Remote Sensing of Environment 304 (2024) 114046

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

OBSUM: An object-based spatial unmixing model for spatiotemporal fusion of remote sensing images

Houcai Guo [a], Dingqi Ye [b], Hanzeyu Xu [c], Lorenzo Bruzzone [a,*]

[a] Department of Information Engineering and Computer Science, University of Trento, Trento 38123, Italy
[b] School of Geosciences and Info-Physics, Central South University, Changsha 410083, China
[c] School of Geography, Nanjing Normal University, Nanjing 210023, China

ARTICLE INFO

Editor name: Jing M. Chen

Keywords:
Spatiotemporal fusion
Object-based image analysis
Sentinel-2
Sentinel-3
Time-series
Remote sensing

ABSTRACT

Spatiotemporal fusion aims to improve both the spatial and temporal resolution of remote sensing images, thus facilitating time-series analysis at a fine spatial scale. However, there are several important issues that limit the application of current spatiotemporal fusion methods. First, most spatiotemporal fusion methods are based on pixel-level computation, which neglects the valuable shape information of ground objects. Moreover, many existing methods cannot accurately retrieve strong temporal changes between the available high-resolution image at base date and the predicted one. This study proposes an Object-Based Spatial Unmixing Model (OBSUM), which incorporates object-based image analysis and spatial unmixing, to overcome the two above-mentioned problems. OBSUM consists of one preprocessing step and three fusion steps, i.e., object-level unmixing, object-level residual compensation, and pixel-level residual compensation. The performance of OBSUM was compared with seven representative spatiotemporal fusion methods at two agricultural sites. The experimental results demonstrated that OBSUM outperformed other methods in terms of both accuracy indices and visual effects over time-series. Furthermore, OBSUM also achieved satisfactory results in crop progress monitoring and crop mapping. Therefore, it has great potential to generate accurate and high-resolution time-series observations for supporting various remote sensing applications.

1. Introduction

Dense satellite image time-series with high spatial resolution can effectively facilitate various remote sensing applications in ecology (Guo et al., 2022a), agriculture (Gao et al., 2017), and disaster (Zhang et al., 2014). However, due to the trade-off between spatial resolution and revisit period, no single satellite-based sensor is able to perform fine-scale Earth observation with a daily frequency in a cost-effective way. For instance, the Multi-Spectral Instrument (MSI) carried by Sentinel-2 satellite constellation has a relatively long revisit period of 5 days, but a high spatial resolution of 10 m (hereafter, fine image) (Drusch et al., 2012). By contrast, the Ocean and Land Colour Instrument (OLCI) carried by Sentinel-3 satellite constellation has a daily revisit frequency, but a coarse spatial resolution of 300 m (hereafter, coarse image) (Donlon et al., 2012). The recently launched PlanetScope satellite constellation is able to make daily observations with 3 m spatial resolution (Kwan et al., 2018). However, the utilization of PlanetScope images in certain remote sensing applications suffers not only from the high price, but also from the relatively low radiometric quality (Houborg and McCabe, 2018). By contrast, the Harmonized Landsat and Sentinel-2 (HLS) (Claverie et al., 2018) project integrates the Operational Land Imager (OLI) and MSI aboard the Landsat 8/9 and Sentinel-2 satellites, respectively, and the HLS data product has a 2–3 days temporal resolution. Nevertheless, its spatial resolution remains 30 m, the same as the Landsat 8/9 OLI images, which limits its applicability to many fine-scale applications (Gu et al., 2023). Considering these existing problems, spatiotemporal fusion provides a cost-effective solution by blending the abovementioned coarse and fine images, thus generating images with both high spatial and temporal resolution (Gao et al., 2015; Ghamisi et al., 2019; Zhu et al., 2018).

Generally, spatiotemporal fusion methods can be divided into five main categories: spatial unmixing-based, weight function-based, Bayesian-based, learning-based, and hybrid methods (Zhu et al., 2018). The Multisensor Multiresolution Technique (MMT) is the first spatial unmixing-based method developed to fuse images with different spatial and temporal resolutions (Zhukov et al., 1999). Based on its

* Corresponding author.
E-mail address: lorenzo.bruzzone@unitn.it (L. Bruzzone).

https://doi.org/10.1016/j.rse.2024.114046
Received 10 October 2023; Received in revised form 5 February 2024; Accepted 7 February 2024

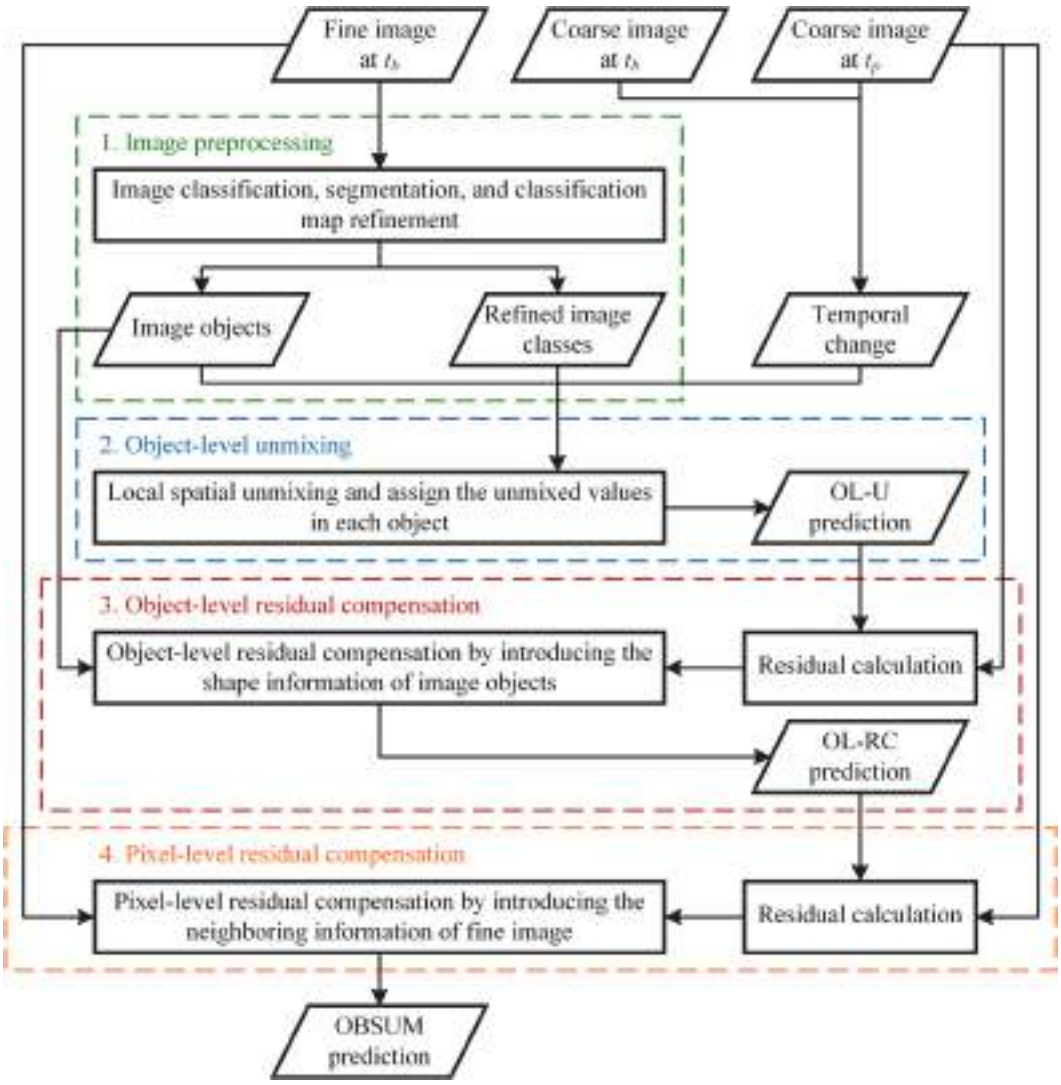

Fig. 1. Flowchart of OBSUM. The four steps are represented by containers with different colors.

framework, many spatial unmixing-based methods have been proposed in the past decades, including Unmixing-Based Data Fusion (UBDF) (Zurita-Milla et al., 2008), Spatial Temporal Data Fusion Approach (STDFA) (Mingquan et al., 2012), Blocks-Removed Spatial Unmixing (SU-BR) (Wang et al., 2021), etc. The Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) is the first weight function-based spatiotemporal fusion method presented in the literature (Feng et al., 2006). It predicts the reflectance of the target fine pixel by combining the reflectance changes of its spatially neighboring similar pixels. Many approaches have been proposed to improve the performance of STARFM, for example, Spatial Temporal Adaptive Algorithm for mapping Reflectance Change (STAARCH) (Hilker et al., 2009), Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model (ESTARFM) (Zhu et al., 2010), and the three-step method (Fit-FC) (Wang and Atkinson, 2018). Bayesian-based methods consider spatiotemporal fusion as a maximum a posterior problem and predict the fine image by optimizing the probability functions (Li et al., 2013; Shen et al., 2016). Learning-based methods employ machine learning techniques, e.g., dictionary pair learning (Huang and Song, 2012), convolutional neural network (Liu et al., 2019) and generative adversarial network (Zhang et al., 2021a) to model the relationship between the input coarse images and the fine image. Hybrid methods, such as Flexible Spatiotemporal DAta Fusion (FSDAF) (Zhu et al., 2016), Reliable and Adaptive Spatiotemporal Data Fusion (RASDF) (Shi et al., 2022), and Variation-based Spatiotemporal Data Fusion (VSDF) (Xu et al., 2022), combine both spatial unmixing technique and the weight function-based methods to improve the prediction robustness. A more detailed review to spatiotemporal fusion methods can be found in Zhu et al. (2018) and Wang et al. (2023).

Despite hundreds of spatiotemporal fusion methods have been developed rapidly in the past decades, there are still several remaining challenges. First, the unmixing-based methods suffers from the block effect, i.e., fine pixels of the same land-cover class inside two adjacent coarse pixels present different reflectance and results in clear footprints of coarse pixels in the predicted image. The SU-BR method introduces an iterative optimization strategy that considers both errors of the unmixing model and reflectance differences to remove the block effects. However, the iterative optimization process also leads to computational inefficiency. The unmixing-based methods also ignore the compensation of residuals, which leads to inaccurate prediction. Second, the weight function-based methods are based on pixel-level computation, which neglects the inherent and valuable shape information of the ground objects. As a result, these methods suffer from both low computational efficiency and intra-class spectral variation (Guo et al., 2022b). Moreover, many existing methods have difficulty in accurately retrieving strong temporal changes. For example, the Fit-FC method assumes the temporal changes of the ground objects are scale-invariant, and models the temporal changes by applying local window-based linear regression

to the input coarse images. After that, the regression models are applied to the input fine image to obtain an initial prediction. Even if the following spatial filtering step can alleviate the block effect introduced by the regression model fitting, there are still spectral distortions in the final prediction caused by the scale inconsistency, especially in heterogeneous areas (Shi et al., 2022).

In recent years, several object-level fusion methods have been proposed and proved their satisfactory performances, including the Object-restricted strategy (Guan et al., 2019), Object-Based SpatioTemporal Fusion Model (OBSTFM) (Zhang et al., 2021b), Object-Level (OL) weight function methods (Guo et al., 2022b), and the Object-Level Hybrid SpatioTemporal Fusion Method (OL-HSTFM) (Guo and Shi, 2023). Actually, spatiotemporal fusion is an ill-posed problem and its accuracy is affected by various factors, such as the time interval between the base date and the prediction date, the scale factor between the coarse and fine images, and the heterogeneity of the land-cover (Zhu et al., 2018). An object is a shape composed of spectrally similar pixels that are spatially adjacent, and all pixels within the object belong to the same land-cover class (Blaschke, 2010; Hossain and Chen, 2019). The object-level shape information is one of the inherent characteristics of the land surface and is valuable for improving the fusion accuracy. For example, the object-level processing takes an object as a homogeneous unit, which helps to retrieve the strong temporal changes and alleviate the uncertainty caused by spatial heterogeneity (Belgiu and Csillik, 2018). However, the Object-restricted strategy simply adds a constraint for selecting similar pixels, and the OL weight function methods are only modifications to existing weight function-based methods. OBSTFM ignores the compensation of prediction residuals, which leads to errors in retrieving temporal changes. Furthermore, the similar pixel searching strategy in OBSTFM may also dismiss the within-object land-cover changes. OL-HSTFM combines the OL-STARFM and OL-Fit-FC methods (Guo et al., 2022b), in which the residual compensation steps simply add the coarse-scale residuals to the preliminary predictions in each object. As a result, this approach ignores the uncertainties in the residual map caused by the scale difference between the coarse and fine images. To the best of our knowledge, no existing method has combined object-level processing, spatial unmixing, and weight functions to obtain a more accurate fusion result. Moreover, it is necessary to develop a residual compensation scheme that considers the shape information of ground objects.

In order to address the aforementioned problems, an Object-Based Spatial Unmixing Model (OBSUM) is proposed and validated in this paper. OBSUM is a hybrid method that utilizes object-level shape information to obtain more accurate fusion results. The object-based image analysis (OBIA), spatial unmixing, and combination of similar pixels are integrated into OBSUM. The object-level unmixing produces an initial prediction in which no block effect appears. After that, a novel object residual index is proposed to calculate and compensate the residual of each object to improve the fusion accuracy significantly and retrieve strong temporal changes. Finally, a weight function-like process is adopted to predict within-object land-cover changes and further improve the fusion accuracy. The performance of OBSUM is compared to those of seven typical spatiotemporal fusion methods, including UBDF, STARFM, Fit-FC, FSDAF, RASDF, VSDF, and OL-HSTFM, using time-series of Sentinel-2 MSI and Sentinel-3 OLCI images. The potential of OBSUM to support various remote sensing applications is also discussed in detail.

The remainder of this paper is organized as follows. Section 2 introduces the methodology of OBSUM. Section 3 presents the comparison experiments. Section 4 introduces two potential application scenarios supported by OBSUM. Finally, Section 5 gives a detailed discussion of OBSUM and draws the conclusion on the main findings of this paper.

2. Methodology

OBSUM requires one pair of coarse and fine images at the base date t_b and one coarse image at the prediction date t_p to predict the fine image at t_p. As shown in Fig. 1, OBSUM is based on four main steps: (1) image preprocessing, (2) object-level unmixing, (3) object-level residual compensation, and (4) pixel-level residual compensation. Table 1 shows the main acronyms and the descriptions used in this section. The detailed description of the method is given below.

Table 1
Main acronyms and related descriptions.

