https://dx.doi.org/10.1016/j.rse.2019.03.012

An Improved Flexible Spatiotemporal DAta Fusion (IFSDAF) method for producing high spatiotemporal resolution NDVI time series

Meng Liu[a,e], Wei Yang[b], Xiaolin Zhu[c], Jin Chen[a*], Xuehong Chen[a], Linqing Yang[d,e]

[a] 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

[b] Center for Environmental Remote Sensing, Chiba University, Chiba 263-8522, Japan

[c] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University

[d] State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

[e] Department of Ecosystem Science and Management, Texas A&M University, College Station, TX 77843, USA

*Corresponding author: Prof. Jin Chen. E-mail address: chenjin@bnu.edu.cn

ABSTRACT

The Normalized Difference Vegetation Index (NDVI) is one of the mostly used vegetation index for ecosystem dynamics monitoring and biosphere process modeling. However, global NDVI products are usually provided with relatively coarse spatial resolutions, being short of important spatial details. Producing NDVI time-series data with a high spatiotemporal resolution is thus indispensable for monitoring land surface and ecosystem changes, especially in heterogeneous areas. An Improved Flexible Spatiotemporal DAta Fusion (IFSDAF) method is developed in this study to overcome the existing issues. In accordance with the distinctive characteristics of NDVI with large data variance and high spatial autocorrelation compared with raw reflectance bands, the IFSDAF method first produces temporal increment with linear unmixing and spatial-dependent increment by thin plate spline (TPS) interpolation, and then obtains final prediction from the optimal integration of two increments by Constrained Least Square (CLS) theory. Moreover, IFSDAF is developed with a capacity of employing all available and partially contaminated fine images. Coarse spatial resolution NDVI (MODIS) and fine spatial resolution NDVI images (Landsat and Sentinel) in areas with great spatial heterogeneity and significant land cover changes were used to test the performance of the new method. The promising results (RMSE 0.0884, rRMSE 22.12% in heterogeneous areas, RMSE 0.0546, rRMSE 25.77% in land cover change areas) demonstrate the strengths and robustness of the proposed method in providing reliable high spatial and temporal resolution NDVI datasets to

support research on land surface processes. The proposed IFSDAF method can be further simplified by only using spatial-dependent increment to improve the efficiency to a great extent. It will make IFSDAF a feasible method for applications in large geographical area and has the potential for global studies.

Keywords: Normalized Difference Vegetation Index (NDVI), Spatiotemporal Data Fusion, High Spatial and Temporal Resolution, Constrained Least Square (CLS) method, Weighted Integration

1. Introduction

The Normalized Difference Vegetation Index (NDVI) enhances the absorptive and reflective features of vegetation and provides a proxy for measuring canopy greenness and vigor (Rouse et al., 1974; Huete et al., 2002). Accordingly, NDVI time-series data derived from spaceborne sensors are widely employed in ecosystem dynamics monitoring and biosphere process modeling, helping to understand responses of ecosystems to climate change (Pettorelli et al., 2005). As the most significant constraint of the available NDVI time-series products (e.g., GIMMS, MODIS, SPOT VGT), coarse spatial resolutions ranging from 250 m to 8 km prevent these products from capturing spatial details necessary for monitoring land surface and ecosystem changes, especially in geographically heterogeneous areas (Gao et al., 2006; Rao et al., 2015). Producing NDVI time-series data with both high spatial and high temporal resolutions is thus critically required for such applications, raising the need for developing spatiotemporal fusion methods by blending the high frequent but low spatial resolution images (e.g., MODIS images, hereinafter referred to as coarse images) with the high spatial resolution but low frequent images (e.g., Landsat images, hereinafter referred to as fine images) (Zhu et al., 2018). Recently with emerging constellations of CubeSats and new satellite systems (e.g. Sentinel 2 with 5 day NDVI at 10 m resolution observations), new opportunities to alleviate the issue of the classical trade-off between spatial and temporal resolution is becoming hoped, however, spatiotemporal fusion is still necessary for long time series analysis as such data are unavailable before 2015.

When using spatiotemporal fusion technology to produce NDVI data with high spatial and temporal resolutions, users need to solve the two puzzles: (I) selecting an appropriate blending strategy: Blend-then-Index (BI) or Index-then-Blend (IB), and (II) selecting a suitable and accurate spatiotemporal fusion method. For the first puzzle, recent studies (Chen et al., 2018; Jarihani et al., 2014; Tian et al., 2013) have demonstrated that the IB strategy consistently yields better or comparable results than the BI, mainly because the IB method has these advantages compared with the BI: (i) less error propagation in the blending process; (ii) less computationally expensive; and (iii) easier to clean the noises (e.g., cloud effects) on NDVI than the raw reflectance bands by the advanced filters (e.g., Chen et al., 2004). Consequently, IB is generally recommended and becomes the dominant blending strategy for producing fused NDVI products.

Regarding the second puzzle, a number of spatiotemporal fusion methods have been proposed and validated over past years (Zhu et al., 2018). These methods need 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 the input. However, the consensus regarding the most suitable method for producing high spatiotemporal resolution NDVI data has not been reached. Generally, as a band combination index for feature enhancement, NDVI enlarges the contrast between vegetated and non-vegetated pixels and therefore displays larger spatial and temporal variance (i.e., larger heterogeneity) than the raw reflectance in most satellite images. Accordingly, a suitable spatiotemporal fusion method for fusing NDVI product is supposed to satisfy the following criteria in practice: (i) obtaining good prediction in areas with large spatial and temporal variance; (ii) requiring only one pair of clear fine and coarse NDVI image at a base date, ensuring its applicability in areas with frequent cloud contamination; (iii) having a capacity to handle land cover change, such as urbanization, deforestation/reforestation, wildfires, floods and land cover transitions caused by other forces. Among the existing spatiotemporal fusion methods, the Flexible Spatiotemporal DAta Fusion method (FSDAF) (Zhu et al., 2016) is the one meeting these criteria and can be considered a potential candidate, while other existing methods fail in at least one criterion, especially the third criterion. For example, 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), unmixing-based spatiotemporal reflectance fusion model (U-STFM, Huang and Zhang, 2014), NDVI linear mixing growth model (NDVI-LMGM, Rao et al., 2015), and spatial and temporal reflectance unmixing model (STRUM, Gevaert and Garcia-Haro, 2015) cannot handle land cover changes occurring between base date and prediction date. The learning-based methods, such as Sparse-representation-based spatiotemporal reflectance fusion model (SPSTFM, Huang and Song, 2012; Song and Huang, 2013), an error-bound-regularized semi-coupled dictionary learning model

(EBSCDM, Wu et al., 2015) and an extreme learning machine based fusion method (Liu et al., 2016) can better capture land cover change but their learning step is time consuming, and the accuracy decreases when the spatial heterogeneity is high and scale differences between coarse and fine images are large (Zhu et al., 2016).

FSDAF is based on the spectral unmixing analysis and further introduces thin plate spline (TPS) interpolation to capture land cover change if the change is detectable in coarse images (Zhu et al., 2016). Compared with two widely used spatiotemporal fusion methods, STAFRM algorithm (Gao et al., 2006) and an unmixing-based data fusion (UBDF) algorithm (Zurita-Milla et al., 2008), FSDAF needs the same input data as these two methods but is superior in producing more accurate predictions especially in the NIR band of heterogeneous landscapes (Table 3 and Table 4 in Zhu et al., 2016). Like NDVI, the NIR band has larger spatial and temporal variances than red band, because the reflectance in NIR band generally has larger difference among different land covers than red band, and it has more significant temporal changes than red band during vegetation growth cycles. Moreover, FSDAF can capture both the gradual and abrupt land cover changes, which is an existing issue with current spatiotemporal fusion methods. Considering many advantages of FSDAF, it could be the appropriate method for producing high spatiotemporal resolution NDVI data. However, there is still space to further improve the FSDAF method. It should be noted that the FSDAF method only relies on the result of TPS interpolation to distribute residuals (ε) between prediction and true values under an assumption that errors mainly depend on the landscape homogeneity. Such an assumption is very empirical and has not been demonstrated by theoretical analysis. It may be not an optimal way to distribute residuals for different scenarios. Furthermore, in practice, many available fine images are partially contaminated by clouds. Clear pixels on these partially contaminated fine images can provide significant information of temporal changes, demonstrated by STAIR method proposed by Luo et al. (2018) with better result of producing daily surface reflectance than STARFM. Consequently, using cloud-free fine images together with partially contaminated fine images will benefit spatiotemporal NDVI fusion and expand its applicability to clouded regions. Unfortunately, the FSDAF method falls short in such a capacity and is not applicable in clouded regions.

To address the abovementioned limitations, we propose an Improved Flexible Spatiotemporal DAta Fusion (IFSDAF) method for producing high spatiotemporal resolution NDVI time series. The IFSDAF incorporates Constrained Least Square (CLS) theory into FSDAF method, by which temporal prediction derived from unmixing procedure and spatial prediction derived from TPS interpolation are combined, thus ensuring final prediction obtained from the optimal integration of temporal and spatial predictions. Moreover, IFSDAF was developed with the capacity of employing all available and partially contaminated fine images (e.g. maximum cloud coverage is less than 70%). To validate the effectiveness of the proposed method, comparison with three popular NDVI fusion methods (i.e., NDVI-LMGM, STARFM and FSDAF) under the IB strategy were performed in several experiment areas, including a site with a heterogeneous landscape, a site with abrupt land cover changes, and a site where satellite images contain a lot of clouds.

2. Methodology

Although the principles of existing spatiotemporal fusion methods have a great variety, the main idea can be framed by Eq. (1), in which the fine increment of NDVI (ΔF) between the predicting date (t_p) and the base date (t_0) is firstly estimated, and then fine NDVI values (F_p) on the predicting date (t_p) are predicted as the sum of the base fine NDVI value (F_0) and the increment (ΔF), plus the residuals ε.

$$F_p = F_0 + \Delta F + \varepsilon \tag{1}$$

Given that F_0 is known, IFSDAF also follows this unified equation but estimates the increment in two ways: the temporal increments using (i) unmixing analysis and spatial-dependent increments using (ii) the Thin Plate Spline (TPS) interpolation method, and then combine the two increments to obtain final ΔF through a Constrained Least Square (CLS) method. The CLS method adopted here purifies the original FSDAF because it can adaptively combine the two increments, allowing the final ΔF approaching the one with higher accuracy.

The flowchart of the proposed IFSDAF is shown in Fig. 1. The input data for IFSDAF include coarse NDVI time series images and all available fine NDVI images within the same period. In these images, coarse NDVI and fine NDVI images acquired at the same dates are named as one pair. The pair with minimal cloud contaminations is selected as the base images (C_0 and F_0) and its acquisition date is the base date t_0. The dates of other pairs are denoted as $\cdots$, p-3, p-2, p-1, p+1, p+2, p+3, $\cdots$. The coarse and fine NDVI images of these pairs are denoted as ($\cdots$, $C_{p\text{-}3}$, $C_{p\text{-}2}$, $C_{p\text{-}1}$, C_{p+1}, C_{p+2}, C_{p+3}, $\cdots$) and ($\cdots$, $F_{p\text{-}3}$, $F_{p\text{-}2}$, $F_{p\text{-}1}$, F_{p+1}, F_{p+2}, F_{p+3}, $\cdots$) respectively. The task of IFSDAF is to predict fine NDVI images at any dates whenever a course NDVI image is available, e.g., the date of t_p. In IFSDAF, the input fine NDVI images are not required to be cloud free except F_0. Like other spatiotemporal fusion methods, all the coarse and fine NDVI images need to be geo-registered and cropped to become the same image size. Besides, coarse NDVI time-series images need to be smoothed by an algorithm based on Savitzky-Golay filter (Chen et al., 2004), which was designed to reconstruct a high-quality NDVI time-series data by keeping the clear-sky values and interpolating clouded values. And cloud pixels in partially cloud-contaminated fine NDVI images are also masked by the Fmask algorithm (Zhu and Woodcock, 2012). A land cover classification map at a fine resolution, which can be derived from either existing land cover products (e.g. Globeland30, Chen et al., 2015) or the classification result of the input clear fine images, is needed to provide fractional cover for the unmixing process. The output of IFSDAF is synthetic fine NDVI images ($\hat{F}_p$) on the prediction date t_p (p=1, 2, 3, …). More detailed description for each implementation step of IFSDAF is given below and a list of notations and explanation is given in Appendix.

