com}}$	Combined fine spatial resolution increment based on $\Delta T$ and $\Delta S$
$\Delta T$	Fine spatial resolution temporal increment	$R(x, y)$	Residual within the coarse pixel $(x, y)$
$\Delta S$	Fine spatial resolution spatial-dependent increment	$C_{0'}(x, y)$	ith coarse pixel in the moving window centered by coarse pixel $(x, y)$ on date $q$
$\Delta C^T$	Upscaled fine spatial resolution temporal increment	$C_p'(x, y)$	ith coarse pixel in the moving window centered by coarse pixel $(x, y)$ on date $t_p$
$\Delta C^S$	Upscaled fine spatial resolution spatial-dependent increment	$w_{q,p}(x, y)$	Contribution coefficient of fine spatial resolution pixels on date $q$ to the final prediction on $t_p$ within coarse pixel $(x, y)$

## Appendix B. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.rse.2019.03.012>.

## References

Busetto, L., Meroni, M., Colombo, R., 2008. Combining medium and coarse spatial

resolution satellite data to improve the estimation of sub-pixel NDVI time series. *Remote Sens. Environ.* 112, 118–131.

Cao, R., Chen, Y., Shen, M., Chen, J., Zhou, J., Wang, C., Yang, W., 2018. A simple method to improve the quality of NDVI time-series data by integrating spatiotemporal