Acronym	Description
OL-U	Object-level unmixing
OL-RC	Object-level residual compensation
OHI	Object homogeneity index
ORI	Object residual index
OL-R	Object-level residual calculated by OL-RC
PL-RC	Pixel-level residual compensation
PL-R	Pixel-level residual calculated by PL-RC

2.1. Image preprocessing

The image preprocessing includes image classification, image segmentation, and classification map refinement. First, the fine image at t_b is classified into several land-cover classes. Either a supervised or unsupervised classifier can be applied to obtain this, depending on the availability of reference land-cover labels. In this study, we employed the simple unsupervised K-Means classifier (Lloyd, 1982) to make OBSUM fully automatic and independent from a training set.

Then, the fine image at t_b is segmented into homogeneous regions, i.e., image objects. In this study, the state-of-the-art segment anything model (SAM) (Kirillov et al., 2023) is used to perform image segmentation. However, any other image segmentation technique can be used with OBSUM. SAM has been trained on a dataset that contains 11 million images and 1.1 billion segmentation masks. It can generate masks for all objects in an image and has impressive zero-shot performance on any segmentation task. Specifically, for the input image, given the number of samples in the image and the output normative parameters (including mask filtering threshold, non-maximum suppression, etc.), SAM can get the object mask of all pixels within the image automatically.

After image classification on the pixel-level and an independent, subsequent image segmentation, the image objects are used to refine the classification map. The OBIA technique assumes all pixels inside an object have similar spectral characteristics, therefore, these pixels should also belong to the same land-cover class. After classification map refinement, the land-cover class of all fine pixels within an object is set to the mode of the original classes of these fine pixels. This drives the subsequent object-level unmixing step.

2.2. Object-level unmixing

The object-level unmixing (OL-U) incorporates the spatial unmixing technique and OBIA to obtain an initial prediction. First, the coarse-scale temporal change is calculated as:

$$\Delta C(x_i, y_i, b) = C_{tp}(x_i, y_i, b) - C_{tb}(x_i, y_i, b) \quad (1)$$

where $\Delta C(x_i, y_i, b)$ is the band b temporal change of the target coarse pixel at location (x_i, y_i), $C_{tb}(x_i, y_i, b)$ and $C_{tp}(x_i, y_i, b)$ are the corresponding band b reflectance of this coarse pixel at t_b and t_p, (respectively).

According to the spatial unmixing theory, assuming that land-cover changes are not scattered within a coarse pixel, the temporal change of the coarse pixel can be modeled as the linear combination of the temporal changes of each class of fine pixels within it:

$$\Delta C(x_i, y_i, b) = \sum_{c=1}^{n_c} f_c(x_i, y_i) \times \Delta F(c, b) \quad (2)$$

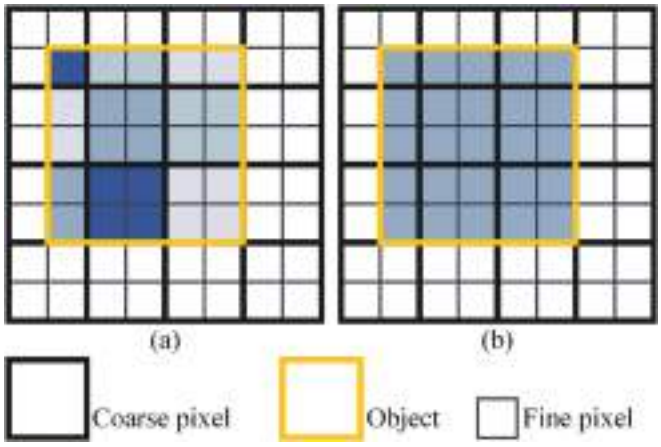

Fig. 2. Fine-scale temporal changes (ΔF) estimated by (a) spatial unmixing and (b) OL-U.

where n_c is the number of land-cover classes, $f_c(x_i, y_i)$ denotes the fraction of class c inside the target coarse pixel, and $\Delta F(c, b)$ is the band b temporal change of class c. The refined classification map is used to calculate $f_c(x_i, y_i)$ by counting the proportion of class c inside the coarse pixel.

For a coarse pixel located at (x_i, y_i), $\Delta F(c, b)$ can be estimated through a local square window with size w centered on this target coarse pixel. The coarse pixels inside this window can compose the following linear equation system:

$$\begin{bmatrix} \Delta C(x_1, y_1, b) \\ \vdots \\ \Delta C(x_i, y_i, b) \\ \vdots \\ \Delta C(x_n, y_n, b) \end{bmatrix} = \begin{bmatrix} f_1(x_1, y_1) & f_2(x_1, y_1) & \cdots & f_{n_c}(x_1, y_1) \\ \vdots & \vdots & & \vdots \\ f_1(x_i, y_i) & f_2(x_i, y_i) & \cdots & f_{n_c}(x_i, y_i) \\ \vdots & \vdots & & \vdots \\ f_1(x_n, y_n) & f_2(x_n, y_n) & \cdots & f_{n_c}(x_n, y_n) \end{bmatrix} \begin{bmatrix} \Delta F(1, b) \\ \vdots \\ \Delta F(c, b) \\ \vdots \\ \Delta F(n_c, b) \end{bmatrix} \tag{3}$$

where n is the number of coarse pixels inside the local window, subject to $n = w^2$. Eq. (3) can be solved through the least square method, and the resulting fine-scale temporal changes are assigned to each land-cover class inside each coarse pixel to obtain a preliminary estimation of the fine-scale temporal changes.

Following the basic principle of OBIA, the temporal change of the oth object is assigned as the mean temporal changes of all fine pixels within the object:

$$\Delta F(o, b) = \frac{1}{p_o} \sum_{k=1}^{p_o} \Delta F(k, b) \tag{4}$$

where p_o is the number of fine pixels within the oth object, and $\Delta F(k, b)$ is the band b temporal change of the kth fine pixel within the oth object.

After that, the OL-U prediction is obtained by adding the fine-scale temporal changes to the fine image at t_b:

$$OL\text{-}U_{tp}(x_{ij}, y_{ij}, b) = F_{tb}(x_{ij}, y_{ij}, b) + \Delta F(x_{ij}, y_{ij}, b) \tag{5}$$

where $OL\text{-}U_{tp}(x_{ij}, y_{ij}, b)$ is the predicted band b reflectance of the fine pixel located at (x_{ij}, y_{ij}), $F_{tb}(x_{ij}, y_{ij}, b)$ and $\Delta F(x_{ij}, y_{ij}, b)$ are the band b reflectance at t_b and the temporal change of this fine pixel, respectively. Fig. 2 shows an example of OL-U. It can be seen that the fine pixels within an object have the same temporal change, which provides a basis for the subsequent object-level residual compensation.

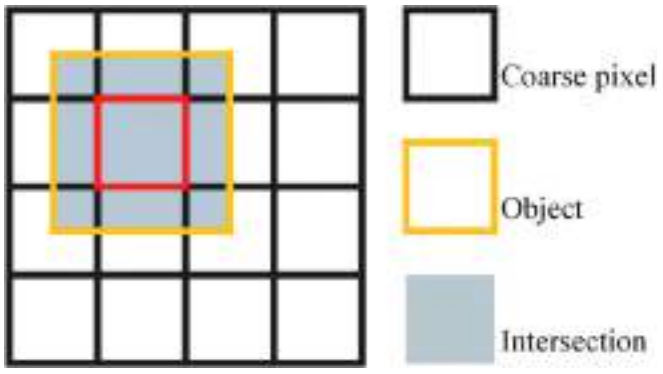

Fig. 3. An example of OL-RC using coarse residuals.

2.3. Object-level residual compensation

The OL-U produces an initial prediction by assuming that all fine pixels within an object have the same temporal change. However, the OL-U prediction is not accurate enough because it also assumes the land-cover classes are stable from t_b to t_p, while ignoring that the possible presence of changed coarse pixels would decrease the accuracy of the local unmixing process. Thus, residual compensation is necessary to recover the spectral information and improve the prediction accuracy. By assuming all fine pixels within an object have the same residual, the objective of object-level residual compensation (OL-RC) is to calculate and compensate the residual for each object.

The OL-U prediction is first upscaled to the resolution of the coarse image to get a coarse prediction, and the coarse residuals are calculated by subtracting the coarse prediction from the coarse image at t_p. As shown in Fig. 3, an object intersects with nine coarse residual pixels, and the residual of the red coarse pixel is most likely to be the residual of this object because this pixel is fully intersected with the object. This is the theoretical basis of the proposed OL-RC. However, considering the complexity and heterogeneity of the land-cover, the coarse residuals are not suitable for residual compensation of small objects, in which no complete coarse residual pixel exists. Therefore, the coarse residuals are downscaled to the resolution of the fine image using a bi-cubic interpolation to obtain the fine residuals $R_1(x_{ij}, y_{ij}, b)$. As inspired by the residual compensation step in FSDAF (Zhu et al., 2016), the scheme of OL-RC needs to be adjusted accordingly.

Inside the coarse pixel located at (x_i, y_i), the normalized spatial distance DC between the jth fine pixel and the center of the coarse pixel is defined as:

$$DC(x_{ij}, y_{ij}) = 1 + \sqrt{\left(x_{ij} - x_{i\frac{m}{2}}\right)^2 + \left(y_{ij} - y_{i\frac{m}{2}}\right)^2} \Big/ (s/2) \tag{6}$$

where (x_{ij}, y_{ij}) is the coordinates vector of the jth fine pixel, m is the number of fine pixels inside one coarse pixel, $(x_{i\frac{m}{2}}, y_{i\frac{m}{2}})$ is the coordinates vector of the coarse pixel center, and s is the scale factor between the coarse image and the fine image, subject to $m = s^2$. DC ranges between 1 and $1 + \sqrt{2}$, and a lower value indicates a higher similarity between the downscaled fine residual at (x_{ij}, y_{ij}) and the original coarse residual at (x_i, y_i).

In order to perform OL-RC, we introduce the object homogeneity index OHI:

$$OHI(x_{ij}, y_{ij}) = \frac{1}{m} \sum_{k=1}^{m} I_k \tag{7}$$

If the kth fine pixel inside a local window (the window size is one coarse pixel, and the window center is the target fine pixel) belongs to the same object as the target pixel located at $\left(x_{ij}, y_{ij}\right)$, $I_k = 1$; otherwise, $I_k = 0$. *OHI* ranges between 0 and 1, and a higher value indicates a more homogeneous object-level land-cover.

Theoretically, if both a fine residual is similar to the residual of the coarse pixel it is located in, and the object-level land-cover inside the coarse pixel is homogeneous, then the fine residual has a higher similarity to the actual residual of the object it is located in. Therefore, *DC* and *OHI* are combined to calculate the object residual index (*ORI*):

$$ORI\left(x_{ij}, y_{ij}\right) = OHI\left(x_{ij}, y_{ij}\right) / DC\left(x_{ij}, y_{ij}\right) \tag{8}$$

ORI ranges between 0 and 1, and a higher value indicates a higher similarity of the fine residual to the actual object residual.

After calculating *ORI*, a possible scheme for OL-RC is to select a fine pixel with the highest *ORI* within each object and then assign the fine pixel's residual as the predicted object-level residual. However, given that bicubic interpolation would smooth the residual map and there are uncertainties caused by errors of the previous steps, only one fine residual pixel is not robust enough to compensate the residual of an object. Therefore, we first select a certain percentage of fine residual pixels that have the highest *ORI* within an object, then combine the selected fine residuals to calculate the object-level residual. The number of selected fine residual pixels for the oth object is defined as r_o:

$$r_o = OR\ percent \times p_o \tag{9}$$

where *OR percent* is the percentage of selected fine residual pixels for the oth object, and p_o is the number of fine pixels inside the oth object. Considering the *OR percent*, a relatively low value is unstable for calculating the object-level residual, whereas a relatively high value could introduce errors since the bi-cubic interpolation would smooth the residual map. The determination of the optimal value of *OR percent* will be discussed in the Section 3.2.

After the selection of fine residual pixels, the weight of the kth selected pixel is calculated as:

$$W_k = ORI_k \Big/ \sum_{k=1}^{r_o} ORI_k \tag{10}$$

where ORI_k is the object residual index of the kth selected fine pixel.

The predicted object-level residual (OL-R) of the oth object is the weighted sum of the residuals of the selected fine pixels:

$$OL\text{-}R(ob) = \sum_{k=1}^{r} W_k \times R_1(k, o, b) \tag{11}$$

where $R_1(k, o, b)$ is the band b residual of the kth selected fine pixel inside the oth object.

After that, OL-R is added to the OL-U prediction to get the OL-RC prediction:

$$OL\text{-}RC_{tp}\left(x_{ij}, y_{ij}, b\right) = OL\text{-}U_{tp}\left(x_{ij}, y_{ij}, b\right) + OL\text{-}R\left(x_{ij}, y_{ij}, b\right) \tag{12}$$

2.4. Pixel-level residual compensation

The OL-RC calculates and compensates the residual for each object, and the prediction is much more accurate than that of the OL-U. However, both OL-U and OL-RC assume the land-cover type inside each object is stable from t_b to t_p. As a result, these two steps cannot recover within-object land-cover changes in the predicted image. The pixel-level residual compensation (PL-RC) employs a strategy similar to the weight function-based spatiotemporal fusion methods to recover the within-object land-cover changes and further improve the prediction accuracy.

The OL-RC prediction is first upscaled to the resolution of the coarse image to get a coarse prediction, and the coarse residuals are calculated by subtracting the coarse prediction from the coarse image at t_p. After that, the coarse residuals are downscaled to the resolution of the fine image using a bi-cubic interpolation to obtain the fine residuals $R_2\left(x_{ij}, y_{ij}, b\right)$. Similar to the residual compensation step in Fit-FC (Wang and Atkinson, 2018), the detailed process of PL-RC is given below.

The spectral distance between a target fine pixel located at $\left(x_{ij}, y_{ij}\right)$ and its kth neighboring fine pixel is defined as:

$$S_k = \frac{1}{n_b} \sum_{b=1}^{n_b} \left| F_{tb}(x_k, y_k, b) - F_{tb}\left(x_{ij}, y_{ij}, b\right) \right| \tag{13}$$

where n_b is the number of spectral bands of the fine image, and $F_{tb}\left(x_k, y_k, b\right)$ and $F_{tb}\left(x_{ij}, y_{ij}, b\right)$ are the band b reflectance of the kth neighboring pixel and the target pixel, respectively.