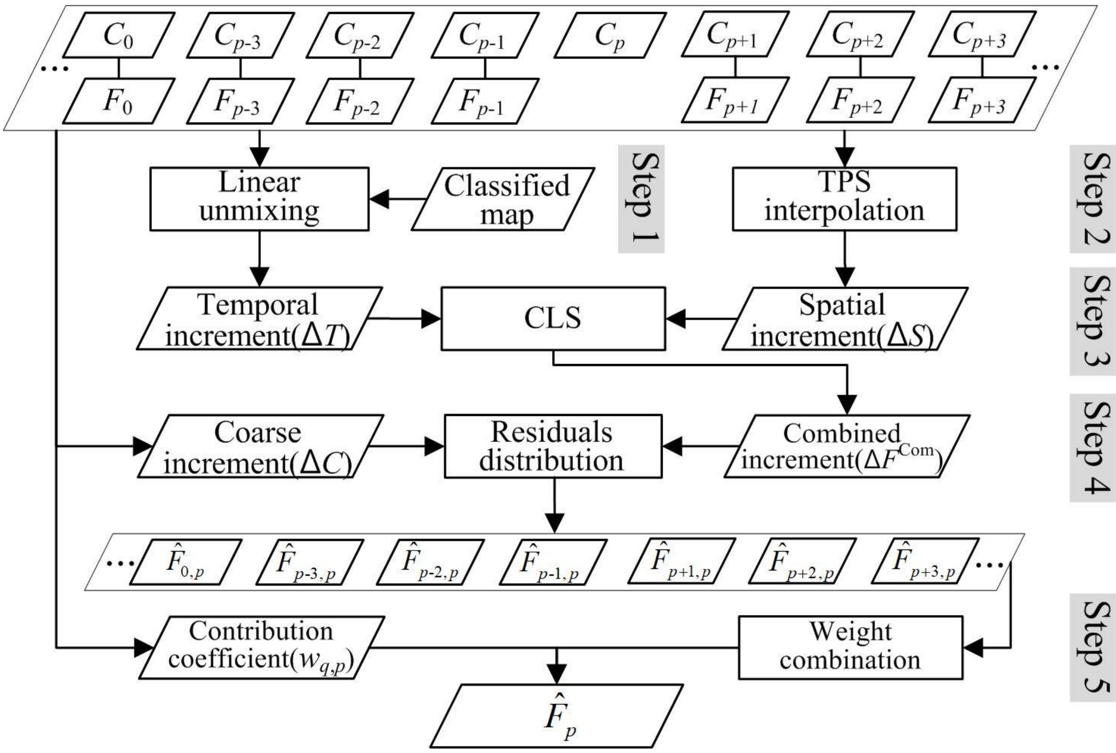

Fig.1. Flowchart of the Improved Flexible Spatiotemporal DAta Fusion method (IFSDAF)

2.1 Generation of temporal increment by unmixing method

Following the linear spectral mixing theory, the temporal NDVI change (increment) of a coarse pixel can be considered as the linear combination of NDVI increments of fine pixels within the coarse pixel during a short period (Rao et al., 2015). Accordingly, a linear mixture model is used to unmix the increment of coarse pixels from the base date t_0 to the prediction date t_p, assuming that fine pixels belonging to the same land cover class have a similar increment within the local region (Busetto et al., 2008; Rao et al., 2015). Neighboring coarse pixels within a moving window centered by coarse pixel (x, y) are used to establish a linear equation system, as shown in Eq. (2).

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

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

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. ΔF_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. ΔC_{window} is the set of all coarse NDVI increments in the window. $\min(\Delta C_{\text{window}})$, $\max(\Delta C_{\text{window}})$, and $\text{std}(\Delta C_{\text{window}})$ are the minimum value, maximum value and standard deviation of ΔC_{window}, respectively. A moving window sized at a 7×7 coarse pixel 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 the abovementioned overdetermined linear equations less influenced by collinearity and land cover changes. By solving the linear equations, the temporal NDVI increment of each class (ΔF_c) in the moving window can be acquired. Then, the fine temporal increment $\Delta T(x_j, y_j)$, where (x_j, y_j) devotes jth fine pixel in the coarse pixel (x, y), is defined by Eq. (3), as following,

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

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, 2017). Users need to set the number of classes in unsupervised classification. According to previous studies, the number of classes ranging from 3 to 6 could get satisfied results for most situations (Rao et.al 2015, Zhu et.al, 2017). Accuracy assessment of the classification map is not included in the fusion process because: (1) aggregation of fine-scale class to coarse-scale fraction will average out some errors in classification so it may not cause large problem in solving Eq. (2); (2) temporal change assigned to a pixel with wrong class labels using Eq. (3) will be compensated by the spatial-dependent increment introduced in the next section; and (3) reference samples selection for accuracy assessment introduces more human-computer interaction. Although the proposed method is not sensitive to classification accuracy, including more accurate and robust classification methods in IFSDAF could further improve its performance.

2.2 Generation of spatial-dependent increment by TPS interpolation

Coarse NDVI image on t_p contains signals of land cover changes when changes are significant enough to be shown in coarse pixels. Therefore, spatial interpolation of coarse NDVI to fine resolution will retain useful information of land cover changes. Accordingly, coarse spatial resolution NDVI images on t_p and t_0 are interpolated to fine spatial resolution respectively, through Thin Plate Spline (TPS) interpolation method (Chen et al., 2014; Zhu et al., 2016). TPS as a spatial interpolation technique for point data based on spatial dependence (Dubrule, 1984), is employed to obtain interpolation result thanks to its high accuracy. Then, another increment from the difference between interpolation results on t_p and t_0 can be acquired. As this increment only uses spatial dependence among coarse pixels, it can be referred to as the spatial-dependent increment $\Delta S(x_j, y_j)$, as shown in Eq. (4), where $F_p^{\mathrm{TPS}}(x_j, y_j)$ and $F_0^{\mathrm{TPS}}(x_j, y_j)$ are TPS interpolated values on t_p and t_0 respectively, and (x_j, y_j) is the jth fine pixel within the coarse pixel (x, y).

$$\Delta S(x_j, y_j) = F_p^{\mathrm{TPS}}(x_j, y_j) - F_0^{\mathrm{TPS}}(x_j, y_j) \tag{4}$$

Compared with the temporal increment, spatial-dependent increment has two advantages. First, coarse NDVI image on date t_p contains signals of land cover changes if the changes are significant enough to be recorded. By TPS interpolation, such land cover change information can be directly captured at a fine resolution. Second, spatial-dependent increment is independent of classification map and unmixing procedure, thus it has the potential to justify errors in the temporal increment resulted from classification or unmixing. In this study, TPS is used to estimate spatial-dependent increment rather than estimate the NDVI value on t_p, a strategy used in FADAF, because the increment reveals the changes of NDVI directly. Zhang et al. (2015) also suggested that using increment yields higher accuracy than predicting the value directly at t_p. The use of this spatial-dependent increment will be further discussed in **Section 5**.

2.3 Combination of two increments by CLS

The abovementioned two increments can be considered to be two independent predictions by two different models. Due to the distinct features used by the two predictions, the former uses the information of temporal changes of NDVI, and the later mainly utilizes the spatial dependence. Their prediction accuracies should be different under different scenarios and spatial-dependencies. Therefore, it is natural to expect 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 temporal increment (ΔT) and spatial-dependent increment (ΔS) should be summing them by reasonable weights. Moreover, an ideal combination should be as close to the true fine NDVI increment (ΔF) as possible. Thus, an objective function of weighted combination can be written as,

$$(\hat{w}_{\mathrm{S}}, \hat{w}_{\mathrm{T}}) = \arg \min_{(w_{\mathrm{S}}, w_{\mathrm{T}}) \in (0,1)} \sum_k \left(w_{\mathrm{S}} \Delta S_k + w_{\mathrm{T}} \Delta T_k - \Delta F_k \right)^2 , \tag{5}$$

where ΔS_k, ΔT_k, and ΔF_k are the spatial-dependent increment, the temporal increment, and the true increment of the kth fine pixel, respectively. w_{S} and w_{T} are weights of the spatial-dependent increment and the temporal increment, respectively. Eq. (5) can be solved by the Constrained Least Square (CLS) method, with constraints of w_S and w_T being nonnegative and summing up to one.

However, as fine NDVI values on t_p are unknown, it is impossible to obtain the true fine increment (ΔF). Fortunately, a real NDVI increment of a coarse pixel (ΔC) from t_0 to t_p is available because coarse observations are available on two dates. Therefore, both the temporal increment and the spatial-dependent increment are up-scaled to the resolution of coarse pixel (ΔC^{T} and ΔC^{S}), as shown in Fig. 2. Then, w_S and w_T in Eq. (5) can be obtained by solving Eq. (6) alternatively:

$$(\hat{w}_{\mathrm{S}}, \hat{w}_{\mathrm{T}}) = \arg \min_{(w_{\mathrm{S}}, w_{\mathrm{T}}) \in (0,1)} \sum_k \left(w_{\mathrm{S}} \Delta C_k^{\mathrm{S}} + w_{\mathrm{T}} \Delta C_k^{\mathrm{T}} - \Delta C_k \right)^2 \tag{6}$$

where ΔC_k^{S}, ΔC_k^{T} and ΔC_k are up-scaled spatial-dependent increment, up-scaled temporal increment and true increment of kth coarse pixel, respectively. Here, the average value of all fine NDVI pixels within the coarse pixel are used to produce up-scaled spatial increment (ΔC^{S}) and up-scaled temporal increment (ΔC^{T}), and ΔC_k is calculated as difference between coarse NDVI values on prediction date t_p and t_0. Considering that the weights w_S and w_T are spatially-dependent, Eq. (6) is solved in a 7×7 moving window at a coarse resolution corresponding to the window size of the unmixing process. Then, with the estimated w_S and w_T, the final fine increment can be calculated as following:

$$\Delta F^{\mathrm{Com}}(x_j, y_j) = w_{\mathrm{S}} \times \Delta S(x_j, y_j) + w_{\mathrm{T}} \times \Delta T(x_j, y_j), \tag{7}$$

where $\Delta F^{\mathrm{Com}}(x_j, y_j)$ is the combined increment of fine pixel (x_j, y_j). w_S and w_T are supposed to be scale-invariant and its rationality will be discussed **Section 5**.

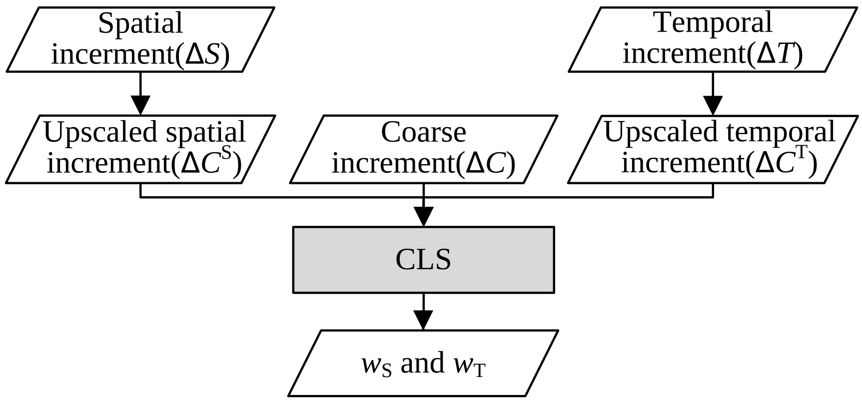

Fig. 2. Illustration of weighted calibration based on Constrained Least Square (CLS) method.

2.4 Distribution of residuals

After CLS optimization, the combined increment can capture most of the fine NDVI increment. However, residuals are inevitable even though they are minimal. The residuals can be mathematically expressed as Eq. (8),

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

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. In order to further improve the accuracy of the combined increment, residual derived above needs to be allocated to each fine pixel (x_j, y_j) within the coarse pixel (x, y). Because the residuals are minimal after the weighted combination of two increments, they can be distributed equally (Chen et al., 2014) as Eq. (9).

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

where $F_0(x_j, y_j)$ is 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 residuals distribution, a smoothing process based on similar pixels (Zhu et al., 2016) is applied to remove block effects in the fused image.