- information with the Savitzky-Golay filter. *Remote Sens. Environ.* 217, 244–257.
- Chen, J., Jonsson, P., Tamura, M., Gu, Z.H., Matsushita, B., Eklundh, L., 2004. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. *Remote Sens. Environ.* 91, 332–344.
- Chen, X., Li, W., Chen, J., Rao, Y., Yamaguchi, Y., 2014. A combination of TsHARP and thin plate spline interpolation for spatial sharpening of thermal imagery. *Remote Sens.* 6, 2845–2863.
- Chen, J., Chen, J., Liao, A.P., Cao, X., Chen, L.J., Chen, X.H., He, C.Y., Han, G., Peng, S., Lu, M., Zhang, W.W., Tong, X.H., Mills, J., 2015. Global land cover mapping at 30 m resolution: a POK-based operational approach. *ISPRS J. Photogramm. Remote Sens.* 103, 7–27.
- Chen, S.L., Chen, X.H., Chen, J., Jia, P.F., Cao, X., Liu, C.Y., 2016. An iterative haze optimized transformation for automatic cloud/haze detection of Landsat imagery. *IEEE Trans. Geosci. Remote Sens.* 54, 2682–2694.
- Chen, X.H., Liu, M., Zhu, X.L., Chen, J., Zhong, Y.F., Cao, X., 2018. “blend-then-index” or “index-then-blend”: a theoretical analysis for generating high-resolution NDVI time series by STARFM. *Photogramm. Eng. Remote. Sens.* 84, 66–74.
- Dubrule, O., 1984. Comparing splines and kriging. *Comput. Geosci.* 10, 327–338.
- Emelyanova, I.V., McVicar, T.R., Van Niel, T.G., Li, L.T., van Dijk, A.I., 2013. 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.* 133, 193–209.
- Gao, F., Masek, J., Schwaller, M., Hall, F., 2006. On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance. *IEEE Trans. Geosci. Remote Sens.* 44, 2207–2218.
- Gevaert, C.M., Garcia-Haro, F.J., 2015. A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion. *Remote Sens. Environ.* 156, 34–44.
- Gozdowski, D., Stepien, M., Samborski, S., Dobers, E., Szatylowicz, J., Chormański, J., 2015. Prediction accuracy of selected spatial interpolation methods for soil texture at farm field scale. *J. Soil Sci. Plant Nutr.* <https://doi.org/10.4067/S0718-951620150050003>.
- Huang, B., Song, H., 2012. Spatiotemporal reflectance fusion via sparse representation. *IEEE Trans. Geosci. Remote Sens.* 50, 3707–3716.
- Huang, B., Zhang, H.K., 2014. Spatio-temporal reflectance fusion via unmixing: accounting for both phenological and land-cover changes. *Int. J. Remote Sens.* 35, 6213–6233.
- Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* 83, 195–213.
- Jarihani, A., McVicar, T., Van Niel, T., Emelyanova, I., Callow, J., Johansen, K., 2014. Blending Landsat and MODIS data to generate multispectral indices: a comparison of “index-then-blend” and “blend-then-index” approaches. *Remote Sens.* 6, 9213–9238.
- Liu, X., Deng, C.W., Wang, S.G., Huang, G.B., Zhao, B.J., Lauren, P., 2016. Fast and accurate spatiotemporal fusion based upon extreme learning machine. *IEEE Geosci. Remote Sens. Lett.* 13, 2039–2043.
- Luo, Y., Guan, K., Peng, J., 2018. STAIR: a generic and fully-automated method to fuse multiple sources of optical satellite data to generate a high-resolution, daily and cloud-gap-free surface reflectance product. *Remote Sens. Environ.* 214, 87–99.
- Meng, J.H., Du, X., Wu, B.F., 2013. Generation of high spatial and temporal resolution NDVI and its application in crop biomass estimation. *Int. J. Digital Earth* 6, 203–218.
- Paradis, E., 2011. Moran's autocorrelation coefficient in comparative methods. *ReCALL*. <https://doi.org/10.1093/jb/mvu081>.
- Pettorelli, N., Vik, J.O., Mysterud, A., Gaillard, J.M., Tucker, C.J., Stenseth, N.C., 2005. Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends Ecol. Evol.* 20, 503–510.
- Rao, Y., Zhu, X., Chen, J., Wang, J., 2015. An improved method for producing high spatial-resolution NDVI time series datasets with multi-temporal MODIS NDVI data and Landsat TM/ETM+ images. *Remote Sens.* 7, 7865–7891.
- Rouse Jr., J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1974. Monitoring vegetation systems in the Great Plains with ERTS. In: Proceedings of Third ERTS Symposium, Washington, DC, USA, 10–14 December 1973, pp. 309–317.
- Song, H., Huang, B., 2013. Spatiotemporal satellite image fusion through one-pair image learning. *IEEE Trans. Geosci. Remote Sens.* 51, 1883–1896.
- Tian, F., Wang, Y.J., Fenstholz, R., Wang, K., Zhang, L., Huang, Y., 2013. Mapping and evaluation of NDVI trends from synthetic time series obtained by blending Landsat and MODIS data around a coalfield on the loess plateau. *Remote Sens.* 5, 4255–4279.
- Van Leeuwen, W.J.D., Huete, A.R., Laing, T.W., 1999. MODIS vegetation index compositing approach. *Remote Sens. Environ.* 69, 264–280.
- Wang, Q.M., Atkinson, P.M., 2018. Spatio-temporal fusion for daily Sentinel-2 images. *Remote Sens. Environ.* 204, 31–42.
- Wang, P., Teng, M., He, W., Tang, C., Yang, J., Yan, Z., 2018. Using habitat selection index for reserve planning and management for snub-nosed golden monkeys at landscape scale. *Ecol. Indic.* 93, 838–846.
- Wu, B., Huang, B., Zhang, L., 2015. An error-bound-regularized sparse coding for spatiotemporal reflectance fusion. *IEEE Trans. Geosci. Remote Sens.* 53, 6791–6803.
- Zhang, H.K., Huang, B., Zhang, M., Cao, K., Yu, L., 2015. A generalization of spatial and temporal fusion methods for remotely sensed surface parameters. *Int. J. Remote Sens.* 36, 4411–4445.
- Zhao, C.M., Chen, W.L., Tian, Z.Q., Xie, Z.Q., 2005. Altitudinal pattern of plant species diversity in Shennongjia Mountains, central China. *J. Integr. Plant Biol.* 47, 1431–1449.
- Zhu, X., Helmer, E.H., 2018. An automatic method for screening clouds and cloud shadows in optical satellite image time series in cloudy regions. *Remote Sens. Environ.* 214, 135–153.
- Zhu, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sens. Environ.* 118, 83–94.
- Zhu, X., Chen, J., Gao, F., Chen, X., Masek, J.G., 2010. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sens. Environ.* 114, 2610–2623.
- Zhu, X., Helmer, E.H., Gao, F., Liu, D., Chen, J., Lefsky, M.A., 2016. A flexible spatio-temporal method for fusing satellite images with different resolutions. *Remote Sens. Environ.* 172, 165–177.
- Zhu, X., Cai, F., Tian, J., Kay-AnnWilliams, T., 2018. Spatiotemporal fusion of multisource remote sensing data: literature survey, taxonomy, principles, applications, and future directions. *Remote Sens.* 10, 527.
- Zurita-Milla, R., Clevers, J., Schaepman, M.E., 2008. Unmixing-based Landsat TM and MERIS FR data fusion. *IEEE Geosci. Remote Sens. Lett.* 5, 453–457.