Inside the local square window with size w_s centered on the target fine pixel, a total number of n_s pixels with the smallest S_k are selected as similar pixels. The spatial distance between the target pixel and its kth similar pixel is defined as D_k:

$$D_k = 1 + \sqrt{\left(x_k - x_{ij}\right)^2 + \left(y_k - y_{ij}\right)^2} \Big/ (w_s/2) \tag{14}$$

where D_k is a relative distance that ranges between 1 and $1 + \sqrt{2}$.

The pixel-level residual (PL-R) of a target pixel is the weighted combination of the residuals of its neighboring similar pixels. According to Tobler's first law of geography (Tobler, 1970), similar pixels that are spatially closer to the target pixel contribute more to the combined residual than similar pixels that are farther from the target pixel. Therefore, the weight of the kth similar pixel is calculated as:

$$W_k = (1/D_k) \Big/ \sum_{k=1}^{n_s} (1/D_k) \tag{15}$$

After that, the residual of the target fine pixel located at $\left(x_{ij}, y_{ij}\right)$ is calculated as:

$$PL\text{-}R\left(x_{ij}, y_{ij}, b\right) = \sum_{k=1}^{n_s} W_k \times R_2(x_k, y_k, b) \tag{16}$$

Finally, the OBSUM prediction is obtained by adding PL-R to the OL-RC prediction:

$$OBSUM_{tp}\left(x_{ij}, y_{ij}, b\right) = OL\text{-}RC_{tp}\left(x_{ij}, y_{ij}, b\right) + PL\text{-}R\left(x_{ij}, y_{ij}, b\right) \tag{17}$$

3. Experiments and results

3.1. Study sites and dataset

Sentinel-2 MSI and Sentinel-3 OLCI images were used as fine and coarse images in the experiments. For Sentinel-2 MSI images, we collected the Level-2A atmospherically corrected surface reflectance products that were provided by the European Space Agency. The 10 m spectral bands B02 (blue), B03 (green), and B04 (red) were used. We also downscaled the 20 m near infra-red (NIR) band B8A to a spatial resolution of 10 m using the Sen2Res plugin (Brodu, 2017) of the SNAP software. For Sentinel-3 OLCI images, we downloaded the full resolution Level-1B top of atmosphere radiance products and performed atmosphere correction using the iCOR plugin (Stefan et al., 2018) of the SNAP software. We considered the 300 m spectral bands Oa4 (blue), Oa6 (green), Oa8 (red), and Oa17 (NIR). All spatiotemporal fusion methods produce fine images with similar reflectance with the coarse images, i.e., the Sentinel-3 OLCI images at the prediction date. In order to achieve fair comparison, we further maximized the reflectance consistency between the Sentinel-2 MSI and the Sentinel-3 OLCI images by applying a global linear transformation to each OLCI spectral band. Such a

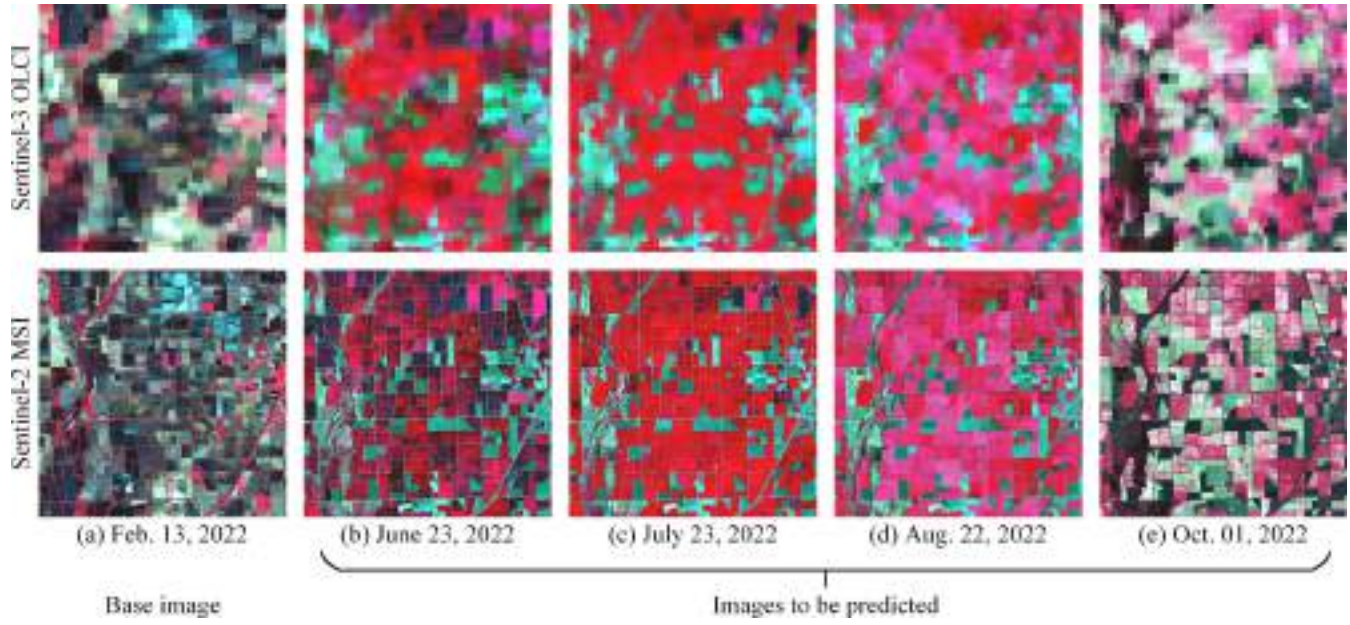

Fig. 4. Experimental data of the BC site. (a) Sentinel-2 MSI and Sentinel-3 OLCI image pair acquired on Feb. 13, 2022, (b)-(e) four Sentinel-2 MSI and Sentinel-3 OLCI image pairs acquired on June 23, 2022, July 23, 2022, Aug. 22, 2022, and Oct. 01, 2022, respectively. All images use NIR-red-green as RGB false colour composition. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

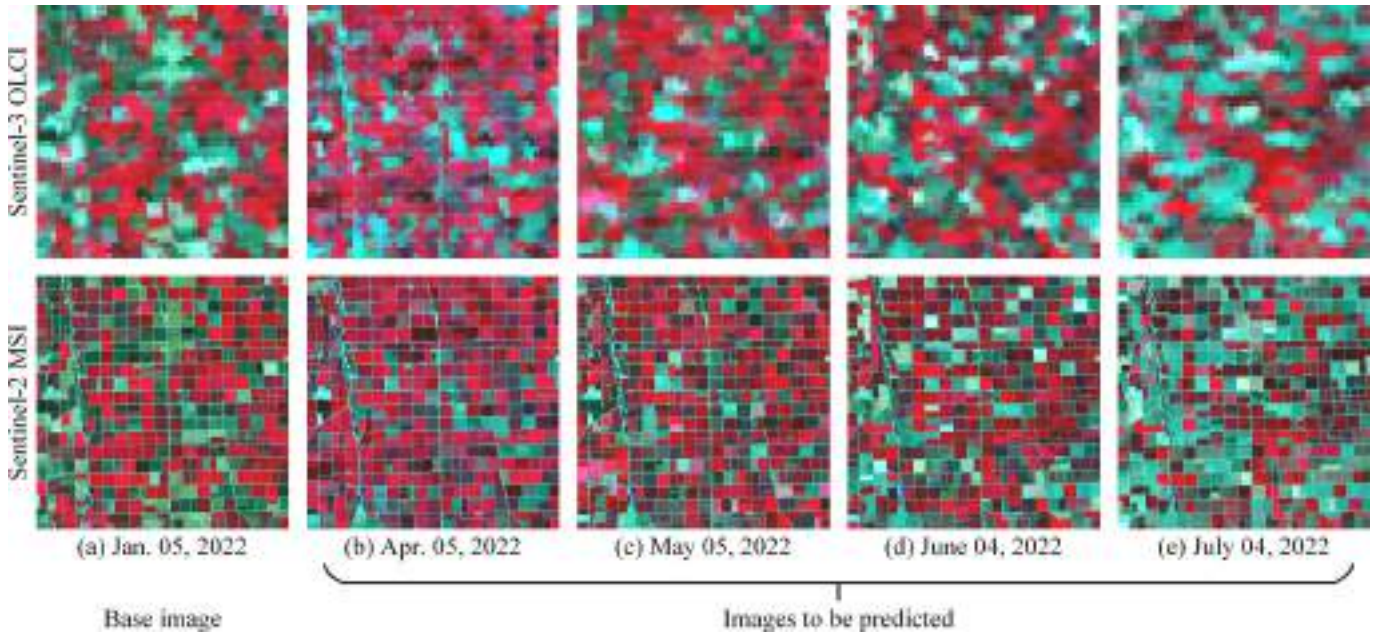

Fig. 5. Experimental data of the IC site. (a) Sentinel-2 MSI and Sentinel-3 OLCI image pair acquired on Jan. 052022, (b)-(e) four Sentinel-2 MSI and Sentinel-3 OLCI image pairs acquired on Apr. 05, 2022, May 05, 2022, June 04, 2022, and July 04, 2022, respectively. All images use NIR-red-green as RGB false colour composition. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

preprocessing was widely used in many spatiotemporal fusion studies (Chen et al., 2023; Gevaert and García-Haro, 2015). The coefficient and interception of the linear regression models were estimated using the upscaled 300 m MSI bands and the corresponding OLCI bands.

Two agricultural regions were selected as experimental sites to validate the performance of OBSUM. Both sites cover an area of 15 km × 15 km, which corresponds to 1500 × 1500 Sentinel-2 MSI pixels and 50 × 50 Sentinel-3 OLCI pixels. In order to test the fusion result over a long time-series, we collected five pairs of cloud-free Sentinel-2 MSI images and Sentinel-3 OLCI images for each site. The first Sentinel-2 MSI image is regarded as the base fine image, and the other four Sentinel-2 MSI images are considered as the images to be predicted. The time intervals between the four images to be predicted at the two experimental sites are one month.

The first site is located in southwest Butte County, California, United States of America (BC site hereafter), and the major crop type is rice. As shown in Fig. 4 (a), the land-cover is dominated by bare land and water on Feb. 13, 2022. According to the 2022 Crop Progress and Conditions report (USDA, 2022a) released by the United States Department of Agriculture (USDA), the rice in California is planted in May, emerges in June, matures in July and August, and is finally harvested in October. The phenology of rice can be clearly observed in Fig. 4 (b)-(e), so the BC

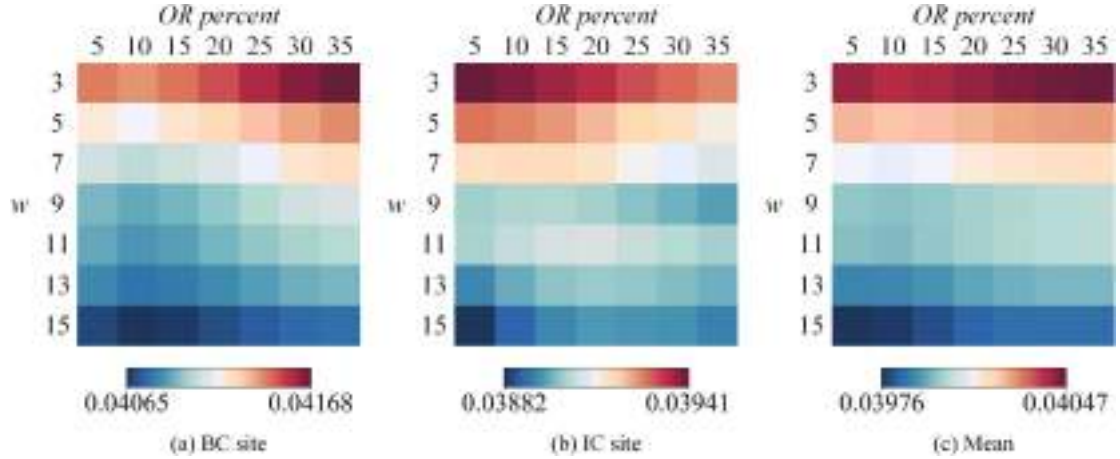

Fig. 6. RMSE of the images fused by OBSUM with different combinations of *w* and *OR percent*. (a) BC site when predicting the fine image on July 23, 2022, (b) IC site when predicting the fine image on June 04, 2022, and (c) mean value of two sites.

Table 2
Parameter settings for seven comparison methods and OBSUM.

Method	n_c	w	w_s	n_s
UBDF	5	3	N/A	N/A
STARFM	5	N/A	31	30
Fit-FC	N/A	3	31	30
FSDAF	5	N/A	31	30
RASDF	3–7	N/A	N/A	N/A
VSDF	N/A	N/A	31	30
OL-HSTFM	N/A	N/A	N/A	N/A
OBSUM	5	15	31	30

site is considered a representative region of phenological changes. The correlation coefficients between the base image and the four images to be predicted are 0.022, −0.076, −0.069, and 0.045, respectively, which indicate the strong temporal changes between the Sentinel-2 MSI images.

The second site is located in the center of Imperial County, California, United States of America (IC site hereafter), and has a mixed planting of dozens of crops. According to the Cropland Data Layer (CDL) (USDA, 2022c) provided by the USDA, the main crop types of this site are alfalfa, sugar beets, durum wheat, onions, and other hay/non alfalfa. As shown in Fig. 5 (a)-(e), the IC site has an uneven trend of land-cover changes because of its complex crop planting structure. Moreover, cropland parcels in the IC site are smaller than those in the BC site. Therefore, the IC site is considered a region with rapid land-cover changes. The correlation coefficients between the base image and the four images to be predicted are 0.579, 0.501, 0.399, and 0.424, respectively.

3.2. Experimental setup

There are two key parameters to be defined in OBSUM: the size of the local unmixing window (w) and the percentage of fine residual pixels selected for OL-RC (*OR percent*). Fig. 6 (a) and (b) show the RMSE of images fused by OBSUM with different combinations of these two parameters. For both sites, the RMSE becomes lower by increasing w. However, as *OR percent* increases, the RMSE in different sites has different trends. For the BC site, the RMSE increases with the increase in *OR percent*. For the IC site, when $w \leq 9$, the RMSE gradually decreases by increasing *OR percent*; when $w = 11$, as the *OR percent* increases, the RMSE first increases and then decreases; when $w \geq 13$, the RMSE becomes higher by increasing w. One can observe from Fig. 6 (c) that the mean RMSE of two sites gradually decreases by increasing w, and the RMSE increases as *OR percent* increases. Moreover, the lowest RMSE was obtained when $w = 15$ and *OR percent* $= 5$. Considering a larger local unmixing window will increase the computational cost with a limited accuracy improvement. Thus, in this paper, we set $w = 15$ and *OR percent* $= 5$.