2.5 Combination of multi-time predictions

Through **Sections 2.1** to **2.4**, a prediction $\hat{F}_{0,p}$ for date t_p based on the fine NDVI on t_0 can be acquired. 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 $p+i$ (i=…, -3, -2, -1, 1, 2, 3, …) in other partially clouded fine NDVI images. Recognition of a pixel is either clear or clouded can be performed based on the Fmask algorithm (Zhu and Woodcock, 2012). Generally, the predictions with a base date too far from t_p are excluded considering that the base NDVI images hold weak relationship with the NDVI image on date t_p. Operationally, the maximum interval between the base date and the prediction date is set as two months. Then, the NDVI difference of coarse pixels between the base date and the prediction date is used to calculate the contribution of each prediction, as shown in Eq. (10).

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

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

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

If $C_q^i(x,y)$ equals $C_p^i(x,y)$, $\hat{F}_p(x_j,y_j)$ will be set as $\hat{F}_{q,p}(x_j,y_j)$ since $w_{q,p}(x, y)$ is infinite under this situation. Finally, for each prediction date in the time series, a final prediction in Eq. (11) can be obtained using the same routine described in **Sections 2.1 through 2.5**.

To assess the performance of the new method, four accuracy indices, Root Mean Square Error (RMSE), relative RMSE (RMSE divided by averaged observation value), 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).

3. Data

3.1 Data for experiments using single cloud-free fine image

We used Landsat images without clouded pixels to evaluate the performance of the proposed IFSDAF model at two sites with different land-cover characteristics. Considering that the performance of the existing spatiotemporal fusion methods generally perform well in homogeneous areas (Zhu et al., 2018), the study only tests the performance of the new method in cases with relative complexities (i.e., a heterogeneous site and a site with significant land cover changes.) The Landsat images covering the two sites were shared by Emelyanova et al. (2013) and were also used to test the NDVI-LMGM and FSDAF algorithms (Rao et al., 2015; Zhu et al., 2016).

This first site is located in the Coleambally irrigated area (34°54′S and 145°57′E), characterized by great heterogeneity in landscape with many small patches of farm land and rapid phenological changes (Fig. 3). Two Landsat ETM+ images (800×800 pixels), acquired on November 25th, 2001 (t_0) and January 12th, 2002 (t_p) during the growing season, were upscaled by the ratio of 8:1 to synthesize MODIS images. In this test, the synthesized MODIS image instead of the real MODIS image was used, because the synthesized MODIS image can exclude the co-registration error (Gevaert and Garcia-Haro, 2015; Wang and Atkinson, 2018; Zhu et al., 2016). This exclusion ensures a fair comparison of different algorithms. The NDVI data were then derived from corresponding reflectance images. Then, the land cover classification map was obtained by the Iterative Self-Organizing Data Analysis Technique (ISODATA) method based on the Landsat image acquired on November 25th, 2001 (t_0).

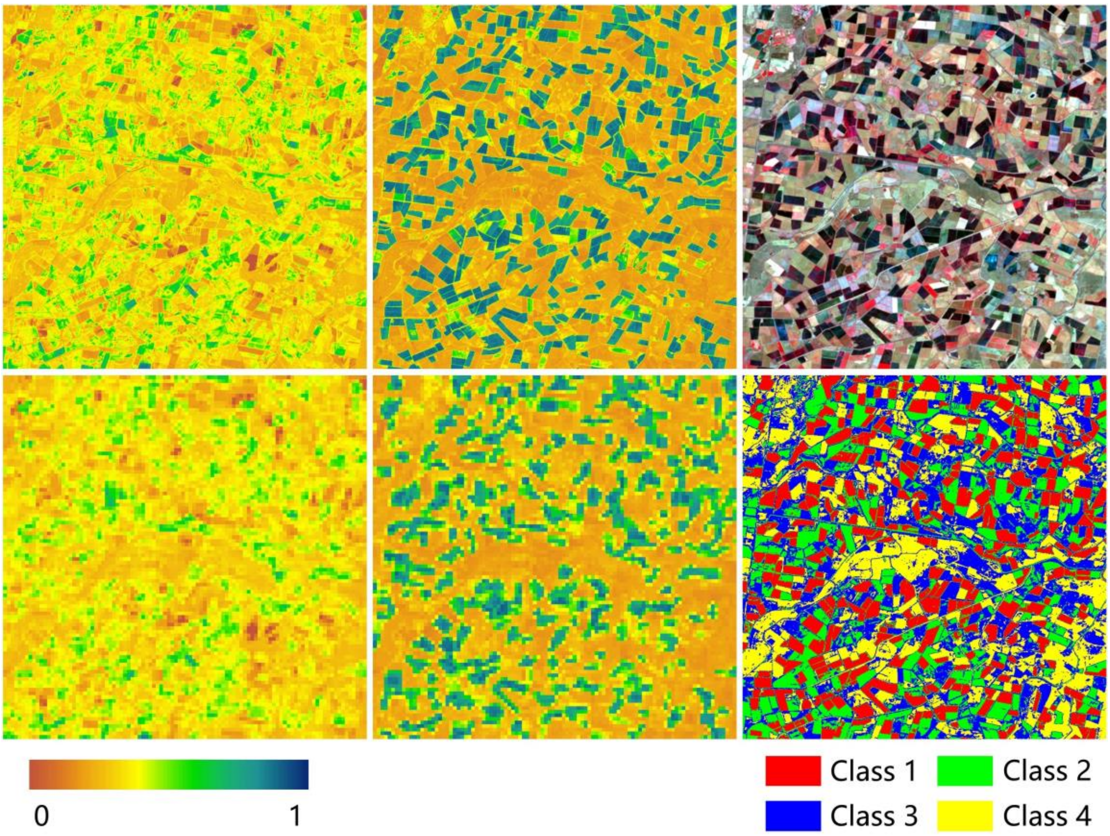

Fig.3. Test data of the heterogeneous site in Coleambally irrigation area: Landsat

NDVI on November 25th, 2001 (a) and January 12th, 2002 (b); false-color-composite Landsat image on November 25th, 2001 (c); MODIS NDVI on November 25th, 2001 (d) and January 12th, 2002 (e); and land cover map on November 25th, 2001 by ISODATA (f).

The second site is located in the Gwydir area (29°07′S and 149°04′E) with a flood event occurred in December 2004. Two Landsat TM images (800×800 pixels) on November 26th, 2004 (t_0) and December 12th, 2004 (t_p) were used at this site (Fig. 4). Abrupt land cover changes can be observed in these two images due to the flood (Emelyanova et al., 2013). Two Landsat images were also upscaled by the ratio of 8:1 to synthesize the MODIS images. Then, the NDVI data were derived from all the original images. A land cover classification map was obtained based on Landsat image on November 26th, 2004 (t_0) by the ISODATA method.

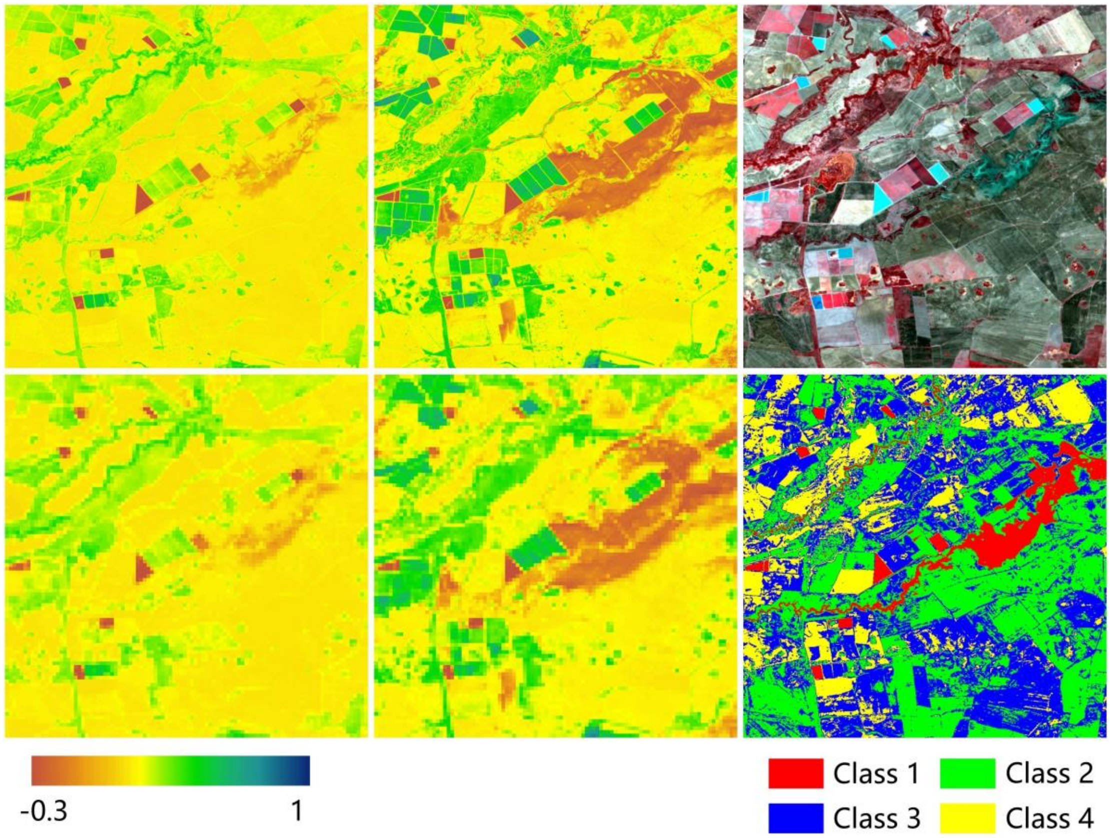

Fig.4. Test data of a site experienced land cover change in Gwydir area: Landsat NDVI on November 26th, 2004 (a) and December 12th, 2004 (b); false-color-composite Landsat image on November 26th, 2004 (c); MODIS NDVI on November 26th, 2004 (d) and December 12th, 2004 (e); and classification map on November 26th, 2004 by ISODATA (f).

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 data set for comparison.

3.2 Data for experiments using multiple clouded fine images

Experiments using multiple cloudy fine images were implemented to assess the performance of the proposed IFSDAF method for predicting the NDVI time series when the input fine images which were partially 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 in two cloudy sites respectively.

The first site is Shennongjia Forestry District (109°59'–110°58' E, 31°15'–31°57' N) located in the western part of Hubei Province, central China (Fig. 5). This area belongs to subtropical monsoon climate; and its elevation ranges from 398 to 3105 m. The vegetation distribution of this area is very heterogeneous with the types of evergreen broadleaf forest, deciduous broadleaf forests, and evergreen coniferous forest. There are also farmlands and artificial surfaces in this area (Wang et al., 2018; Zhao et al., 2005). The 250 m (resampled to 240 m) 16-day composite MODIS NDVI products (MOD13Q1) covering this site in 2015 were acquired from NASA (https://ladsweb.nascom.nasa.gov/search/) and then resampled to the resolution of 240 m. Landsat 8 level 2A surface reflectance products and their cloud masks by Fmasks in 2015 were downloaded from the USGS (https://espa.cr.usgs.gov/ordering/new/). All Landsat images were co-registered to MODIS images. M osaic of two adjacent Landsat-8 scenes can cover the whole area of this site. When mosaicking two Landsat 8 images with close acquisition dates, pixels in the overlapped part have two NDVI values and the higher one is kept because higher NDVI is less likely affected by poor atmospheric condition. Those Landsat images with clouds, shadows, and snow more than 70% were discarded. Finally, one clear Landsat image (t_0, on October 14th, 2015) and nine partially contaminated Landsat images (Fig. 6) were selected as the input of fine-resolution NDVI images for data fusion.