Seven spatiotemporal fusion methods that require one pair of fine and coarse images are selected to compare with OBSUM, including UBDF, STARFM, Fit-FC, FSDAF, RASDF, VSDF, and OL-HSTFM. These seven methods belong to different categories: UBDF is an unmixing-based method, STARFM and Fit-FC are weight function-based methods, and the others are hybrid methods. Table 2 shows the parameter settings of the comparison methods and OBSUM, including the number of land-cover classes (n_c), the size of the local moving window (w), the size of the window for similar pixel selection (w_s), and the number of similar pixels (n_s). Please note that in Fit-FC, w refers to the size of the local window for regression model fitting (RM); while in UBDF and OBSUM, w represents the size of the local window for spatial unmixing. Through visual interpretation of the Sentinel-2 MSI images, the number of land-cover classes (n_c) is set to 5 for both sites. A more detailed discussion on the impact of n_c on the fusion results is presented in Appendix A. For UBDF and Fit-FC, the size of local widow is set to 3. In most previous spatiotemporal fusion methods, w_s was empirically set to an odd number close to the scale factor between the coarse and fine images, and n_s was set to 30 for fusing Sentinel-2 MSI and Sentinel-3 OLCI images (Erdem and Avdan, 2023; Wang and Atkinson, 2018). Therefore, in this paper we set w_s and n_s to 31 and 30, respectively.

Four indices are used for quantitative evaluation of the fusion accuracy, including average difference (AD), root mean squared error (RMSE), correlation coefficient (r), and structural similarity (SSIM) (Zhou et al., 2004). AD ranges from −1 to 1, and the optimal value is 0. A positive AD value indicates overestimation of the temporal change between the based image and the image to be predicted, whereas a negative AD value indicates underestimation of the temporal change between the two images (Zhu et al., 2022). RMSE ranges between 0 and 1, and a lower value indicates higher fusion accuracy. Both r and SSIM range between 0 and 1, and a larger value indicates higher fusion accuracy. Moreover, fusion accuracy is also evaluated visually by comparing the similarity between fused images and the reference image. Considering the large number of images to be predicted at each site, only the fine image that has the lowest r to the base fine image was selected. This strategy can effectively evaluate the performance in retrieving strong temporal change of all fusion methods. For the BC site, the fine image on July 23, 2022 was selected, and on June 04, 2022 for the IC site.

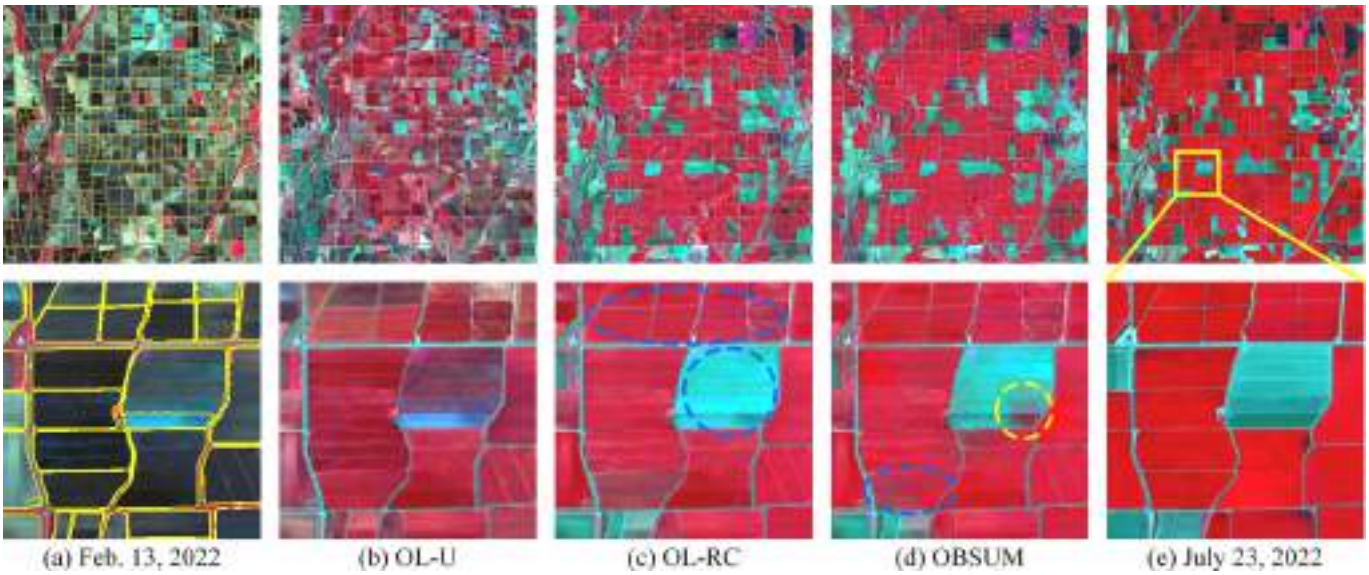

Fig. 7. Results of the three different steps of OBSUM at the BC site on July 23, 2022. (a) Base image acquired on Feb. 13, 2022 and the referred segmentation result, (b)-(d) OL-U, OL-RC, and OBSUM predictions, and (e) reference image. The second row shows the zoomed-in results of the sub-area marked in the yellow rectangle in the upper right sub-figure. The blue dashed line ellipses show the recovery of spectral information by two residual compensation steps, while the yellow dashed line ellipse highlights spectral distortion caused by geometric inconsistency between the coarse and fine images. All images use NIR-red-green as RGB false colour composition. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

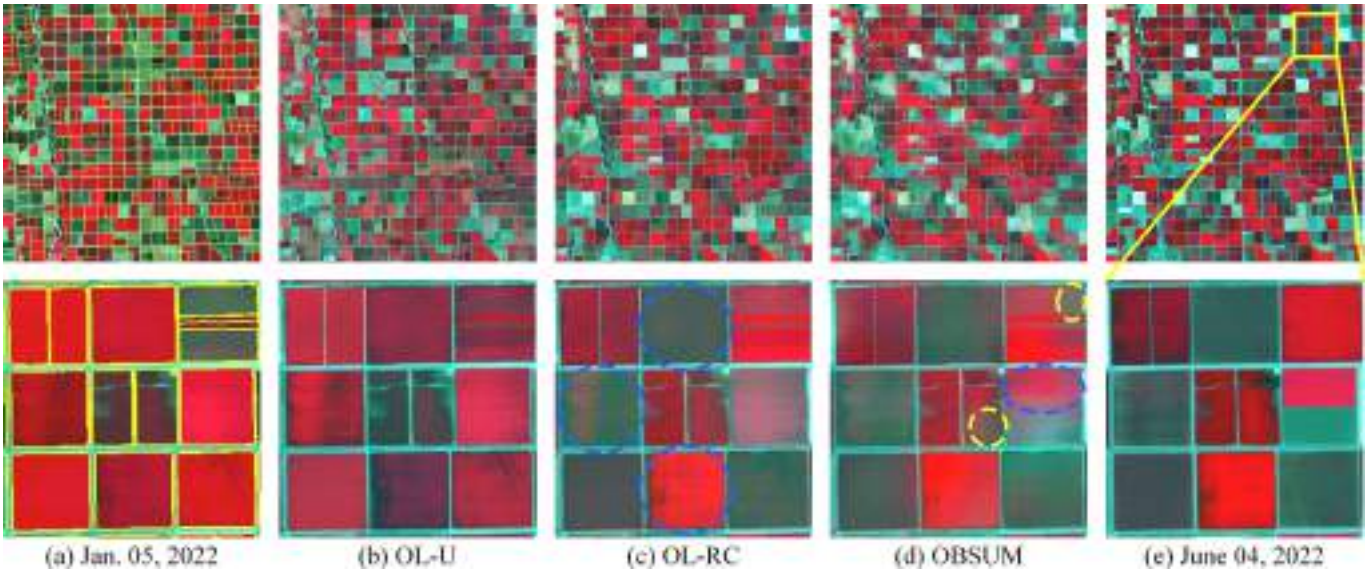

Fig. 8. Results of the three different steps of OBSUM at the IC site on June 04, 2022. (a) Base image acquired on Jan. 05, 2022 and the referred segmentation result, (b)-(d) OL-U, OL-RC, and OBSUM predictions, and (e) reference image. The second row shows the zoomed-in results of the sub-area marked in the yellow rectangle in the upper right sub-figure. The blue dashed line ellipses show the recovery of spectral information by two residual compensation steps, while the yellow dashed line ellipses highlight spectral distortion caused by geometric inconsistency between the coarse and fine images. All images use NIR-red-green as RGB false colour composition. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.3. Comparison of the three fusion steps of OBSUM

Figs. 7 and 8 show the base image, the results of the three fusion steps of OBSUM, and the reference image when predicting the fine image at the BC site on July 23, 2022 and the fine image at the IC site on June 04, 2022, respectively. In both figures, the first row shows the complete images, while the second row shows the enlarged sub-area marked in the yellow rectangle in the first row of sub-figure (e). One can see from Figs. 7 (b) and 8 (b) that there is no block effect in the images predicted by OL-U, whereas there are several spectral distortions in the OL-U predictions when compared to the reference image. These spectral distortions are mainly introduced by errors in the local unmixing process and land-cover changes between t_b and t_p. The following OL-RC step improved the fusion accuracy through computing and compensating the residual for each object. As shown in Figs. 7 (c) and 8 (c), the OL-RC predictions are visually closer to the reference images, and the spectral information in the blue dashed ellipses is recovered properly. However, the OL-RC is an object-level processing, which neglects the pixel-level spectral details and the within-object land-cover changes. After the final pixel-level residual compensation step, the OBSUM predictions are more similar to the reference images because PL-RC recovered the pixel-level spectral details and the within-object land-

Table 3
Accuracy metric values of the three fusion steps of OBSUM and accuracy improvement over the previous step. The bold values indicate the highest accuracy in each term.

Site	t_p	Metric	OL-U	OL-RC (improvement)	OBSUM (improvement)
BC	June 23, 2022	AD	−0.00021	0.00078 (−0.00057)	**0.00006** (0.00072)
		RMSE	0.05619	0.04517 (0.01102)	**0.04278** (0.00239)
		r	0.23173	0.54372 (0.31199)	**0.59799** (0.05427)
		SSIM	0.67726	0.72394 (0.04667)	**0.73042** (0.00648)
	July 23, 2022	AD	**0.00002**	0.00086 (−0.00084)	0.00029 (0.00057)
		RMSE	0.06111	0.04355 (0.01756)	**0.04069** (0.00285)
		r	0.28430	0.68450 (0.40020)	**0.73439** (0.04989)
		SSIM	0.61516	0.70633 (0.09117)	**0.71634** (0.01001)
	Aug. 22, 2022	AD	−0.00011	0.00057 (−0.00047)	**0.00008** (0.00049)
		RMSE	0.05765	0.03964 (0.01801)	**0.03686** (0.00279)
		r	0.21457	0.64435 (0.42978)	**0.70235** (0.05800)
		SSIM	0.72006	0.77641 (0.05635)	**0.78114** (0.00473)
	Oct. 01, 2022	AD	−0.00042	−0.00111 (−0.00068)	**−0.00001** (0.00109)
		RMSE	0.07391	0.05154 (0.02237)	**0.05037** (0.00117)
		r	0.41205	0.75137 (0.33932)	**0.76564** (0.01427)
		SSIM	0.68505	**0.73677** (0.05172)	0.73312 (−0.00365)
IC	Apr. 05, 2022	AD	−0.00018	0.00209 (−0.00190)	**0.00008** (0.00200)
		RMSE	0.05030	**0.03227** (0.01803)	0.03316 (−0.00088)
		r	0.68982	**0.88176** (0.19194)	0.87125 (−0.01051)
		SSIM	0.82250	**0.86656** (0.04406)	0.85882 (−0.00774)
	May 05, 2022	AD	−0.00020	0.00166 (−0.00146)	**0.00005** (0.00161)
		RMSE	0.05644	**0.04059** (0.01585)	0.04150 (−0.00091)
		r	0.61347	**0.81659** (0.20312)	0.80279 (−0.01380)
		SSIM	0.80708	**0.83104** (0.02397)	0.82516 (−0.00588)
	June 04, 2022	AD	−0.00033	0.00183 (−0.00150)	**0.00007** (0.00176)
		RMSE	0.06294	**0.03723** (0.02571)	0.03882 (−0.00159)
		r	0.58428	**0.87094** (0.28665)	0.85595 (−0.01498)
		SSIM	0.79304	**0.84136** (0.04832)	0.83085 (−0.01051)
	July 04, 2022	AD	−0.00004	0.00297 (−0.00293)	**0.00008** (0.00290)
		RMSE	0.05792	**0.03838** (0.01954)	0.03933 (−0.00095)
		r	0.55226	**0.81681** (0.26455)	0.80524 (−0.01157)
		SSIM	0.80813	**0.84243** (0.03430)	0.83591 (−0.00652)

cover changes properly. As shown in Fig. 7 (d), the object in the blue dashed ellipse is brighter than that in the OL-RC prediction and is visually closer to the reference image. In Fig. 8 (d), the within-object land-cover change in the blue dashed ellipse is recovered by PL-RC in the final OBSUM prediction. However, as shown in the yellow dashed ellipses in Figs. 7 (d) and 8 (d), the PL-RC also introduced some slight spectral distortions in the final predictions. These distortions are mainly caused by the spatial inconsistencies between the coarse image and the real fine image at t_p.

Table 3 shows the accuracies of the three fusion steps of OBSUM and the accuracy gains over the previous step when predicting all fine images. At the BC site, one can see that the fusion accuracies are gradually improved through the OL-RC and PL-RC steps. For example, when predicting the fine image at the BC site on June 23, 2022, the RMSE decreased by 0.01102 from OL-U to OL-RC, and then decreased by 0.00239 from OL-RC to OBSUM. It can also be noticed that the accuracy improvements achieved by OL-RC (gain over OL-U) are much higher than those by PL-RC (gain over OL-RC). For example, the RMSE reduction obtained by OL-RC (0.01756) is approximately five times more than that obtained by PL-RC (0.00285) when predicting the fine image at the BC site on July 23, 2022. Such a behavior can be explained by the fact that the OL-RC can improve the spectral accuracy for almost all objects in the image, whereas PL-RC is designed only for recovering within-object land-cover change, which is relatively rare. However, the OL-RC predictions at the IC site are more accurate than the final OBSUM predictions, which indicates that the PL-RC process decreased the fusion accuracy. The reason is that the IC site has a more heterogeneous land-cover, so the PL-RC process would introduce more noticeable spectral distortions, which are caused by the spatial inconsistencies between the coarse image and the real fine image at t_p.

In order to further test the performance of the three fusion steps in the ideal fusion condition, we conducted experiments with simulated Sentinel-3 OLCI-like images (obtained by upscaling the Sentinel-2 MSI images to 300 m resolution). Figs. 9 and 10 show the visual comparison of OBSUM predictions when feeding the real Sentinel-3 OLCI image and the simulated Sentinel-3 OLCI-like image at the BC and IC sites, respectively. Additionally, the accuracy metrics of the OBSUM predictions are also provided in Table 4. One can see that feeding simulated Sentinel-3 OLCI-like images produces more accurate OBSUM predictions than feeding real Sentinel-3 OLCI images. In Figs. 9 and 10, the yellow dashed ellipses show spatial inconsistencies between the Sentinel-2 MSI images and the real Sentinel-3 OLCI images, as well as the spectral distortions in the OBSUM predictions caused by those inconsistencies. In contrast, there are no such inconsistencies between the Sentinel-2 MSI images and the simulated Sentinel-3 OLCI-like images, as well as in the corresponding OBSUM predictions (see pixels inside the blue dashed ellipses). Moreover, one can also observe from Table 4 that the fusion accuracy was gradually improved by the OL-RC and PL-RC steps when feeding simulated Sentinel-3 OLCI-like images at the IC site. This indicates that the PL-RC can effectively improve fusion accuracy when the input coarse-fine image pairs are reliable.

Both the visual and quantitative evaluations indicate that all three fusion steps in OBSUM are indispensable, and the two residual compensation steps can recover object-level spectral information and within-object land-cover changes, respectively. However, the PL-RC process is sensitive to geometric errors between the coarse image and the fine image. The proposed OL-RC process is further discussed in Section 3.5.

3.4. Comparison with other methods

Figs. 11 and 12 show the fusion results of seven comparison methods and OBSUM, and the reference images on four prediction dates at the BC site and the IC site, respectively. For convenience, only the fusion results on July 23, 2022 at the BC site and the fusion results on June 04, 2022 at the IC site were selected for visual comparison, i.e., Figs. 11 (b) and 12 (c).

Fig. 13 shows the results of different spatiotemporal fusion methods at the BC site on July 23, 2022, and Fig. 14 shows the enlarged sub-areas marked in the yellow rectangle in Fig. 13 (j). One can observe from Fig. 13 that almost all methods predicted a fused image similar to the

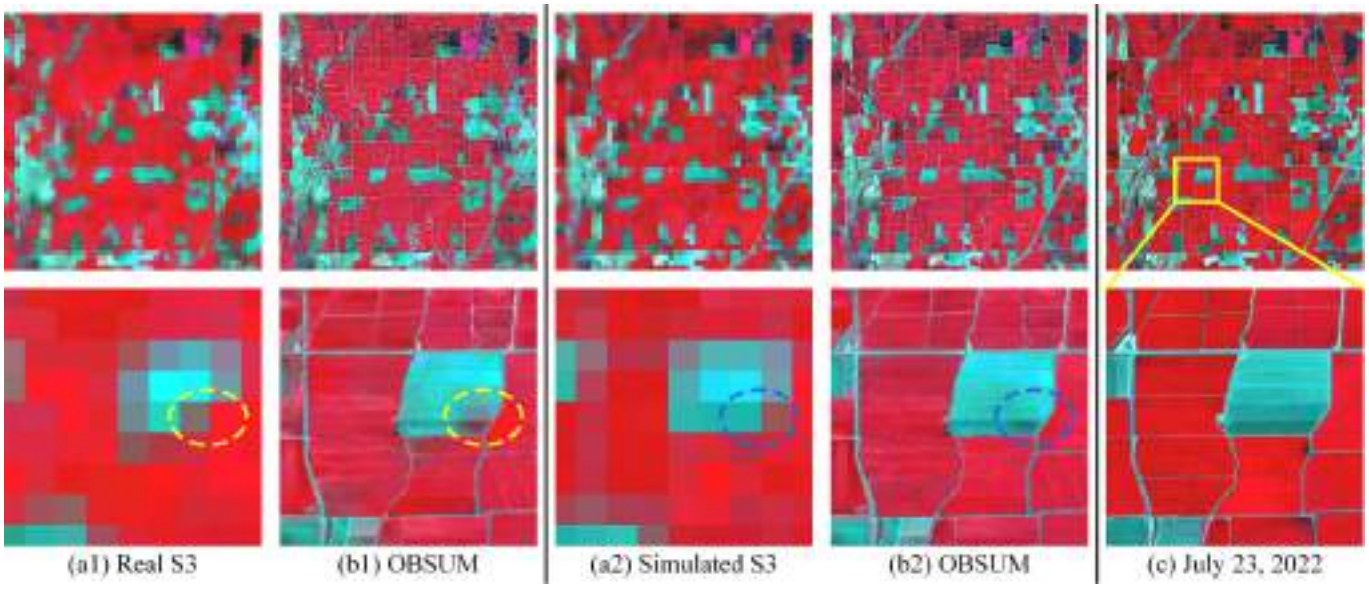

Fig. 9. Fusion results of OBSUM at the BC site on July 23, 2022 when feeding the real Sentinel-3 OLCI image and the simulated Sentinel-3 OLCI-like image. (a1) and (b1) show the real Sentinel-3 OLCI image and the fusion result, (a2) and (b2) present the simulated Sentinel-3 OLCI-like image and the fusion result, and (c) is the reference image. The second row shows the zoomed-in results of the sub-area marked in the yellow rectangle in the upper right sub-figure. The yellow dashed line ellipses highlight spatial inconsistency between the real Sentinel-2 MSI and Sentinel-3 OLCI images, and the spectral distortion in the OBSUM prediction. The blue dashed line ellipses show the spatial consistency between the Sentinel-2 MSI image and the simulated Sentinel-3 OLCI-like image, and the accurate OBSUM prediction. All images use NIR-red-green as RGB false colour composition. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

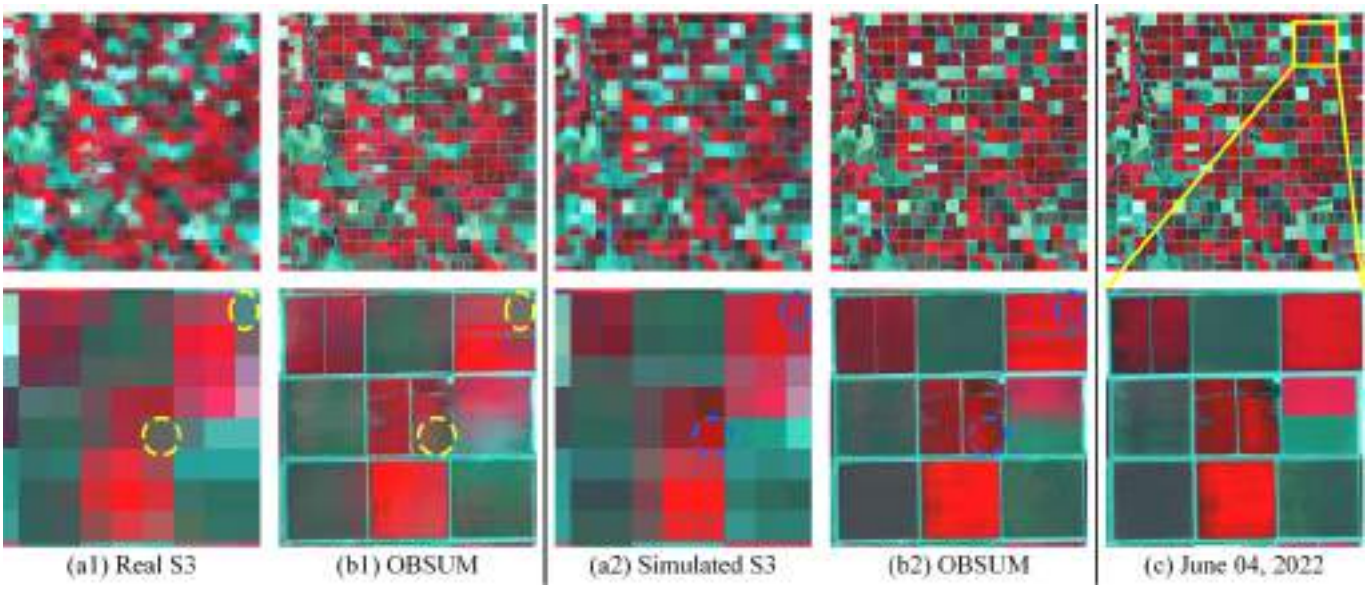

Fig. 10. Fusion results of OBSUM at the IC site on June 04, 2022 when feeding the real Sentinel-3 OLCI image and the simulated Sentinel-3 OLCI-like image. (a1) and (b1) show the real Sentinel-3 OLCI image and the fusion result, (a2) and (b2) present the simulated Sentinel-3 OLCI-like image and the fusion result, and (c) is the reference image. The second row shows the zoomed-in results of the sub-area marked in the yellow rectangle in the upper right sub-figure. The yellow dashed line ellipses highlight spatial inconsistency between the real Sentinel-2 MSI and Sentinel-3 OLCI images, and the spectral distortion in the OBSUM prediction. The blue dashed line ellipses show the spatial consistency between the Sentinel-2 MSI image and the simulated Sentinel-3 OLCI-like image, and the accurate OBSUM prediction. All images use NIR-red-green as RGB false colour composition. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

reference image and captured the strong temporal change between Feb. 13, 2022 and July 23, 2022. In Fig. 14 (b), the UBDF prediction contains clear block effect and spectral distortions with the footprints of the classified fine pixels. Taking advantage of the classification map refinement and the object-level unmixing, these spectral distortions are not present in the OBSUM prediction. Moreover, the UBDF prediction also lacks spatial details. One can also notice that STARFM, FSDAF, RASDF, and VSDF failed to accurately predict the temporal change of the croplands (see the gray and white pixels in the yellow dashed ellipses). As shown in Fig. 14 (d), Fit-FC wrongly estimated the reflectance of roads (see the light pink pixels in the yellow dashed ellipses), and the prediction is over-smoothed. Moreover, there are some obvious salt-and-

Table 4
Accuracy metric values of the three fusion steps of OBSUM and accuracy gains over the previous step at the BC site on July 23, 2022, at the IC site on June 04, 2022, when feeding the real Sentinel-3 OLCI image and the simulated Sentinel-3 OLCI-like image. The bold values indicate the highest accuracy in each term.

Site	Mode	Metric	OL-U	OL-RC (improvement)	OBSUM (improvement)
BC	Real	AD	**0.00002**	0.00086 (−0.00084)	0.00029 (0.00057)
		RMSE	0.06111	0.04355 (0.01756)	**0.04069** (0.00285)
		r	0.28430	0.68450 (0.40020)	**0.73439** (0.04989)
		SSIM	0.61516	0.70633 (0.09117)	**0.71634** (0.01001)
	Simulated	AD	0.00041	0.00079 (−0.00038)	**0.00036** (0.00043)
		RMSE	0.05853	0.03850 (0.02003)	**0.03302** (0.00548)
		r	0.39889	0.77119 (0.37230)	**0.83740** (0.06621)
		SSIM	0.63478	0.73000 (0.09522)	**0.75275** (0.02275)
IC	Real	AD	−0.00033	0.00183 (−0.00150)	**0.00007** (0.00176)
		RMSE	0.06294	**0.03723** (0.02571)	0.03882 (−0.00159)
		r	0.58428	**0.87094** (0.28665)	0.85595 (−0.01498)
		SSIM	0.79304	**0.84136** (0.04832)	0.83085 (−0.01051)
	Simulated	AD	−0.00052	−0.00035 (0.00017)	**−0.00001** (0.00034)
		RMSE	0.06195	0.02902 (0.03294)	**0.02624** (0.00277)
		r	0.59173	0.92386 (0.33213)	**0.93853** (0.01467)
		SSIM	0.78768	0.86504 (0.07736)	**0.86907** (0.00403)

pepper noise-like spectral distortions in the OL-HSTFM prediction in Fig. 14 (h). By contrast, OBSUM successfully retrieved the temporal change and obtained the highest visual accuracy. The reason is that OBSUM compensates the residuals at both the object-level and pixel-level, and these two residual compensation steps can capture strong temporal change and recover the spectral details properly.

Fig. 15 shows the results of different spatiotemporal fusion methods at the IC site on June 04, 2022, while Fig. 16 shows the enlarged sub-areas marked in the yellow rectangle in Fig. 15 (j). One can observe from Fig. 15 (b) that UBDF underestimated the temporal change of croplands, and the prediction is inaccurate. Moreover, there are clear block effects in the UBDF prediction in Fig. 16 (b), as well as the spectral distortions in the yellow dashed ellipses. The complete fused images of the other seven methods are similar to the reference image. However, STARFM, FSDAF, RASDF, VSDF, and OL-HSTFM failed to predict the reflectance of small cropland parcels (see the black and gray pixels marked in yellow dashed ellipses). In Fig. 16 (d), the Fit-FC prediction in the yellow dashed ellipse is lighter than its counterpart in the reference image, which indicates that the temporal change was wrongly estimated. Moreover, the Fit-FC prediction seems to be over-smoothed and lacks spatial details compared to the reference image (see the colour difference of roads in the predicted image and the reference image). There are also some salt-and-pepper noise-like spectral distortions in the OL-HSTFM prediction in Fig. 16 (h). Taking advantage of object-level processing and two residual compensation steps, OBSUM recovered the land-cover change in the fused image and obtained the most accurate prediction among these eight spatiotemporal fusion methods. In the OBSUM prediction in Fig. 16 (i), the spatial details are clearly presented, and the colors of ground objects are similar to their counterparts in the reference image.