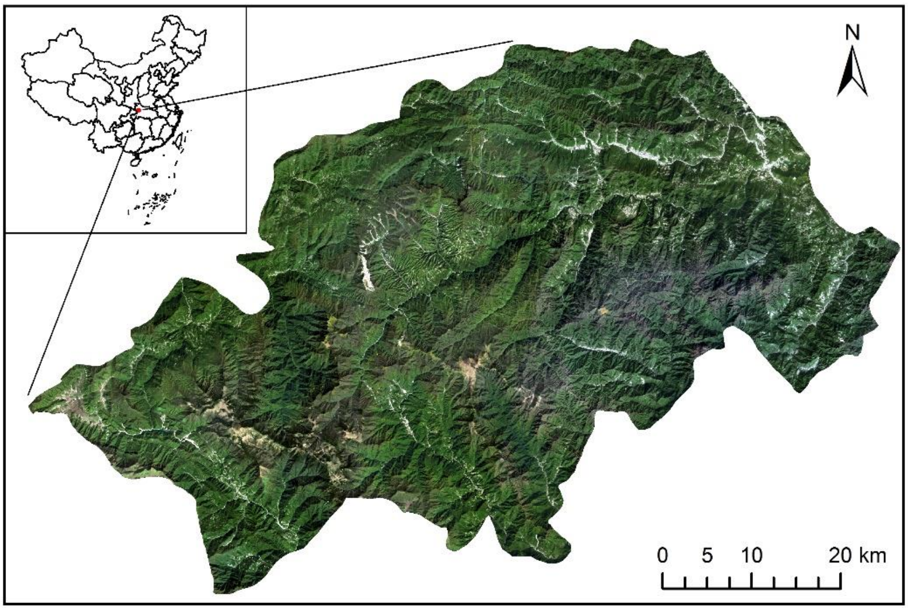

Fig. 5. Shennongjia Forestry District in Hubei Province, central China. The image is a true-color-composite Landsat 8 OLI image acquired on the day of year 287, October 14^{th}, 2015.

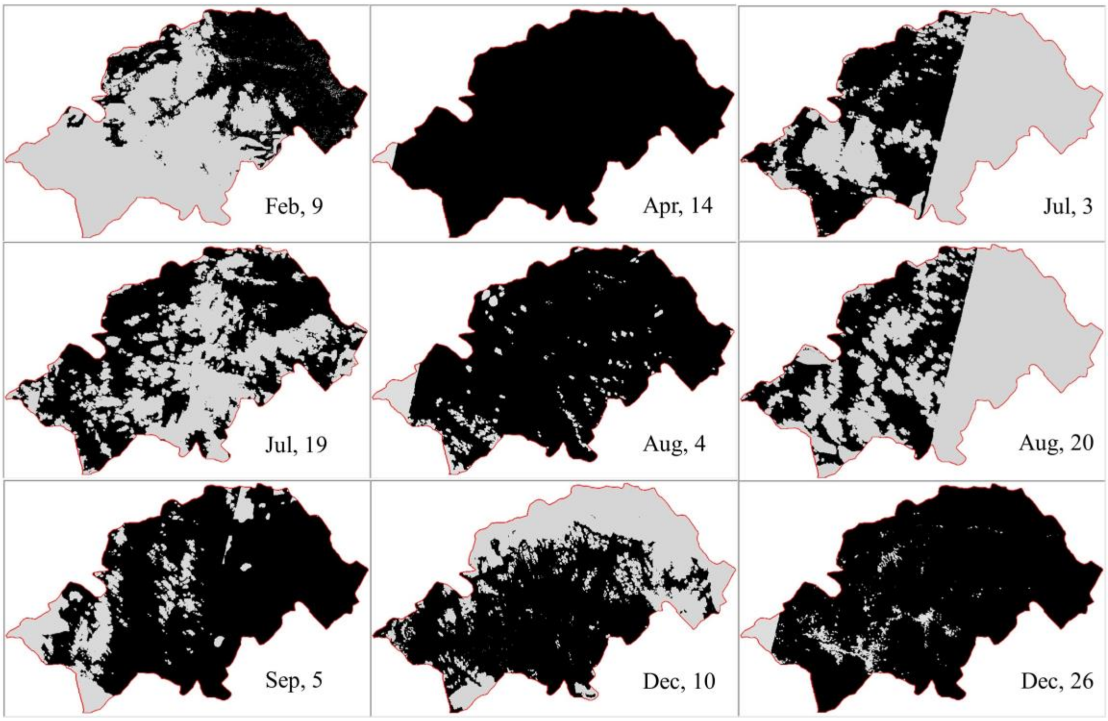

Fig. 6. Cloud masks of nine partially contaminated Landsat 8 images in Shennongjia Forestry District by Fmask method, where gray color indicates pixels contaminated by clouds and cloud shadows, and black color represents clear pixels.

The second site is in Southeast Asia which has complex landscapes with croplands, water, forest and urbans. This site is covered by one Sentinel-2A scene (size 10980×10980 10m pixels) with the tile number of T48QUD. We acquired Sentinel-2A satellite level 1C products from EarthExplorer (https://earthexplorer.usgs.gov/) and 8-day composite MODIS surface reflectance products (MOD09Q1) from NASA. Both images were acquired in 2017. Atmospheric correction of Sentinel-2A images was done with the tool provided by European Space Agency, Sen2Cor (http://step.esa.int/main/third-party-plugins-2/sen2cor/). Cloud masks of sentinel-2A images were produced by the Fmask software (https://github.com/gersl/fmask) and images with cloud cover more than 70% were discarded. Finally, we obtained three clear Sentinel-2A images on Feb. 13th, Mar. 5th and Dec. 20th in 2017 respectively (Fig. 7), and 21 partially cloud contaminated images (Fig. 8). Sentinel-2A NDVI and MODIS NDVI images were then calculated from the surface reflectance data.

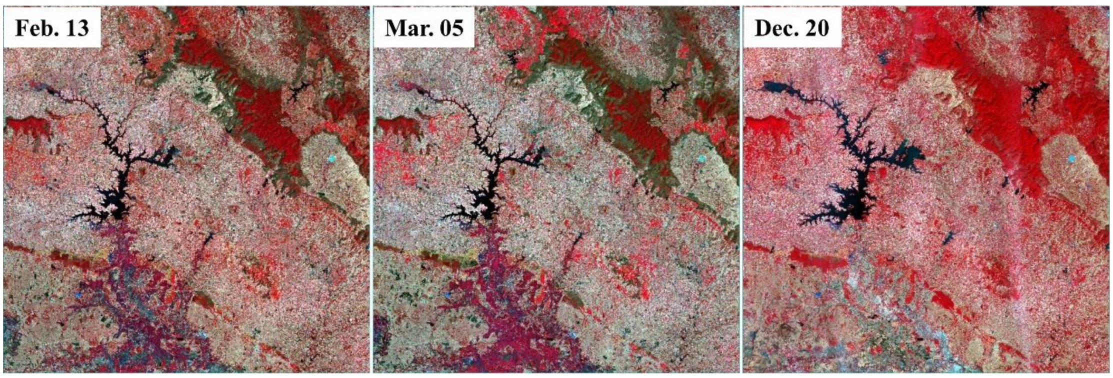

Fig. 7. False-color-composite of clear Sentinel 2A images on Feb. 13th, Mar. 05th and Nov. 20th in 2017 with the tile number of T48QUD, respectively.

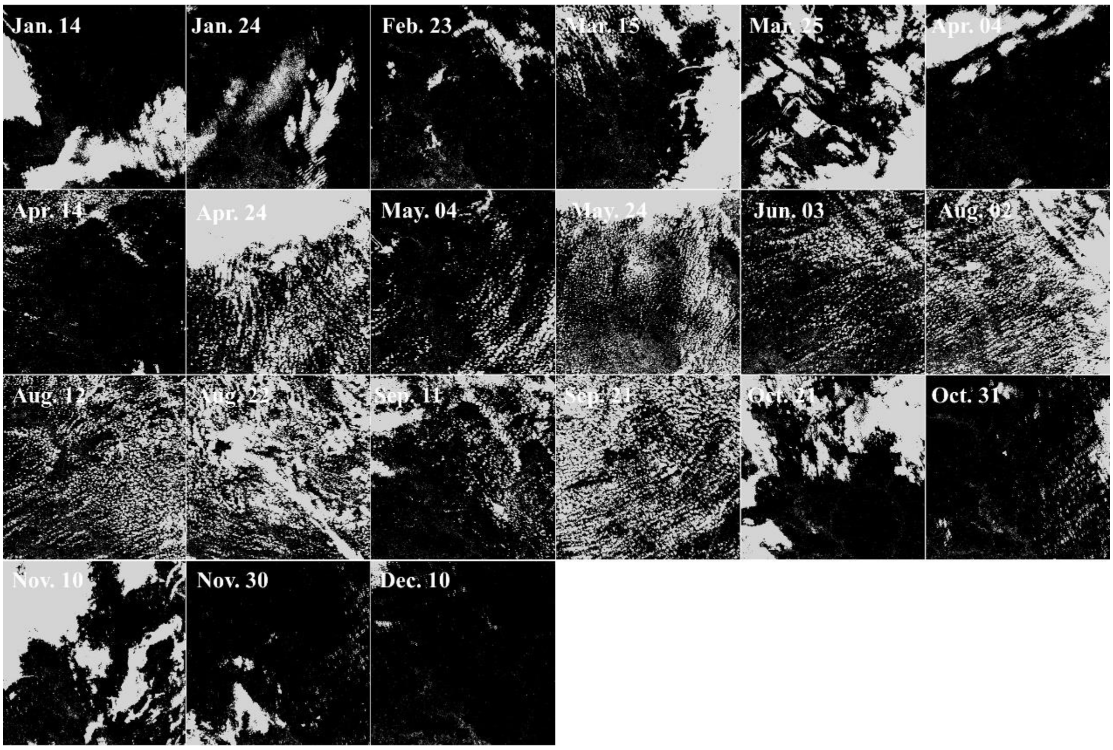

Fig. 8. Cloud masks of 21 partially cloud contaminated Sentinel-2A images with the tile number of T48QUD by Fmask method, where black color is clear pixel and gray color is cloudy pixel.

4. Results

4.1 Fusion using single cloud-free fine image

Fig. 9 shows the visual comparison of predicted Landsat NDVI by IFSDAF and the three existing methods with observed Landsat NDVI on January 12th, 2002 (t_p) for the farmland site with great heterogeneity and rapid phenological changes. Compared with the other three methods, the fused image by IFSDAF (Fig. 9d) is more similar to the actual NDVI image (Fig. 9e) (e.g., the zoomed-in sub-region). On the contrary, the NDVI-LMGM (Fig.9a) and FSDAF (Fig.9c) methods yield large errors in some pixels leading to discontinuity in the fused images; and STARFM (Fig.9b) leads to an unsatisfactory blurring effect for small objects. Scatter plots (Fig. 10) and quantitative assessment also confirms that the proposed method obtains the highest accuracy (Table 1). The IFSDAF has the lowest RMSE (0.0884), lowest rRMSE (22.12%) and highest r (0.9376) for the whole image. Furthermore, the AD (-0.0001) of the newly proposed method is closer to zero, indicating less biased. In addition, the accuracies of the NDVI-LMGM (RMSE= 0.1300 and rRMSE= 32.54%) and STARFM (RMSE= 0.1646 and rRMSE= 41.19%) are much lower for the whole image compared with the FSDAF (RMSE= 0.1002 and rRMSE= 25.06%).

As we mentioned, the NDVI normally has a larger variance than raw reflectance bands. Fig. 11a shows histogram distributions of bands Red, NIR, and corresponding NDVI from Landsat image on January 12th, 2002 (t_p). The NDVI displays two peaks with significantly a greater variance than that of band Red or band NIR due to 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 magnitudes, based on histogram distribution of NDVI (Fig.11a), the whole image is thus divided into three parts: low NDVI (< 0.4), medium NDVI (0.4-0.7), and high NDVI (>0.7). It can be seen that NDVI-LMGM and STARFM have relatively lower accuracies compared with IFSDAF and FSDAF. And IFSDAF has a better performance than FSDAF in medium NDVI and high NDVI sections.