Since the land-cover at both BC and IC sites is dominated by highly vegetated croplands, we further tested the accuracy of all fusion methods by comparing the normalized difference vegetation index (NDVI) calculated by the fused images (in Figs. 13 and 15) with that derived by the reference image. The scatter plots of NDVI of the predicted images (X axis) and NDVI of the reference images (Y axis), along

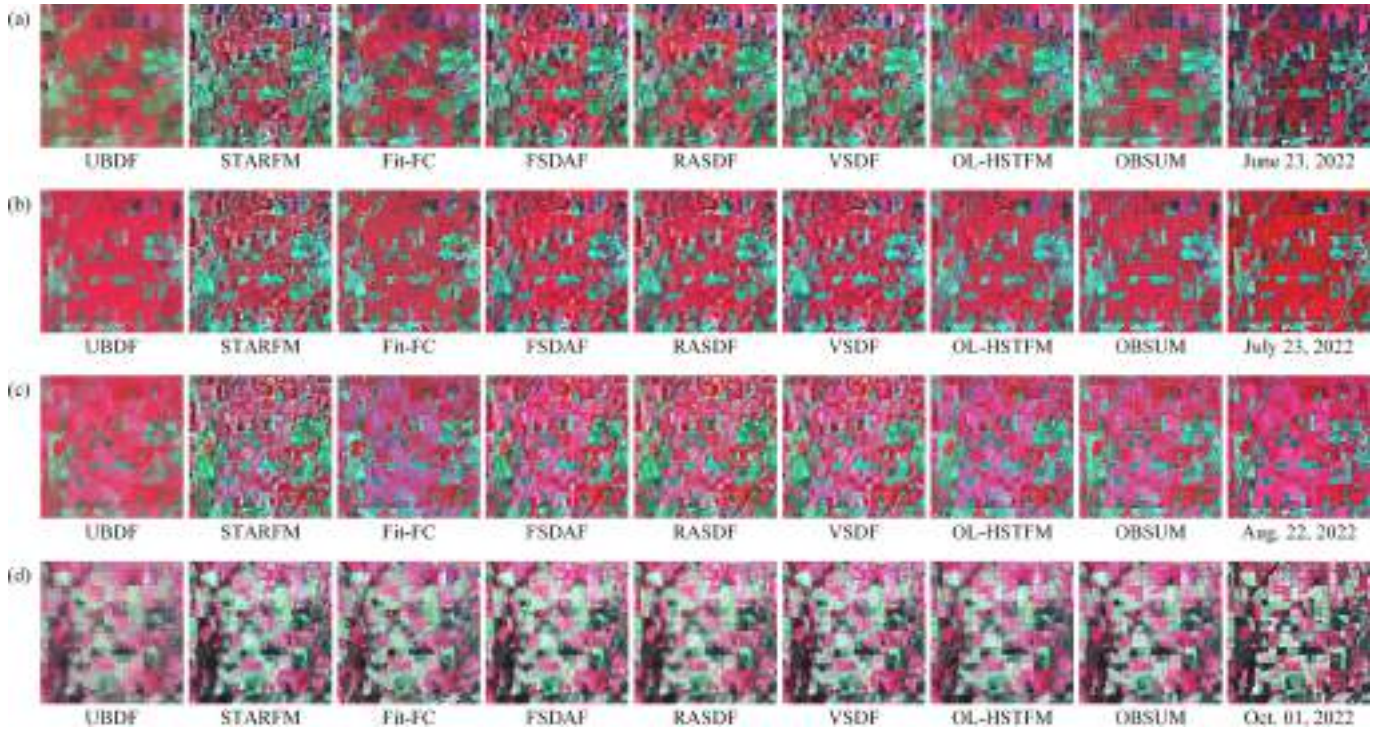

Fig. 11. Results of different spatiotemporal fusion methods on four prediction dates and the reference images at the BC site. All images use NIR-red-green as RGB false colour composition. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

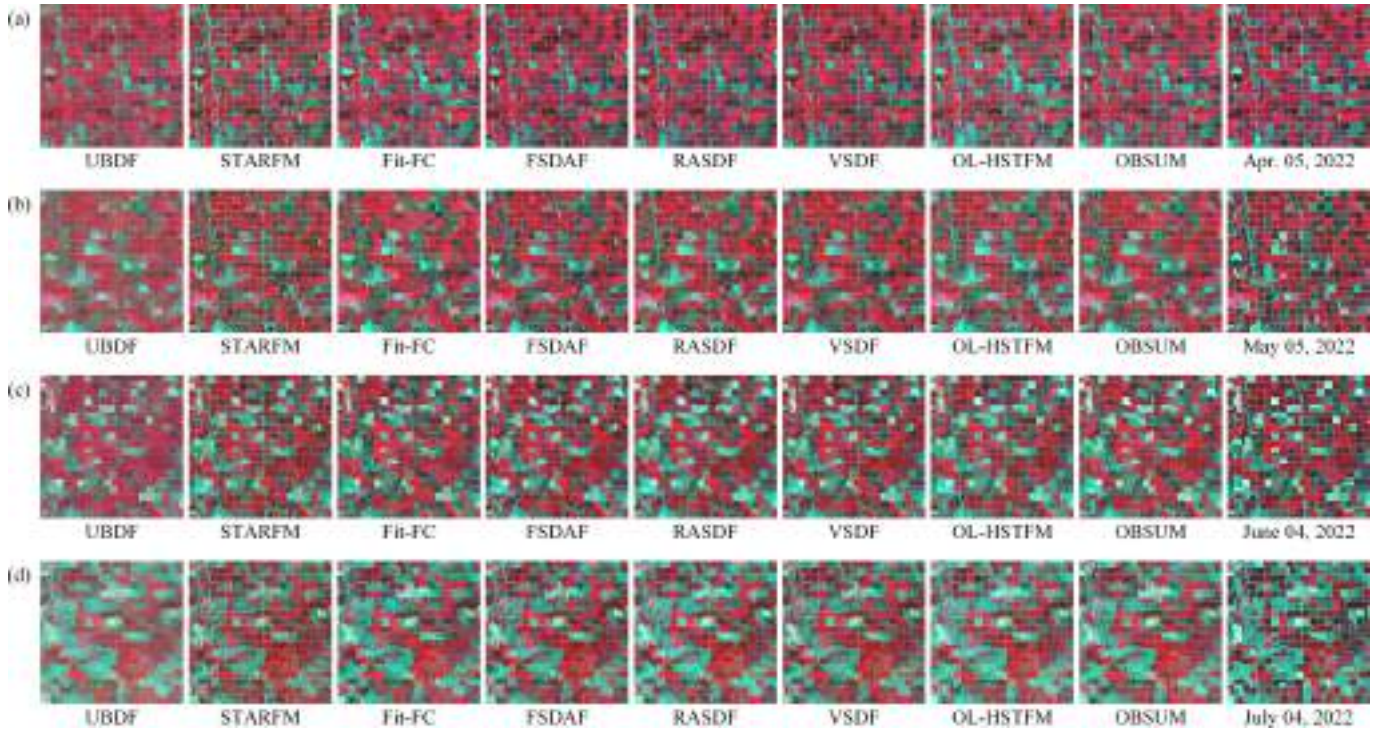

Fig. 12. Results of different spatiotemporal fusion methods on four prediction dates and the reference images at the IC site. All images use NIR-red-green as RGB false colour composition. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

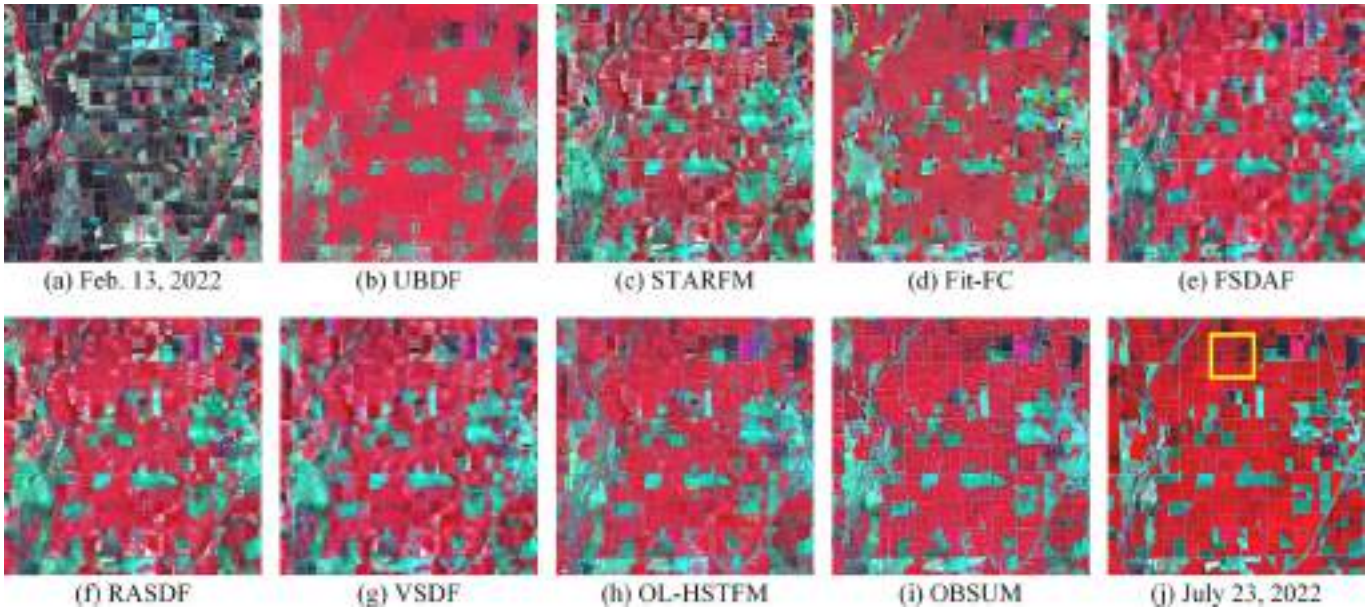

Fig. 13. Results of different spatiotemporal fusion methods at the BC site on July 23, 2022. (a) Base image acquired on Feb. 13, 2022, (b)-(i) fusion results of UBDF, STARFM, Fit-FC, FSDAF, RASDF, VSDF, OL-HSTFM, and OBSUM, and (j) reference image. All images use NIR-red-green as RGB false colour composition. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

with the statistical accuracies at the BC and IC sites, are shown in Figs. 17 and 18, respectively. In each subfigure, the red dashed line represents the 1:1 line. One can see that the NDVI calculated by the OBSUM predictions have the lowest RMSE and highest r over the comparison methods, which indicates the high accuracy of OBSUM-derived NDVI.

Table 5 gives the accuracies of different spatiotemporal fusion methods on four prediction dates at both sites and the mean accuracy over the time-series. Since a positive and a negative AD at two prediction dates could cancel each other out, the mean AD value over the time-series is not reported for method comparison. One can observe in Table 5 that in most fusion tasks OBSUM outperformed the other methods in terms of four accuracy indices, and obtained the highest average fusion accuracy over the time-series. Over the BC site, only OBSUM achieved a mean RMSE <0.045 and a mean r larger than 0.700. Over the IC site, only OBSUM achieved a mean RMSE <0.040, and the

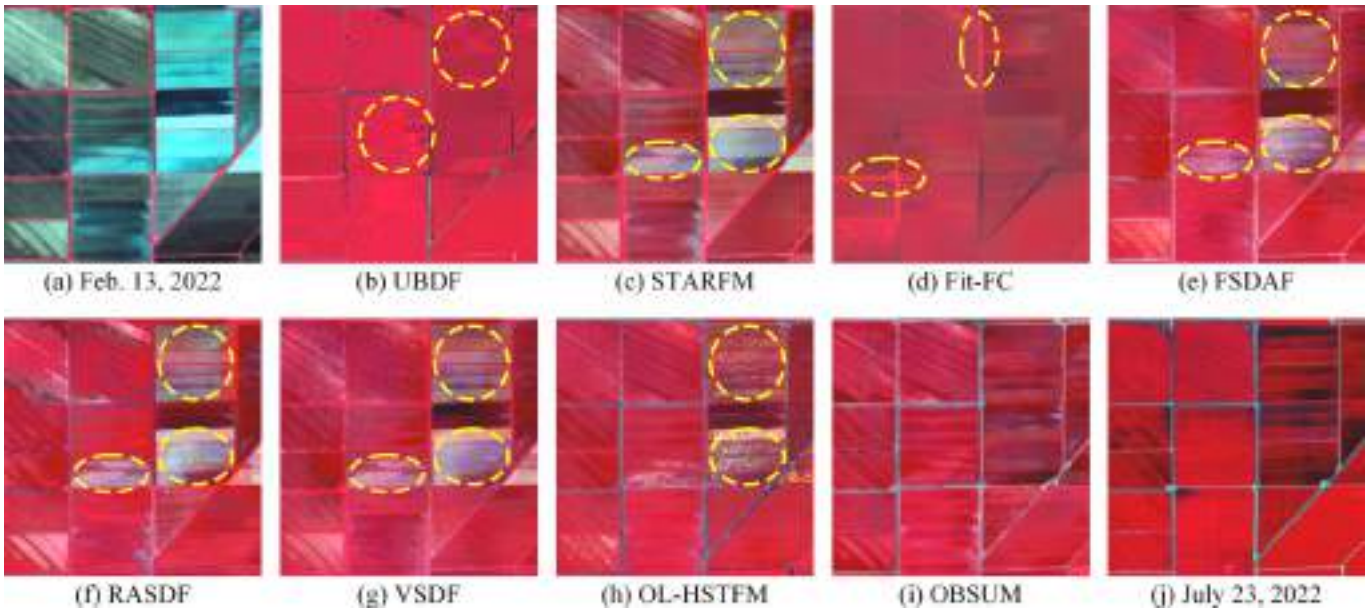

Fig. 14. Zoomed-in results of the sub-area marked in the yellow rectangle in Fig. 13 (j). The ellipses represented by the yellow dashed line highlight spectral distortions in the fused images. All images use NIR-red-green as RGB false colour composition. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

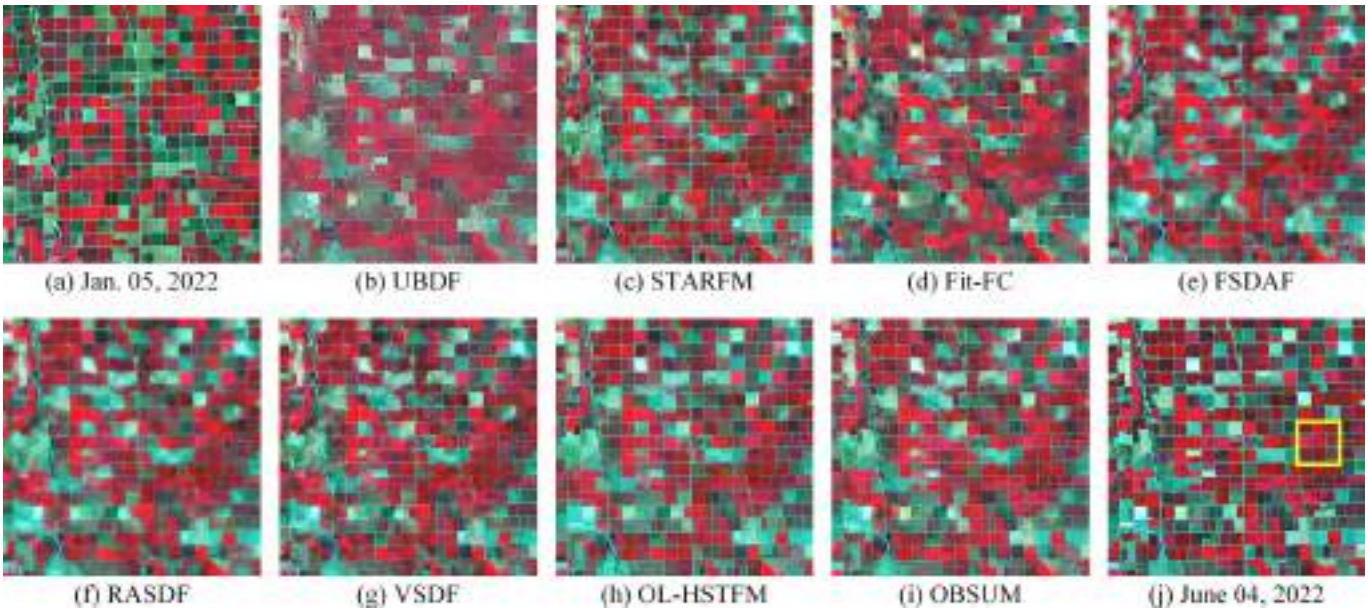

Fig. 15. Results of different spatiotemporal fusion methods at the IC site on June 04, 2022. (a) Base image acquired on Jan. 05, 2022, (b)-(i) fusion results of UBDF, STARFM, Fit-FC, FSDAF, RASDF, VSDF, OL-HSTFM, and OBSUM, and (j) reference image. All images use NIR-red-green as RGB false colour composition. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

mean r and mean SSIM of OBSUM are superior to those of the other methods. Both the visual and quantitative evaluations indicate that OBSUM can retrieve strong temporal changes and recover land-cover changes of ground objects, thereby obtaining the highest fusion accuracy in the comparison.