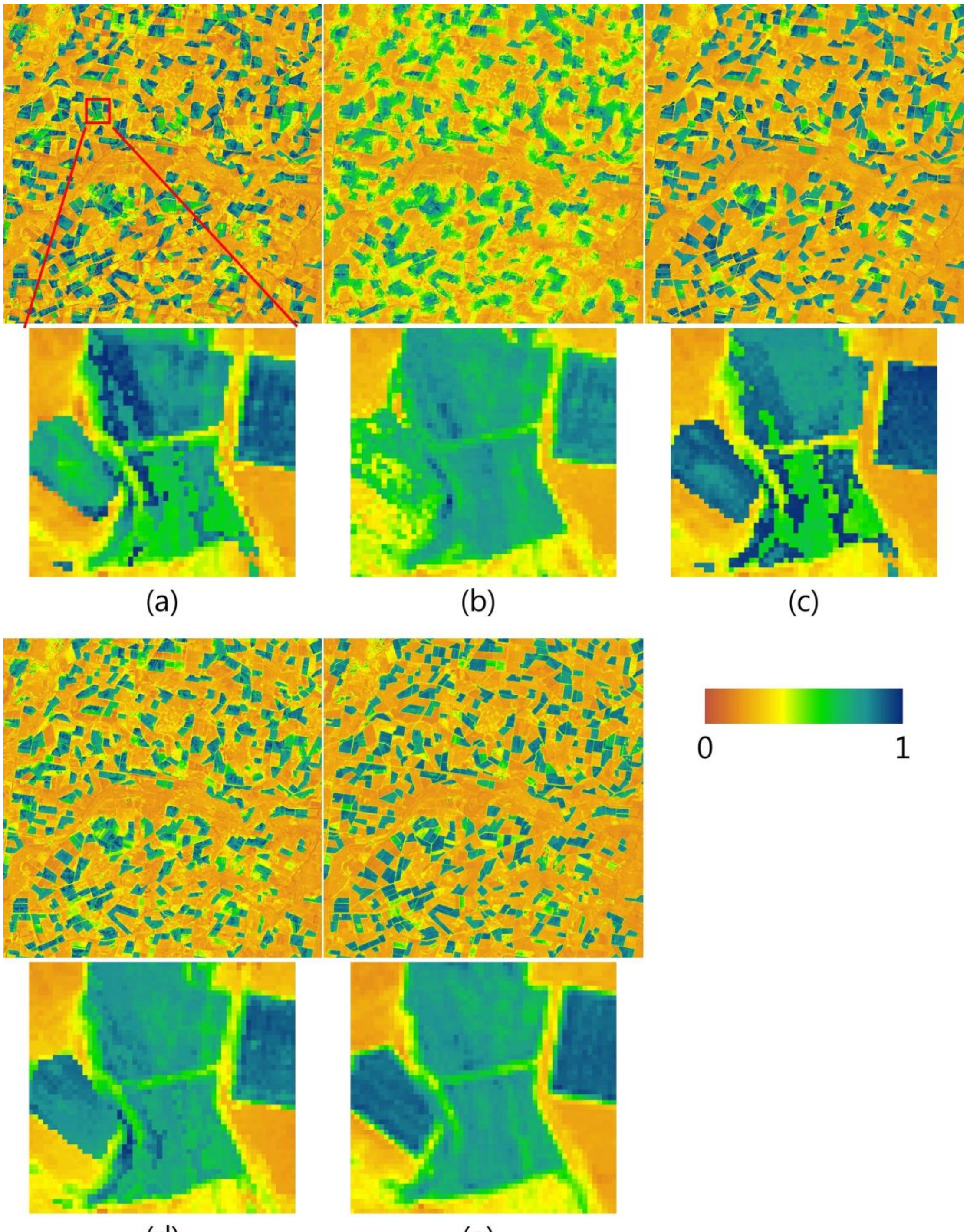

Fig. 9. Landsat NDVI on January 12th, 2002: predictions by NDVI-LMGM (a), STARFM (b), FSDAF (c), IFSDAF (d) and the actual NDVI (e).

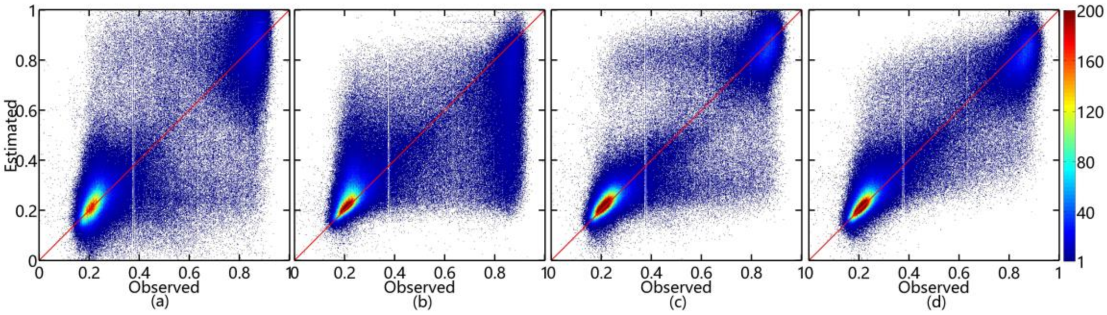

Fig. 10. Scatter plots of estimated results compared with observed value of Landsat NDVI on January 12th, 2002: NDVI-LMGM (a), STARFM (b), FSDAF (c) and IFSDAF (d).

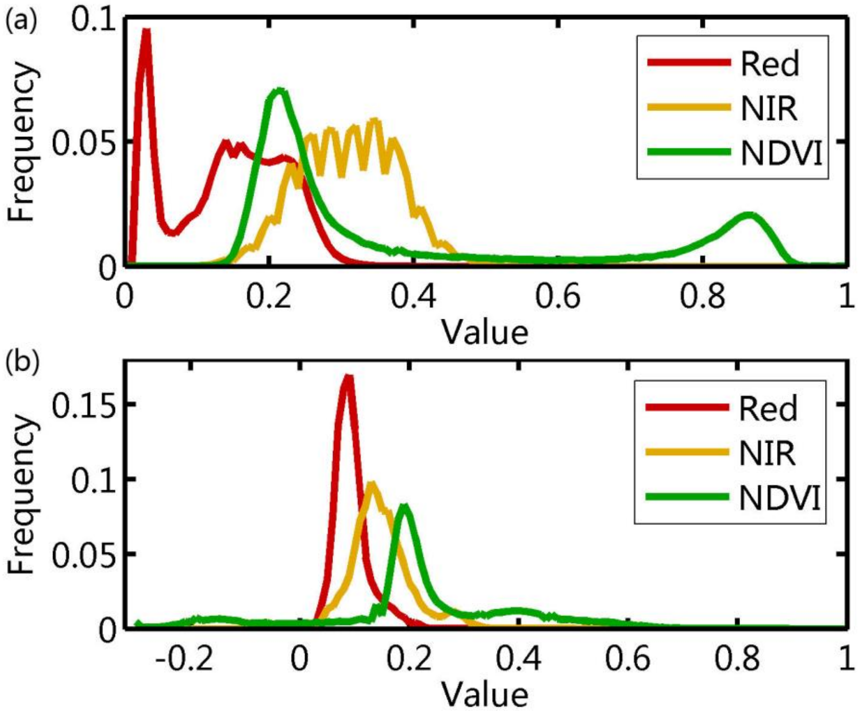

Fig. 11. Histograms of band Red, band NIR and NDVI in Coleambally irrigation area on January 12th, 2002 (a) and the Gwydir area on December 12th, 2004 (b).

Table.1. RMSE, rRMSE, *r* and AD between predicted NDVI and observed NDVI of NDVI-LMGM, STARFM, FSDAF and the IFSDAF method in the Coleambally irrigation area.

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	**0.0148**	0.1068	45.29%	0.5214	0.0564	0.0669	28.36%	0.5576	0.0160	**0.0664*****	**28.16%**	**0.6334**	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	**0.1473*****	**27.51%**	**0.4328**	**-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	**0.1116*****	**13.36%**	**0.4493**	**-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	**0.0884*****	**22.12%**	**0.9376**	**-0.0001**

Note: for t test, * means $p < 0.05$; ** means $p < 0.01$; *** means $p < 0.001$ compared with results of FSDAF.

For the Gwydir site where a flood event occurred, as shown in Fig. 12, the fusion result of IFSDAF (RMSE = 0.0546) captures the change (Fig. 12d), being more similar to the actual NDVI pattern than the other three methods. NDVI-LMGM gets fused image (RMSE = 0.0794) with significant block effects (Fig. 12a). The result of STARFM (RMSE = 0.0686) is generally similar to the actual NDVI image as shown in Fig. 12b. FSDAF also has high accuracy (RMSE= 0.0617) in fusion but it has abnormal predictions for some pixels shown in the selected area (Fig. 12c). The blue arrow in Fig. 12c indicates the error edges produced by FSDAF. In fact, before the big flood, there is a small river in the zoomed-in area, resulting in the edge (marked by the blue arrow) between water and barren land. However, after the flood, the river overflowed and covered nearby farmland. Thus, the original edge of the river disappeared as shown in the actual Landsat NDVI image of Fig. 12e. IFSDAF is the only among the four methods which can capture this phenomenon. Scatter plots in Fig. 13a-d show no obvious bias of these four methods; but points of FSDAF and IFSDAF are closer to the 1:1 line than the other two methods which reveal the comparable capacity of both IFSDAF and FSDAF in capturing land cover changes.

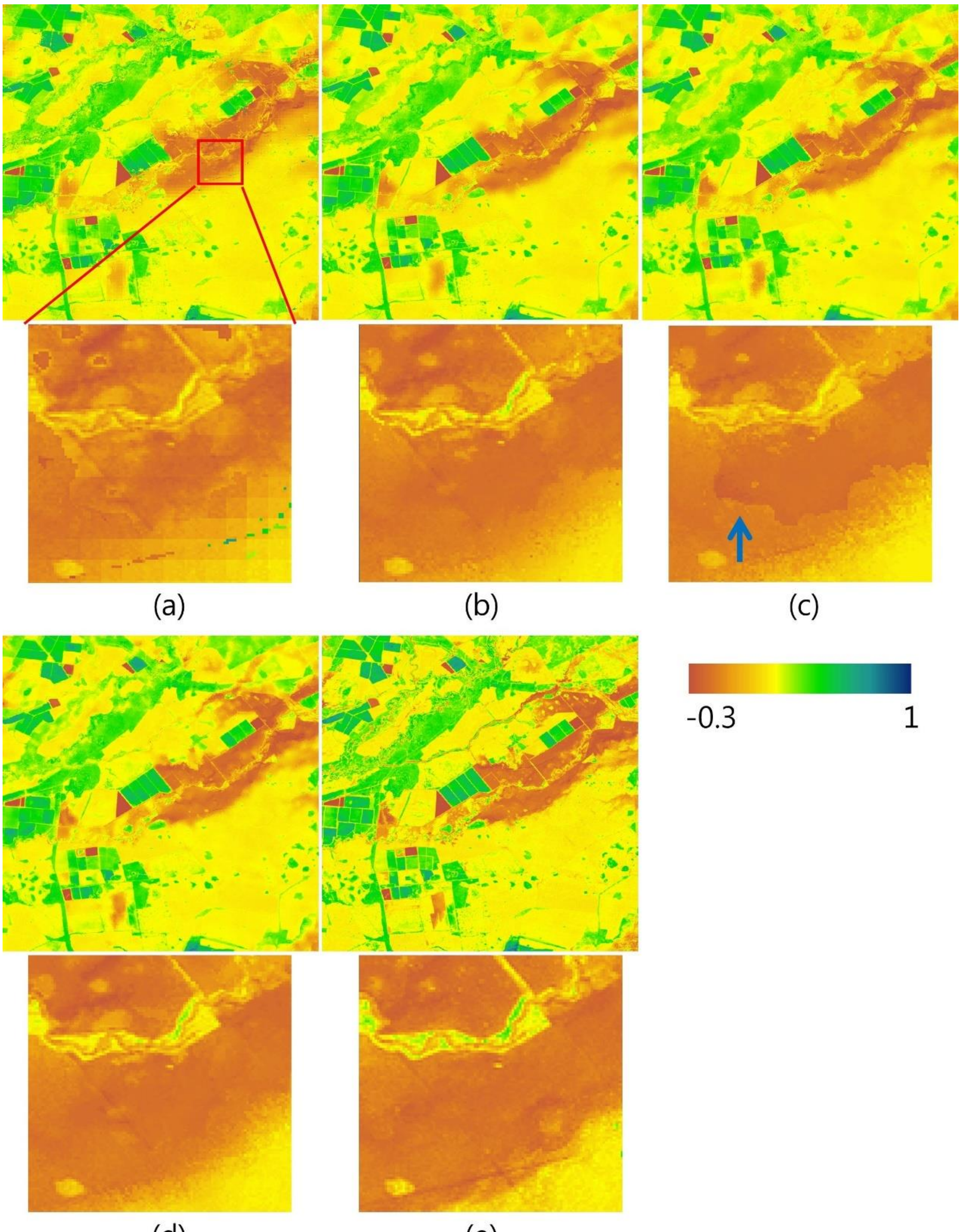

Fig. 12. Landsat NDVI on December 12^{th}, 2004: predictions by NDVI-LMGM (a), STARFM (b), FSDAF (c), IFSDAF (d) and the actual NDVI (e).

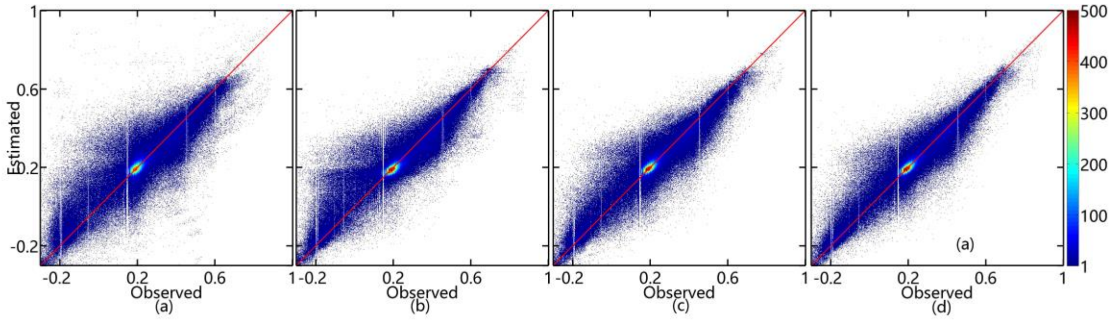

Fig. 13. Scatter plots of estimated results compared with observed value of Landsat NDVI on December 12th, 2004: NDVI-LMGM (a), STARFM (b), FSDAF (c) and IFSDAF (d).

Fig. 11b shows histogram distributions of band Red, band NIR, and NDVI on December 12th, 2004 (t_p) respectively. The variance of NDVI is also significantly higher than that of band Red or band NIR. Moreover, due to the flood event, there are many negative values in NDVI, causing three peaks around NDVI = -0.2, NDVI = 0.2 and NDVI = 0.4 in the histogram distribution of NDVI. The whole image is divided into three parts (low NDVI < 0, medium NDVI 0-0.3, and high NDVI > 0.3) to quantitatively assess the accuracy (Table 2). It is clear that the new method yields higher accuracy with lower RMSE of 0.0546, higher r of 0.9527 among the whole image than the other three methods. In the three separate NDVI parts, the new method also displays higher accuracy with RMSE = 0.0798, 0.0467, and 0.0584, respectively.

Table.2. RMSE, rRMSE, *r* and AD between predicted NDVI and observed NDVI of NDVI-LMGM, STARFM, FSDAF and the IFSDAF method in the Gwydir area.

Methods	NDVI-LMGM				STARFM				FSDAF				IFSDAF			
	RMSE	rRMSE	*r*	AD	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	**0.0798*****	**-56.42%**	**0.7821**	**0.0337**
Medium NDVI	0.0636	33.46%	0.4872	0.0061	0.0526	27.65%	0.5794	**0.0042**	0.0516	27.11%	0.6040	0.0066	**0.0467*****	**24.55%**	**0.6524**	0.0045
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	**0.0584*****	**13.38%**	**0.8658**	**-0.0270**
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	**0.0546**	**25.77%**	**0.9527**	**0.0001**

Note: for t test, * means $p < 0.05$; ** means $p < 0.01$; *** means $p < 0.001$ compared with results of FSDAF.

4.2 Fusion using multiple fine images partially covered by clouds

In the site of Shennongjia, each of the four Landsat NDVI images captured on Apr 14th, Jul 3th, Sep 5th, and Dec 10th in 2015 serves as reference data for the independent validation. For example, Apr 14th was predicted by IFSDAF using all other eight partially contaminated fine NDVI images as input, and then clear pixels in the true Apr 14th image were used to assess the accuracy of predicted Apr 14th image. For the comparison, FSDAF only used one clear Landsat NDVI image on Oct. 14th of 2015 to predict the above four Landsat NDVI images. The accuracies of fusion results of the four images are summarized in Table 3 and the predictions are shown in Fig. 14. For the purpose of simplification, results of NDVI-LMGM and STARFM are not shown in this experiment because they yielded lower accurate results than FSDAF.

It is evident from Fig. 14 that IFSDAF can produce fused images more similar with real Landsat NDVI than FSDAF. In Table 3, RMSE values of IFSDAF on all dates are lower than that of FSDAF. These improvements of accuracy are mainly attributed to the extra information provided by the partially contaminated Landsat images, which can be well used in IFSDAF but not in FSDAF. On the contrary, FSDAF only used one fine image on Oct 14th in 2015 which is far away from some prediction dates, leading to low accuracy on these prediction dates. More important, the improvement of IFSDAF on Jul 3th and Sep 5th during the peak stage of vegetation growth is more significant than other two dates, indicating that IFSDAF may be more effective for fusing images with medium to high NDVI values. This result is similar to the experiment in the Coleambally irrigation area.

Table.3. RMSE, rRMSE, *r* and AD between the predicted NDVI and observed partially contaminated fine NDVI on Apr 14^{th}, Jul 3^{th}, Sep 5^{th} and Dec 10^{th} in year 2015, in the Shennongjia Forestry District.

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

Note: for t test, * means $p < 0.05$; ** means $p < 0.01$; *** means $p < 0.001$ compared with results of FSDAF.

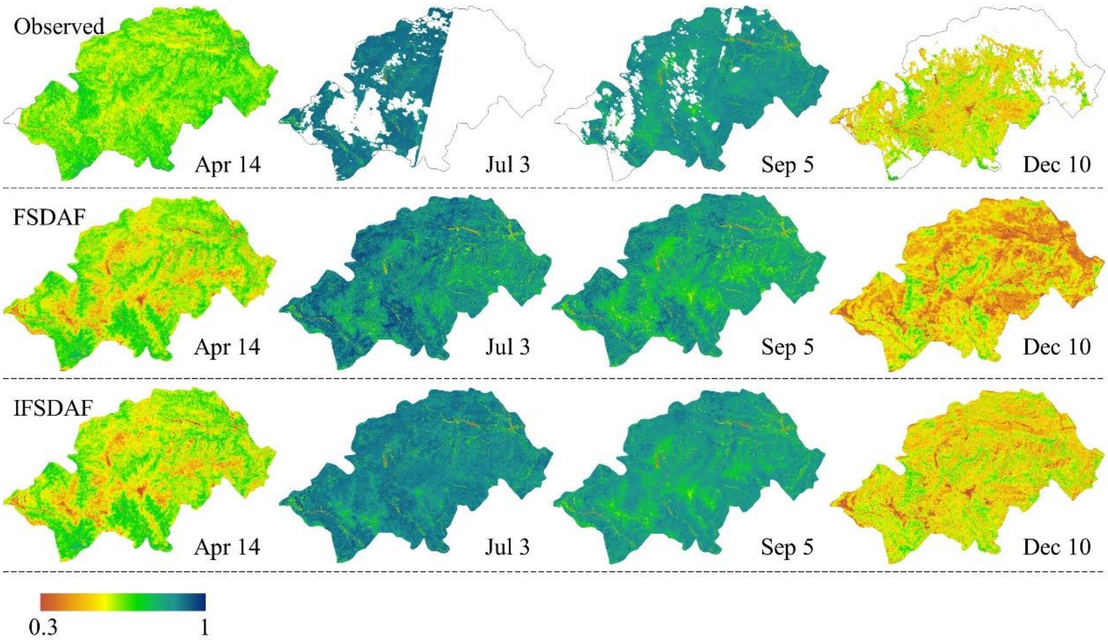

Fig. 14. Landsat 8 NDVI in Shennongjia Forestry District on Apr 14th, Jul 3th, Sep 5th, and Dec 10th in year 2015 predicted by FSDAF and IFSDAF, respectively.

In the South Asia site, clear Sentinel-2A NDVI image on Mar 5th, 2017 was selected as the base image. The other two clear Sentinel NDVI images (Feb. 12th and Dec. 20th) were used as reference data to assess the accuracy of IFSDAF and FSDAF. The base fine spatial resolution NDVI image and the 21 partially cloud contaminated fine NDVI images were used as input for IFSDAF, while only the base fine NDVI image was input to FSDAF. Results in Table. 4 shows that IFSDAF produces more accurate predictions with lower RMSE in both dates (0.0863 and 0.0740) compared with the results by using FSDAF (0.0999 and 0.1469).

Table. 4. RMSE, rRMSE, *r* and AD between the predicted NDVI and observed fine NDVI on Feb. 12th and Dec. 20th, 2017 with Sentinel-2A data.

Date	Methods	RMSE	rRMSE	*r*	AD

Feb 13th	FSDAF	0.0999	24.04%	0.8885	-0.0463
	IFSDAF	**0.0863*** **	**20.76%**	**0.9305**	**-0.0427**
Dec 20th	FSDAF	0.1469	33.69%	0.7401	**-0.0082**
	IFSDAF	**0.0740*** **	**17.69%**	**0.9584**	-0.0141

Note: for t test, * means $p < 0.05$; ** means $p < 0.01$; *** means $p < 0.001$ compared with results of FSDAF.

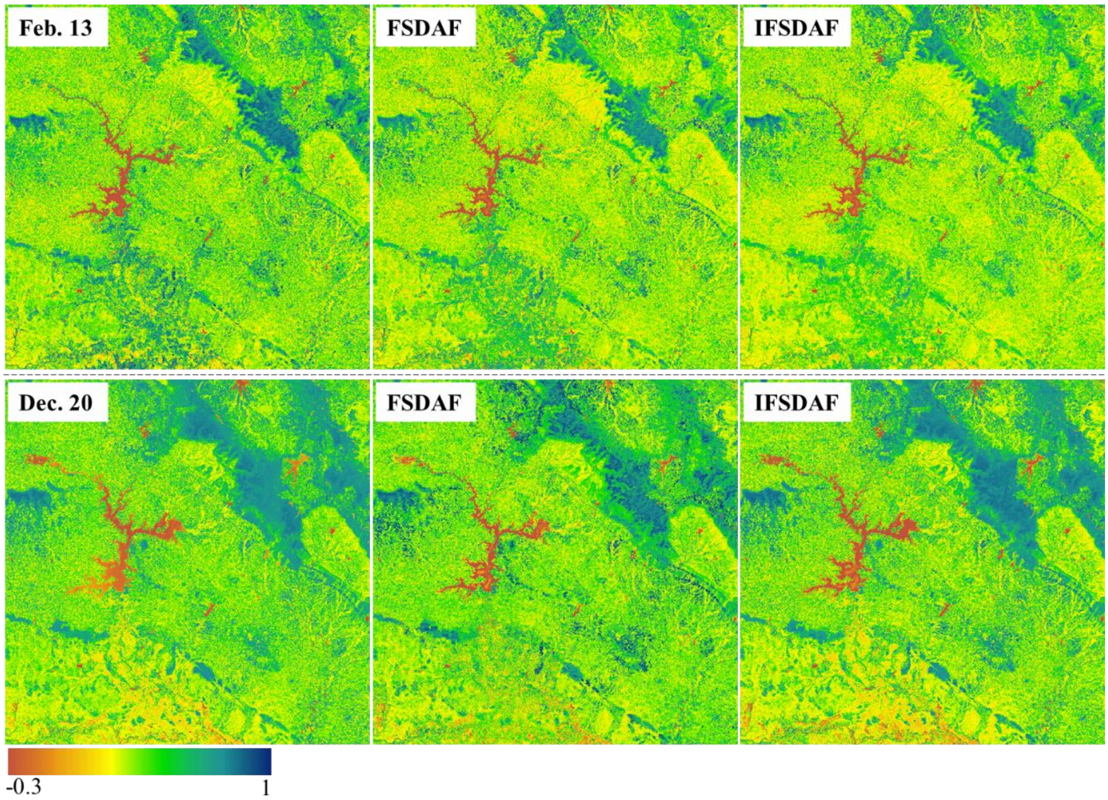

Fig. 15. Sentinel-2A NDVI images on Feb. 13th and Dec. 20th (left) and the results predicted by FSDAF (middle) and IFSDAF (right), respectively.