3.5. *Effectiveness of the OL-RC and PL-RC steps*

The experiment in Section 3.3 demonstrates the effectiveness of both the OL-RC and PL-RC. In this section, we further discuss the mechanism and explain the effectiveness of these two residual compensation steps. Figs. 19 and 20 show different residual maps of the NIR band when predicting the fine image on July 23, 2022 at the BC site and the fine image on June 04, 2022 at the IC site, respectively.

As illustrated in Section 2.3, the coarse residuals are downscaled using a bi-cubic interpolation to get the fine residuals to improve the robustness of OL-RC in heterogeneous areas. However, as shown in subfigure (b) of Figs. 19 and 20, the fine residuals are over-smoothed by the interpolation and have poor structural information of the ground objects. The objective of OL-RC is to calculate the residual for each object by combining the fine residuals and the segmentation result, thus reconstructing a residual map that contains rich structural information of the ground objects. With the guidance of object residual index (*ORI*), the OL-RC selects the most reliable fine residual pixels and combines the

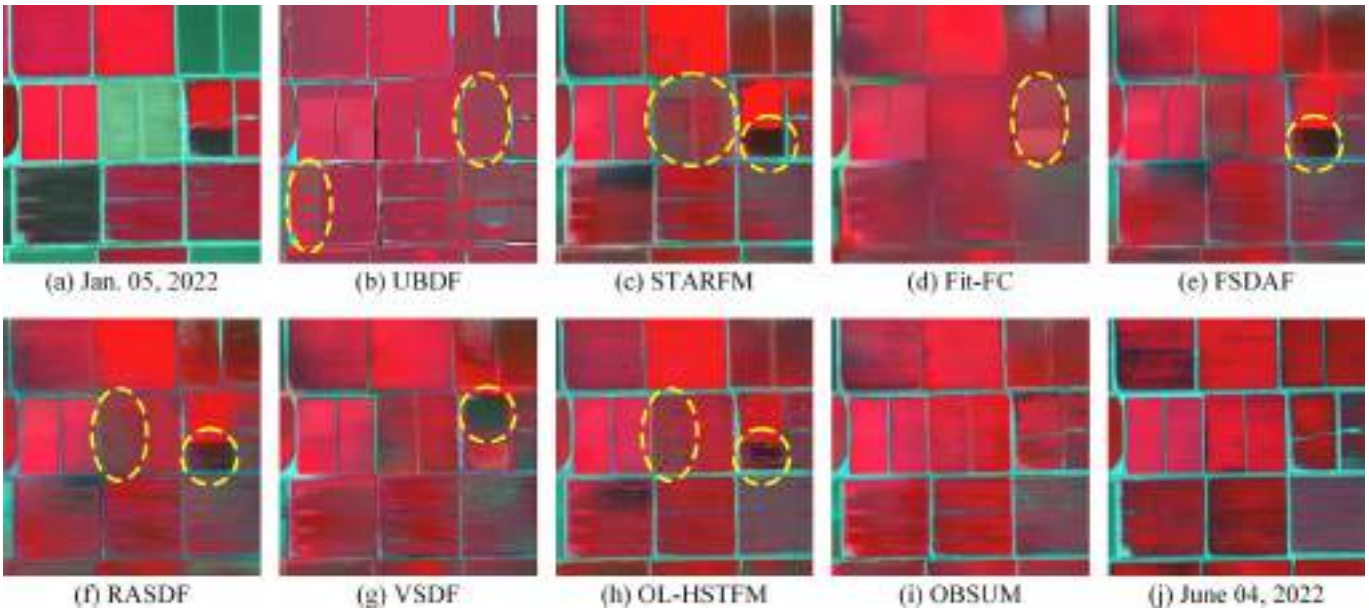

Fig. 16. Zoomed-in results of the sub-area marked in the yellow rectangle in Fig. 15 (j). The ellipses represented by the yellow dashed line highlight spectral distortions in the fused images. All images use NIR-red-green as RGB false colour composition. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

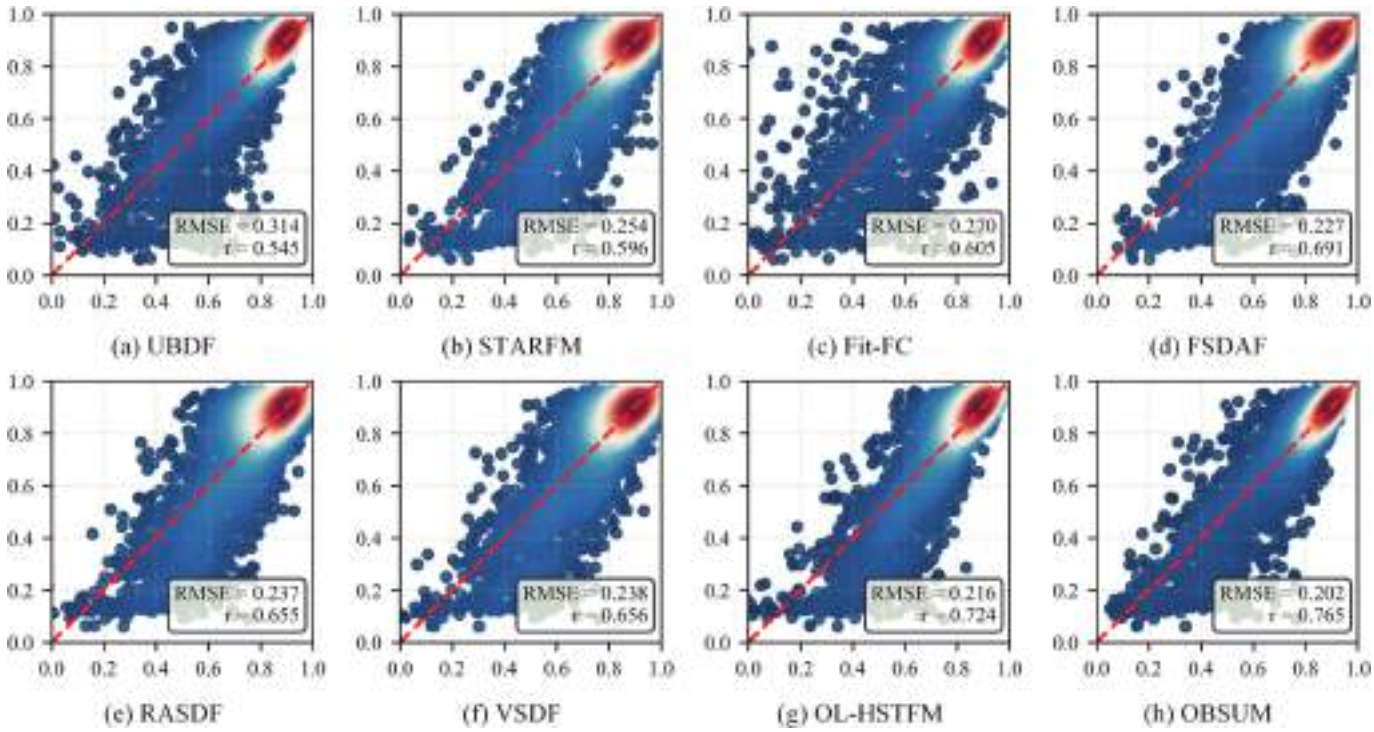

Fig. 17. Scatter plots of the NDVI calculated from the images fused by different methods and the NDVI calculated from the reference image at the BC site on July 23, 2022.

selected residuals to calculate and compensate the residual for each object. The residual map produced by OL-RC (object-level residuals, OL-R) is actually an approximation of the object residuals (mean value of actual residuals in each object). As shown in sub-figure (c) of Figs. 19 and 20, the OL-R predicted by OL-RC are very similar to the object residuals and contain rich structural information of the ground objects. Moreover, as shown in Table 6, the OL-R has the highest similarity (r = 0.87213 at the BC site, 0.91362 at the IC site) to the object residuals compared to other types of residuals.

OL-RC assumes all fine pixels within an object have the same residual, which neglects the pixel-level spectral details of the ground objects. Moreover, the within-object land-cover changes will also reduce the accuracy of OL-RC. Therefore, the PL-RC is adopted to improve the estimation of residuals by introducing neighborhood information for each target fine pixel. By adding the PL-RC predicted residuals (PL-R) to OL-R, the predicted total residual map (OL-R + PL-R) is actually an approximation of the actual residuals, i.e., difference between the OL-U prediction and the real fine image at t_p. As shown in sub-figure (h) of

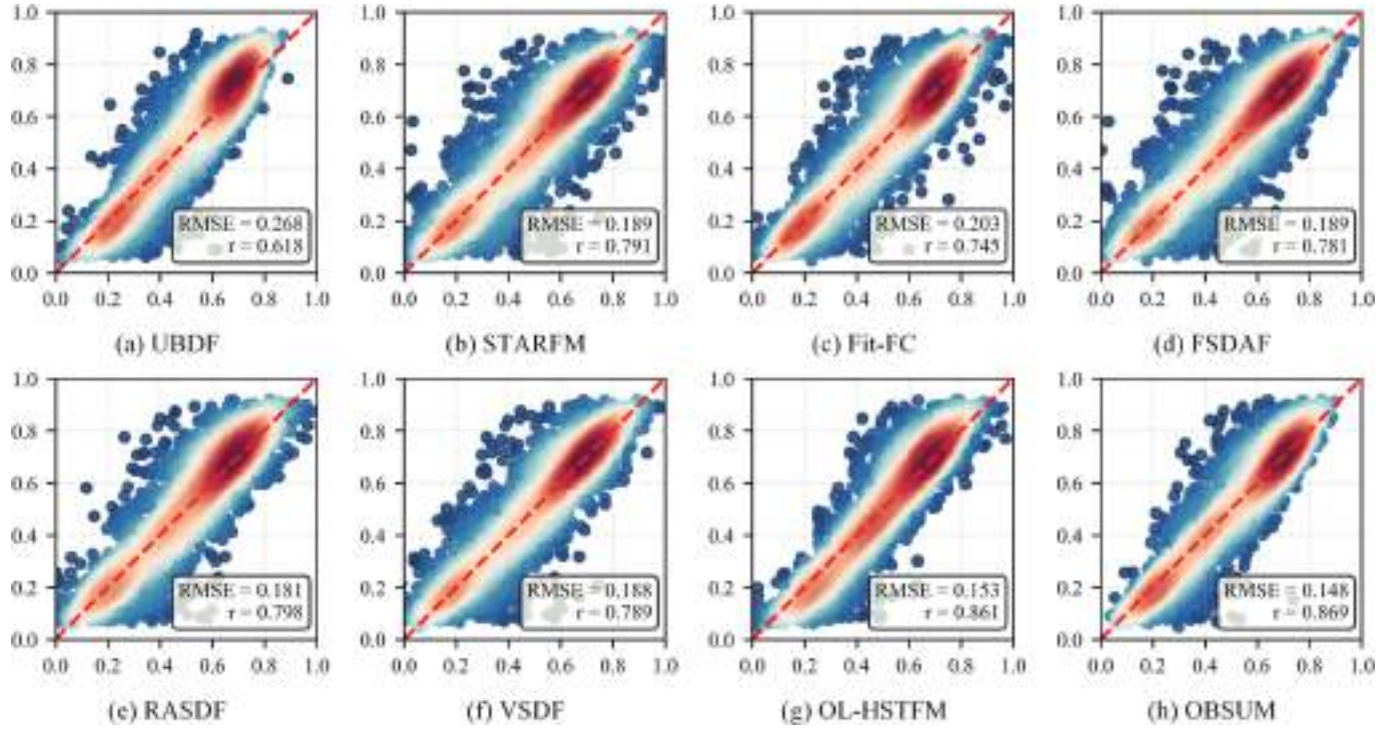

Fig. 18. Scatter plots of the NDVI calculated from the images fused by different methods and the NDVI calculated from the reference image at the IC site on June 04, 2022.

Figs. 19 and 20, the OL-R + PL-R contain more pixel-level spectral information than the OL-R, and is more similar to the actual residuals. Moreover, as shown in Table 6, the OL-R + PL-R has the highest similarity to the actual residuals compared to other types of residuals at the BC site. It could also be noticed that the similarities between the OL-R + PL-R and the actual residuals are lower than those between the OL-R and the object residuals (0.74941 versus 0.87213 at the BC site, 0.82509 versus 0.91362 at the IC site). This is because the pixel-level residuals have a finer spatial scale and are more complex than the object-level residuals, therefore, it is more difficult to predict the PL-R from the over-smoothed fine residual map.