5. Discussion

5.1 The way of deriving spatial-dependent increment

In this study, the spatial-dependent increment (ΔS) is acquired based on difference between interpolation results of coarse NDVI on date t_p and date t_0, respectively, as shown in Eq. (4). However, there is also another way of obtaining ΔS in Eq. (12), since the F_0 is available on date t_0.

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

where $F_0(x_j, y_j)$ is fine NDVI value of pixel (x_j, y_j) on date t_0. However, ΔS derived from Eq. (4) is a better indicator than that from Eq. (12). A theoretical comparison of these two types of ΔS is explained as below. Eq. (9) can be simplified as shown in Equation (13), where residuals R are ignored as they are small.

$$\hat{F}_{0,p} = F_0 + \Delta F^{\mathrm{Com}} = F_0 + w_{\mathrm{S}}\Delta S + w_{\mathrm{T}}\Delta T \tag{13}$$

For simplification, the notation (x_i, y_i) is removed by replacing Eq. (9) with Eq. (13). And then, replacing F_0 by $w_{\mathrm{S}}F_0 + w_{\mathrm{T}}F_0$, as $w_{\mathrm{S}} + w_{\mathrm{T}} = 1$, specifically,

$$\hat{F}_{0,p} = w_{\mathrm{S}}(F_0 + \Delta S) + w_{\mathrm{T}}(F_0 + \Delta T) \tag{14}$$

Based on ΔS in IFSDAF as Eq. (4), Eq. (14) can be written as below,

$$\hat{F}_{0,p} = w_{\mathrm{S}}(F_0 + F_p^{\mathrm{TPS}} - F_0^{\mathrm{TPS}}) + w_{\mathrm{T}}(F_0 + \Delta T) \tag{15}$$

Based on ΔS in Eq. (12), Eq. (14) can also be written as below,

$$\begin{aligned} \hat{F}_{0,p} &= w_{\mathrm{S}}(F_0 + F_p^{\mathrm{TPS}} - F_0) + w_{\mathrm{T}}(F_0 + \Delta T) \\ &= w_{\mathrm{S}}(F_p^{\mathrm{TPS}}) + w_{\mathrm{T}}(F_0 + \Delta T) \end{aligned} \tag{16}$$

Difference between Eq. (15) and Eq. (16) is the term $F_0 - F_0^{\mathrm{TPS}}$ in Eq. (15). TPS prediction is a spatially smoothed prediction which loses spatial details to some degree. As a result, $F_0 - F_0^{\mathrm{TPS}}$ functions similarly as a high-pass modulation to model the spatial contrast at t_0. This spatial contrast is assumed relatively stable from the t_0 to t_p in several fusion models (Song and Huang, 2013; Luo et al., 2018), so $F_0 - F_0^{\mathrm{TPS}}$ in Eq. (15) can better capture spatial details in the fused image. To demonstrate the abovementioned theoretical analysis, an experiment based on these two types of ΔS was conducted in the Gwydir area. Result shows that the prediction based on ΔS derived from Eq. (4) (Fig. 16a) is more accurate than that from Eq. (12) with RMSEs being 0.0546 and 0.0596, respectively. Moreover, as the zoomed-in pictures in Fig. 16 illustrate, the prediction result using Eq. (4) (Fig. 16a) contains more spatial details (e.g., the road marked by blue arrows), providing corroborative support to the analysis.

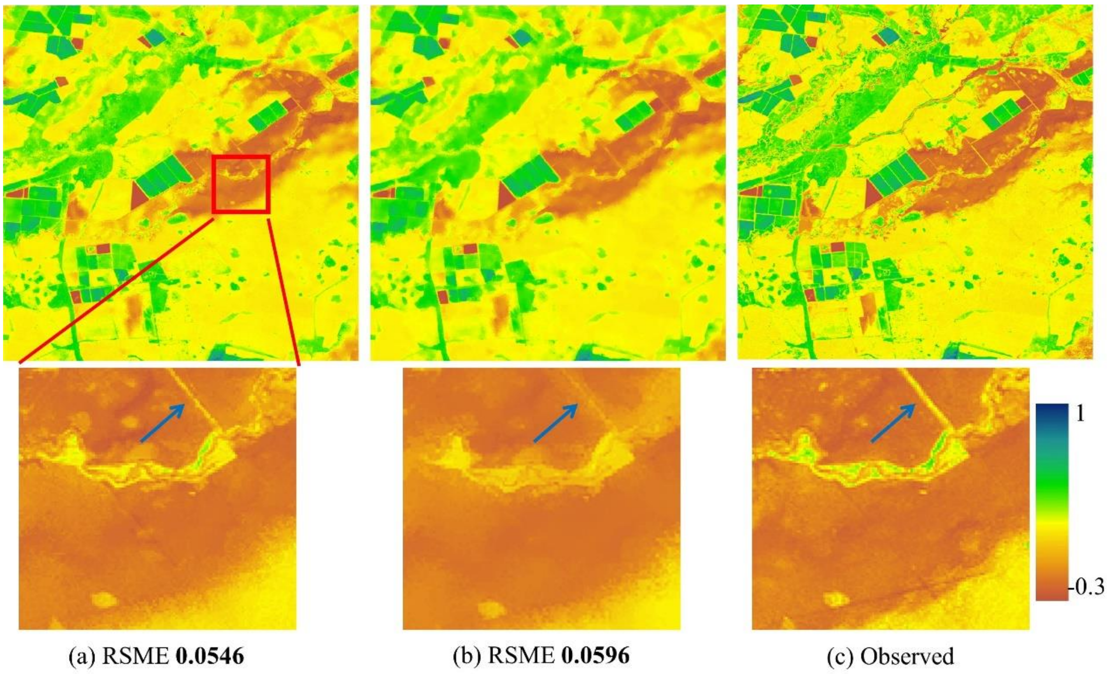

Fig. 16. Comparison of NDVI value on t_p in the Gwydir area predicted from Eq. (4) (a) Eq. (12) (b), and the real Landsat NDVI image(c).

5.2 The weights scale-invariant assumption

In the proposed IFSDAF method, spatial-dependent increment and temporal increment are combined by optimized weights. As the fine NDVI image on t_p is unknown in the real-world application, coarse NDVI increment (ΔC), upscaled spatial-dependent increment (ΔC^{S}), and upscaled temporal increment (ΔC^{T}) are used to derive w_S and w_T in Eq. (5). Such operation assumes that weight is scale-invariant. To verify the assumption, an experiment was conducted in the Coleambally irrigation area, a moving window of 7×7 at a coarse resolution was used to calculate the weights (w_S and w_T) for two increments for the center coarse pixel. Because the fine NDVI image F_p actually exists at the site, the two weights at a fine resolution can also be derived based on fine increment ($\Delta F=F_p-F_0$) using the CLS method. Fig. 17a displays the scatter plot of weights derived from the two approaches. All points are close to the 1:1 line, where x-axis represents the weight of the spatial-dependent increment at the fine resolution and y-axis represents the same weight at the coarse resolution, suggesting that the weights derived from both the coarse and fine images are substitutable. Then, combined increments calculated using the two types of weight are very similar (Fig. 17b). RMSE values of combined increments based on weights from the coarse resolution and fine resolution are 0.0941 and 0.0934, respectively. t-test shows that there is no significant difference between the two combined increments. Consequently, it can be concluded that the scale effect on the derived weights is minimal, and it will not cause significant errors on the combined increment. Thus, the assumption that weights (w_S and w_T) are scale-invariant is reasonable.

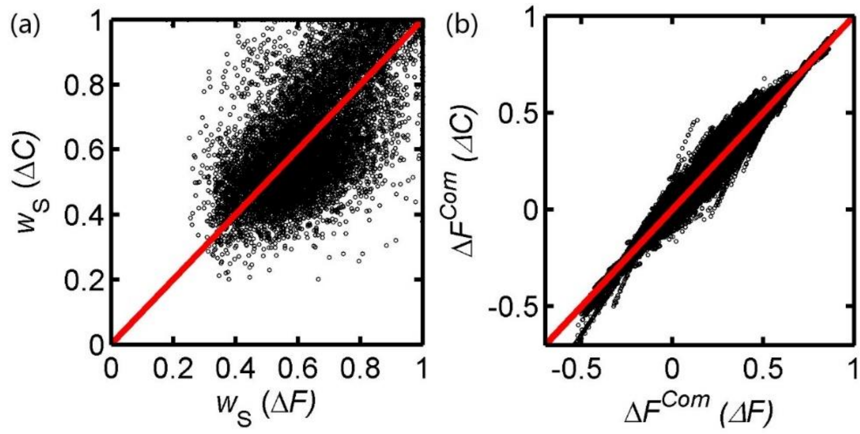

Fig.17. Scatter plot of weights of spatial-dependent increment based on (a) coarse resolution increment (Δ*C*) and fine resolution increment (Δ*F*); (b) comparison of combined increments using weights derived from coarse and fine images.

5.3 Combination of temporal and spatial-dependent increments

Temporal increment and spatial-dependent increment are combined in IFSDAF by CLS method in moving windows. Such a combination is based on the assumption that the accuracies of two increment estimations are different under different scenarios; thus the weighted combination is able to improve the accuracy of NDVI prediction through balancing biases in the estimate of two increments. We verified the assumption by comparing the temporal increment, the spatial-dependent increment and the combined increment with the real increment (Fig.18), in which RMSE was used to represent the error of estimation. Theoretically, a good combination is expected to obtain smaller RMSE value than either of the two increment estimations. As shown in Fig.18, the performance of CLS-based combination agrees with our expectation with decreased RMSE values at both study sites, demonstrating the necessity of combining two increments. Moreover, the residual of spatial-dependent increment (ΔS-ΔF) is much more similar to the residual of the combined increment (ΔF^{Com}-ΔF) than that of the temporal increment (ΔT-ΔF), suggesting that the spatial-dependent increment contributes more to the combined increment than the temporal increment at these two sites.

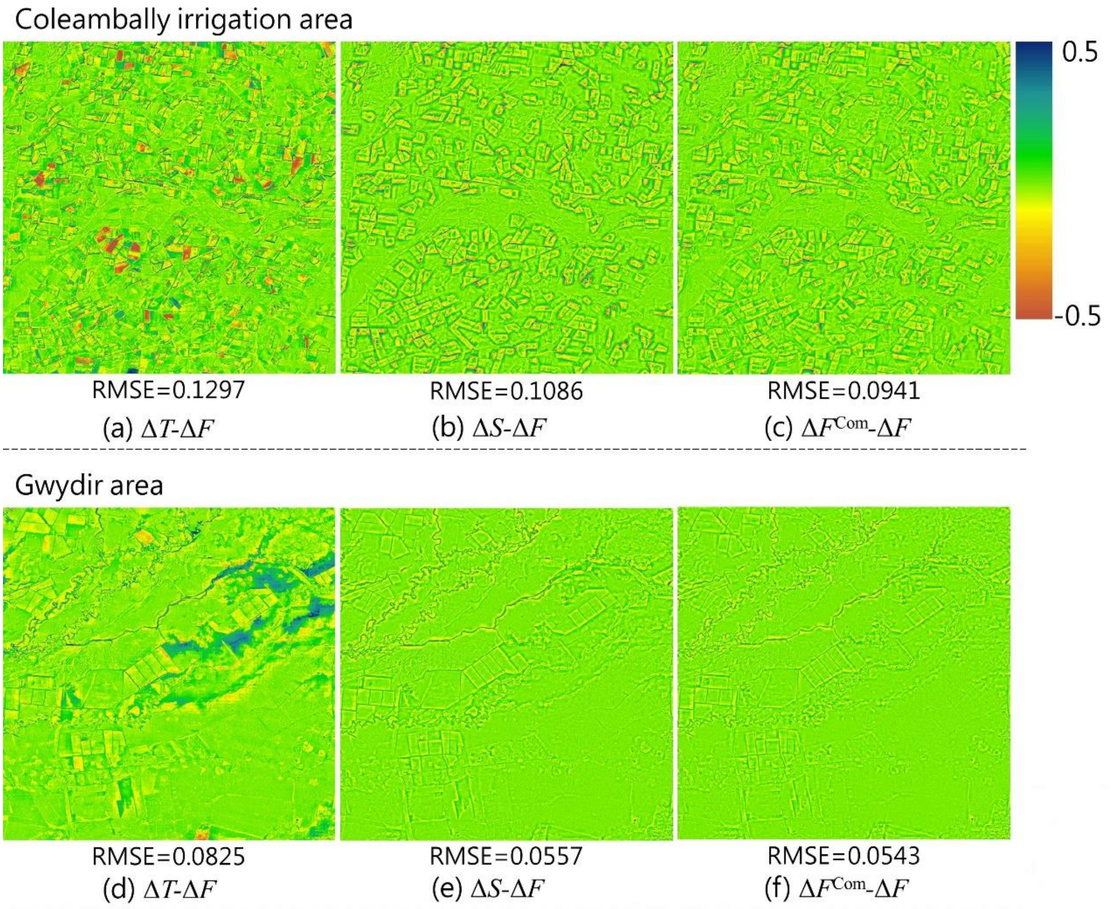

Fig. 18. Difference between predicted increment and observed increment: difference between (a) temporal increment ΔT, (b) spatial-dependent increment ΔS, (c) combined increment ΔF^{Com} and the observed increment ΔF in the Coleambally irrigation area; difference between (d) temporal increment ΔT, (e) spatial-dependent increment ΔS, (f) combined increment ΔF^{Com} and the observed increment ΔF in the Gwydir area.

5.4 Improvements of IFSDAF compared with FSDAF

Compared with FSDAF, IFSDAF has improved in the following aspects. First, the increment estimation in FSDAF mainly produced by the unmixing process, and the TPS interpolation result is only used to guide the distribution of residuals rather than producing spatial-dependent increment. However, as shown in Fig.18, the spatial-dependent increment estimated by the TPS interpolation may be more accurate than the temporal increment by the unmixing process. FSDAF underestimates the contribution of the TPS interpolation to some extent. The reason why the spatial-dependent increment is superior to the temporal increment can be found from Table 5, where we calculated the global Moran's I index of the coarse images for band Red, band NIR, and NDVI on the base date t_0 and prediction date t_p, respectively. The global Moran's I index was used here to measure the spatial autocorrelation of the image, i.e., the relationship of pixel values between neighboring pixels. Larger Moran's I index indicates higher spatial autocorrelation. Table 5 shows that the spatial autocorrelation of NDVI represented by Moran's I index is greater than both the Red and NIR bands, because NDVI, as a feature-enhancing index, can enlarge the data variance compared with the Red and NIR bands (Fig.11a-b). As well known, greater spatial autocorrelation can yield more accurate result in spatial interpolation. Accordingly, the spatial-dependent increment estimated by the TPS interpolation for NDVI should be more accurate than that of Red and NIR bands. Therefore, spatial-dependent increment is more important for fusing NDVI than the raw bands, which greatly benefits the NDVI fusion in IFSDAF. This study implies that without combined with temporal increment, spatial-dependent increment itself may obtain acceptable fusion results. This solution can greatly simplify the fusion process and reduce the computing cost. The simplified fusion model is more effective for the applications in larger areas and when the scale difference between coarse and fine images are not too large, which ensures adequate number of sample points (the center of coarse pixel) to obtain accurate prediction of the fine image by TPS interpolation.

Table. 5. Moran's I of band Red, band NIR, and NDVI on the base date t_0 and prediction date t_p in both the Coleambally irrigation area and Gwydir area at the 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

Second, IFSDAF uses a better way to combine two increments while FSDAF uses only one increment. As we know, the collinearity effect impacts the accuracy of unmixing for temporal increment estimation. Moreover, errors in the classification map and change of land cover also cause uncertainties of temporal increment. In order to correct the potential errors in the temporal increment, FSDAF introduces a homogeneity index $HI(x_j, y_j)$ (over the range of 0-1, derived from the classification map at t_0) to help allocate residuals $R(x, y)$ within the coarse pixel. However, when there are land cover changes and misclassification, $HI(x_j, y_j)$ calculated from the classification map at t_0 will not be suitable for allocating residuals on date t_p. Under this circumstance, the effectiveness of residuals distribution in FSDAF is restricted. Unlike FSDAF, the IFSDAF employs CLS method in moving windows and avoids the use of the homogeneity index; moreover, it allows the final increment estimation with local and adaptive capacity to better combine temporal and spatial-dependent increment.

The third improvement of IFSDAF is that it can employ fine NDVI images partially 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 date to estimate the fine NDVI values at the prediction date respectively, and all predictions on date t_p are finally integrated by weights based on the temporal change magnitude in NDVI between base and prediction dates. This weighted prediction can reduce the critical dependency to the clear fine NDVI image and alleviate prediction uncertainties if date of the clear fine NDVI image is far from the prediction date. Of course, the better use of partially contaminated images needs accurate cloud labeling method (e.g., Fmask method). If there are mistakes in cloud labels, estimation results of IFSDAF will be impacted. For instance, land surface with high reflectance (e.g., sand or snow) is possibly misidentified as clouds (Chen et al., 2016). Moreover, Fmask sometimes omits thin clouds, resulting in cloudy pixels being used in the process of data fusion. Fortunately, effect of the errors in cloud mask can be minimized in IFSDAF because of the weighted combination of predictions from multiple dates. Moreover, with the advance of cloud screening methods (Zhu and Helmer, 2018), the issue can be greatly alleviated. Besides, like other existing spatiotemporal fusion methods, a pure clear image guarantees all pixels to be predicted by IFSDAF. However, in some areas such as tropical areas (e.g., Amazonia), it is difficult to obtain such clear fine image during a long period. Under this condition, using all partially contaminated fine images instead of a purely clear image is a practical choice in IFSDAF although it may leave some pixels not predicted in the fused images if these pixels do not have any one cloud-free observation in the time series.

5.5 Applications to other remote sensing products

Although IFSDAF is designed for the spatiotemporal fusion of NDVI time series, it can also be applied to fuse other vegetation indices like Enhanced Vegetation Index (EVI) and other products such as surface reflectance. To test the applicability of IFSDAF to other products, we assess the performance of IFSDAF in fusion of EVI, Red and NIR bands in Coleambally Irrigation area where have great heterogeneity. RMSEs of the fused images on Jan. 12th, 2002 (Table. 6) suggest that IFSDAF produces higher accuracy that FSDAF when fusing EVI, while when predicting surface reflectance (Red and NIR bands) the accuracy of IFSDAF and FSDAF does not differ too much. These results confirm that IFSDAF is more suitable than the original FSDAF model for fusing remote sensing products with high spatial autocorrelation.

Table. 6. RMSE (rRMSE) of predictions by IFSDAF and FSDAF on 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%)

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 different sensors. Experiments show that the fused NDVI images by IFSDAF is more accurate than FSDAF as well as other two 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 producing spatial-dependent increment by the TPS interpolation, employing CLS method in moving windows to adaptively combine the temporal increment and the spatial-dependent increment, as well as the better use of partially contaminated fine images. Such significant improvements are made in accordance with the characteristics of NDVI with larger data variance and spatial autocorrelation compared with raw reflectance bands. Considering the significant contribution of spatial-dependent increment by the TPS interpolation, when the scale difference between coarse and fine images is not very large, the proposed IFSDAF method can be further simplified by only using spatial-dependent increment to improve the efficiency. This result of the study also supports the IFSDAF to be a feasible method for applications in a large area and different sensors. Moreover, it 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 areas.

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), and the National Natural Science Foundation of China (project no.41701378), and CEReS Oversea Joint Research Program, Chiba University (No.CI17-103).

Reference

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 Sensing of Environment, 112*, 118-131

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 Journal of Photogrammetry and Remote Sensing, 103*, 7-27

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 Sensing of Environment, 91*, 332-344

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 Transactions on Geoscience and Remote Sensing, 54*, 2682-2694

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 Sensing, 6*, 2845-2863

Chen, X., Wang, D., Chen, J., Wang, C., & Shen, M.G. (2018). The Mixed Pixel Effect in Land Surface Phenology: A Simulation Study. *Remote Sensing of Environment*, *211*, 388-344.

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. *Photogrammetric Engineering and Remote Sensing, 84*, 66-74

Dubrule, O. (1984). Comparing Splines and Kriging. *Computers & Geosciences, 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 Sensing of Environment, 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 Transactions on Geoscience and Remote Sensing, 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 Sensing of Environment, 156*, 34-44

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

Huang, B., & Song, H. (2012). Spatiotemporal Reflectance Fusion via Sparse Representation. *IEEE Transactions on Geoscience and Remote Sensing, 50*, 3707-3716

Huang, B., & Zhang, H.K. (2014). Spatio-temporal reflectance fusion via unmixing: accounting for both phenological and land-cover changes. *International Journal of Remote Sensing, 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 Sensing of Environment, 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 Sensing, 6*, 9213-9238

Lillesaeter, O. (1982). Spectral reflectance of partly transmitting leaves: Laboratory measurements and mathematical modeling. *Remote Sensing of Environment*, 12, 247-254

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 Geoscience and Remote Sensing Letters, 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 Sensing of Environment, 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. *International Journal of Digital Earth, 6*, 203-218

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 in Ecology & Evolution, 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 Sensing, 7*, 7865-7891

Rouse, J.W., Jr., 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*, 309–317

Song, H., & Huang, B. (2013). Spatiotemporal Satellite Image Fusion Through One-Pair Image Learning. *IEEE Transactions on Geoscience and Remote Sensing, 51*, 1883-1896

Tian, F., Wang, Y.J., Fensholt, 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 Sensing, 5*, 4255-4279

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. *Ecological Indicators, 93,* 838-846

Wang, Q.M., & Atkinson, P.M. (2018). Spatio-temporal fusion for daily Sentinel-2 images. *Remote Sensing of Environment, 204*, 31-42

Wu, B., Huang, B., & Zhang, L. (2015). An Error-Bound-Regularized Sparse Coding for Spatiotemporal Reflectance Fusion. *IEEE Transactions on Geoscience and Remote Sensing, 53*, 6791-6803

Zhang, H.K.K., Huang, B., Zhang, M., Cao, K., & Yu, L. (2015). A generalization of spatial and temporal fusion methods for remotely sensed surface parameters. *International Journal of Remote Sensing, 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. *Journal of Integrative Plant Biology, 47,* 1431-1449

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 Sensing, 10*, 527

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 Sensing of Environment, 114*, 2610-2623

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 Sensing of Environment, 214*, 135-153

Zhu, X., Helmer, E.H., Gao, F., Liu, D., Chen, J., & Lefsky, M.A. (2016). A flexible spatiotemporal method for fusing satellite images with different resolutions. *Remote Sensing of Environment, 172*, 165-177

Zhu, Z., & Woodcock, C.E. (2012). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment, 118*, 83-94

Zurita-Milla, R., Clevers, J., & Schaepman, M.E. (2008). Unmixing-Based Landsat TM and MERIS FR Data Fusion. *IEEE Geoscience and Remote Sensing Letters, 5*, 453-457

Appendix

Useful notations.

t_0	Base date	$f_l(x, y)$	Fraction of class l within coarse pixel (x, y)
t_p	Prediction date	ΔF_c	Fine spatial resolution increment of class c within the moving window
(x, y)	Location of coarse spatial resolution pixel (x, y)	F_0^{TPS}	Result of TPS interpolation based on coarse NDVI on t_0
(x_j, y_j)	Location of jth fine spatial resolution pixel within coarse pixel (x, y)	F_p^{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
ΔF	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

ΔC	Coarse spatial resolution NDVI increment	ΔF^{Com}	Combined fine spatial resolution increment based on ΔT and ΔS
ΔT	Fine spatial resolution temporal increment	$R(x, y)$	Residual within the coarse pixel (x, y)
ΔS	Fine spatial resolution spatial-dependent increment	$C_q^i(x, y)$	ith coarse pixel in the moving window centered by coarse pixel (x, y) on date q
ΔC^{T}	Upscaled fine spatial resolution temporal increment	$C_p^i(x, y)$	ith coarse pixel in the moving window centered by coarse pixel (x, y) on date t_p
ΔC^{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)