4. Examples of agricultural applications supported by OBSUM

As illustrated in Sections 3.4 and 3.5, OBSUM can retrieve strong temporal changes and recover land-cover changes properly. Therefore, OBSUM has great potential to support various agricultural remote sensing applications. Crop progress monitoring (Gao et al., 2017; Gao et al., 2015) and crop mapping (Belgiu and Csillik, 2018; Gu et al., 2023) are two important application scenarios of spatiotemporal fusion, and the capabilities of OBSUM in these two applications are discussed below.

4.1. Crop progress monitoring

Monitoring the crop progress through the growing season is of great significance for precision agricultural management and understanding vegetation responses to climate change (Zhang et al., 2003). Spatiotemporal fusion provides a solution to improve both the spatial and temporal resolution of crop progress monitoring. We collected 349 valid MOD09GA surface reflectance records and 17 cloud-free Landsat 9 Collection 2 Level-2 surface reflectance observations of the BC site in 2022. The MOD09GA images were first reprojected using the MODIS Conversion Toolkit (MCTK), and then resampled to the spatial resolution of 480 m. Moreover, the 480 m cloud masks of MOD09GA images were also extracted and then used to mask the observations contaminated by clouds in the following fusion process. For each MOD09GA image, the Landsat 9 image nearest in time to it was selected as the base image for predicting the 30 m Landsat-like image using OBSUM. After that, the fused image was used to calculate the Enhanced Vegetation Index (EVI) for monitoring the progress of rice. Compared to other vegetation indices, EVI is more sensitive to the high biomass and varies greatly throughout the growth period, which could benefit the monitoring of crop progress (Zhao et al., 2023). Considering the gaps in the OBSUM predictions and the EVI maps that were caused by cloud contamination in the MOD09GA observations, we applied the vegetation index Savitzky-Golay filter (Chen et al., 2004) to reconstruct a high-quality EVI time-series. Fig. 21 illustrates the spatiotemporal fusion of the Landsat image and the cloudy MOD09GA image using OBSUM and the reconstruction of the gap-free EVI map at day of year (DOY) 193, 2022.

The USDA's 30 m CDL covering the BC site in 2022 was collected for extracting the rice pixels in the fused Landsat-like images. After that, the mean EVI value of the rice pixels was calculated for each DOY. The EVI values were then analyzed by the TIMESAT (Jonsson and Eklundh, 2002) software to extract the seasonality parameters, including the start of season (SOS), middle of season (DOY of maximum EVI, i.e., MAX), end of season (EOS), and duration of season (DUR). A threshold of 0.1 on the EVI amplitude was adopted to extract the start of season and end of season. Fig. 22 shows the EVI curve smoothed by the double logistic function and the extracted seasonality parameters. According to the TIMESAT analysis, the growing season of rice at the BC site in 2022 started on May 24 (DOY 144), reached its peak at Aug. 01 (DOY 213), ended at Oct. 27 (DOY 300), and the duration of season was 156 days. We also collected the Crop Progress (CP) (USDA, 2022b)report released by the USDA as the reference for crop progress. The state-level CP report is publicly available at a weekly frequency. According to the CP report, by May 23 (DOY 143), 90% of the rice in California had been planted, and 30% of the rice had been emerged. By Aug. 01 (DOY 213), 60% of the rice in California had been headed, with 65% in good condition and 30% in excellent condition. By Oct. 24 (DOY 297), 75% of the rice in California had been harvested, and 90% by Oct. 30 (DOY 303). The growth progress of rice obtained by OBSUM plus TIMESAT is generally consistent with those of the CP report. In addition, we collected the

Table 5
Accuracy metric values of different spatiotemporal fusion methods. The bold values indicate the highest accuracy in each term.

Site	t_p	Metric	Method							
			UBDF	STARFM	Fit-FC	FSDAF	RASDF	VSDF	OL-HSTFM	OBSUM
BC	June 23, 2022	AD	−0.00011	−0.00019	0.00034	0.00043	0.00021	0.00126	−0.00050	**0.00006**
		RMSE	0.04858	0.05318	0.05340	0.04639	0.04602	0.04817	0.04468	**0.04278**
		r	0.46774	0.36127	0.42013	0.48938	0.49290	0.44993	0.51070	**0.59799**
		SSIM	0.68775	0.65556	0.70935	0.71414	0.70859	0.67607	0.70382	**0.73042**
	July 23, 2022	AD	**−0.00008**	−0.00026	−0.00016	0.00062	0.00049	0.00155	−0.00038	0.00029
		RMSE	0.05201	0.05596	0.06107	0.04660	0.04994	0.04923	0.04368	**0.04069**
		r	0.57381	0.46852	0.50626	0.62366	0.53997	0.57750	0.66038	**0.73439**
		SSIM	0.66137	0.57389	0.64636	0.67693	0.62419	0.62786	0.68456	**0.71634**
	Aug. 22, 2022	AD	**0.00006**	−0.00037	−0.00027	0.00042	0.00025	0.00181	−0.00059	0.00008
		RMSE	0.04884	0.05204	0.05660	0.04334	0.04682	0.04568	0.03988	**0.03686**
		r	0.51956	0.42733	0.51020	0.55751	0.46568	0.52530	0.62255	**0.70235**
		SSIM	0.72589	0.69193	0.75178	0.76031	0.70442	0.72403	0.76313	**0.78114**
	Oct. 01, 2022	AD	0.00033	−0.00049	−0.00019	0.00054	0.00021	0.00246	−0.00017	**−0.00001**
		RMSE	0.06849	0.06284	0.07480	0.05715	0.05624	0.05954	0.05332	**0.05037**
		r	0.56910	0.63800	0.54281	0.69167	0.69817	0.67475	0.72048	**0.76564**
		SSIM	0.60385	0.67456	0.65544	0.69886	0.69534	0.66961	0.71546	**0.73312**
	Mean	RMSE	0.05448	0.05600	0.06147	0.04837	0.04975	0.05066	0.04539	**0.04267**
		r	0.53255	0.47378	0.49485	0.59056	0.54918	0.55687	0.62853	**0.70009**
		SSIM	0.66971	0.64898	0.69074	0.71256	0.68314	0.67439	0.71674	**0.74025**
IC	Apr. 05, 2022	AD	−0.00020	−0.00029	**0.00007**	0.00042	−0.00017	0.00075	−0.00114	0.00008
		RMSE	0.05722	0.04137	0.04855	0.04246	0.03790	0.03836	0.03594	**0.03316**
		r	0.59722	0.81448	0.67159	0.79031	0.83107	0.83295	0.85277	**0.87125**
		SSIM	0.69530	0.82818	0.75665	0.80941	0.82402	0.82684	0.81556	**0.85882**
	May 05, 2022	AD	−0.00010	−0.00034	0.00021	0.00032	−0.00024	0.00081	−0.00121	**0.00005**
		RMSE	0.06114	0.04734	0.05381	0.04752	0.04682	0.04484	0.04360	**0.04150**
		r	0.54759	0.76081	0.60814	0.74217	0.74716	0.76828	0.78568	**0.80279**
		SSIM	0.67711	0.82154	0.73996	0.79375	0.78360	0.80255	0.78290	**0.82516**
	June 04, 2022	AD	−0.00010	−0.00050	0.00103	0.00030	−0.00029	0.00077	−0.00081	**0.00007**
		RMSE	0.06438	0.04969	0.05733	0.04889	0.04751	0.04785	0.04162	**0.03882**
		r	0.61659	0.77450	0.67476	0.77092	0.78155	0.78615	0.83598	**0.85595**
		SSIM	0.66078	0.81147	0.73013	0.77689	0.76274	0.78157	0.79028	**0.83085**
	July 04, 2022	AD	**−0.00004**	−0.00039	0.00033	0.00044	−0.00016	0.00091	−0.00101	0.00008
		RMSE	0.05855	0.04888	0.05097	0.04656	0.04434	0.04562	0.04031	**0.03933**
		r	0.56336	0.72866	0.66037	0.73081	0.75223	0.74665	0.79871	**0.80524**
		SSIM	0.68492	0.82312	0.77319	0.80717	0.80700	0.79937	0.80594	**0.83591**
	Mean	RMSE	0.06032	0.04682	0.05267	0.04636	0.04414	0.04417	0.04037	**0.03820**
		r	0.58119	0.76961	0.65372	0.75855	0.77800	0.78351	0.81829	**0.83381**
		SSIM	0.67952	0.82108	0.74998	0.79680	0.79434	0.80258	0.79867	**0.83768**

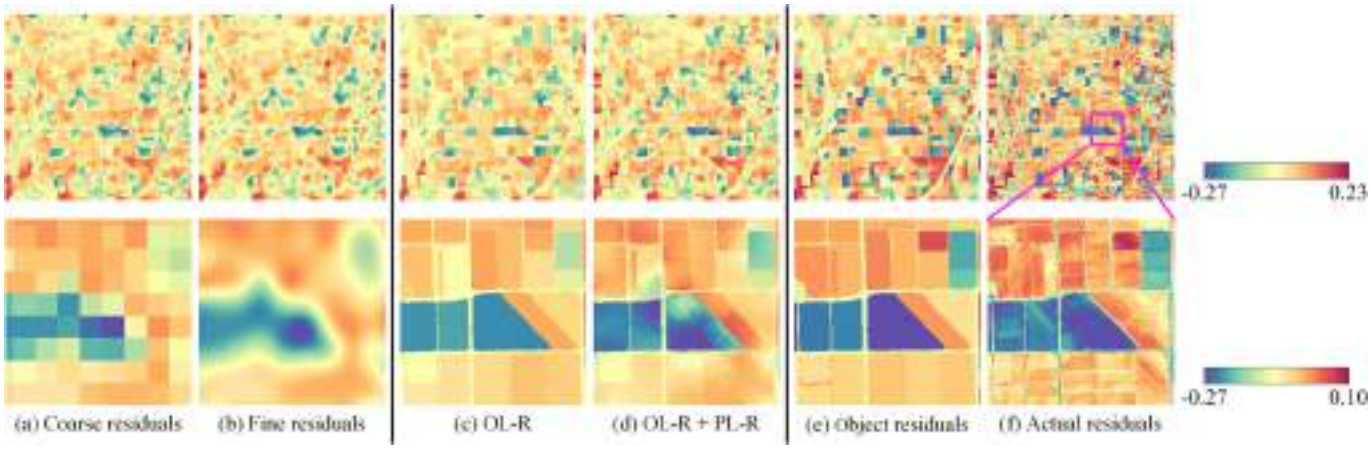

Fig. 19. Different types of residuals of the NIR band at the BC site when predicting the Sentinel-2 MSI image on July 23, 2022. (a) Coarse residuals, (b) fine residuals, (c) OL-R, (d) OL-R + PL-R, (e) object residuals, and (f) actual residuals. The second row shows the zoomed-in results of the sub-area marked in the magenta rectangle in the upper right sub-figure. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

eVIIRS phenology data (USGS, 2022) released by the United States Geological Survey (USGS), which provides SOS, MAX, EOS, and DUR seasonality parameters with a spatial resolution of 375 m. The modes of eVIIRS phenology data layers were extracted as the seasonality parameters of rice pixels. One can see from Table 7 that the seasonality parameters obtained by OBSUM plus TIMESAT are generally consistent with those of the eVIIRS phenology data. The differences in all seasonality parameters are within 10 days, indicating the potential of OBSUM for retrieving crop phenology at a fine scale.

Fig. 23 shows the reconstructed gap-free EVI time-series at the BC site from the start of season (DOY 144) to the end of season (DOY 300) in 2022, and the time intervals between the EVI maps is 14 days. The phenology of rice and the Landsat-like spatial details can be clearly observed in the EVI maps. The above discussion suggests that OBSUM

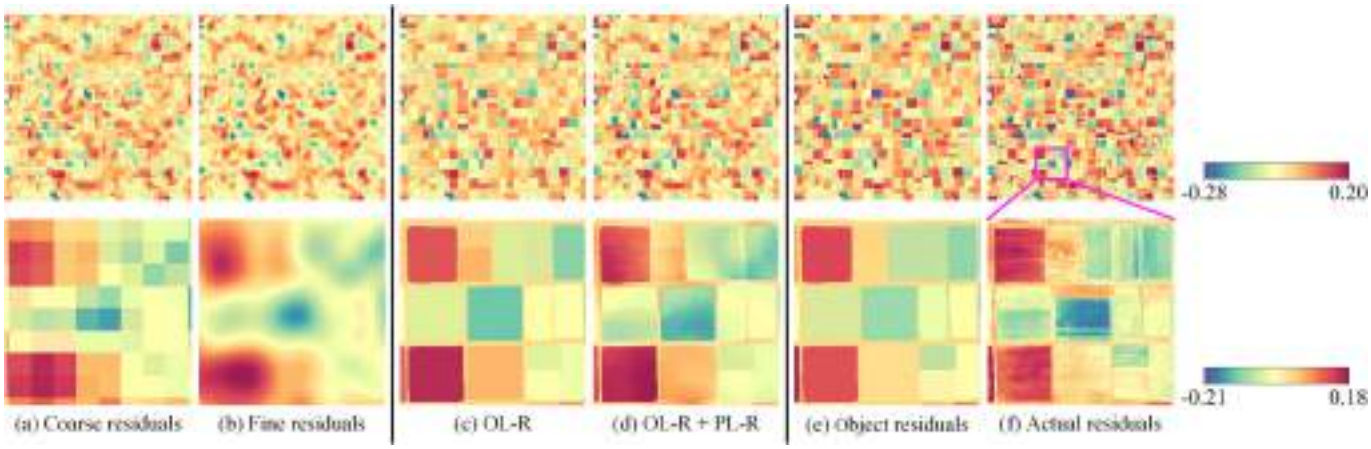

Fig. 20. Different types of residuals of the NIR band at the IC site when predicting the Sentinel-2 MSI image on June 04, 2022. (a) Coarse residuals, (b) fine residuals, (c) OL-R, (d) OL-R + PL-R, (e) object residuals, and (f) actual residuals. The second row shows the zoomed-in results of the sub-area marked in the magenta rectangle in the upper right sub-figure. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 6
Correlation coefficient (r) between different residual maps and the two reference residual maps. The bold values indicate the highest accuracy in each term.

Site	Reference	Coarse residuals	Fine residuals	OL-R	OL-R + PL-R
BC	Object residuals	0.65949	0.70276	**0.87213**	0.81810
	Actual residuals	0.65008	0.68910	0.72333	**0.74941**
IC	Object residuals	0.71370	0.76784	**0.91362**	0.88490
	Actual residuals	0.69734	0.75025	**0.82577**	0.82509

can support dynamic monitoring of crop progress at a fine field-scale.

4.2. Crop mapping

Spatiotemporal fusion can improve the temporal resolution of remote sensing images, thus facilitating crop mapping by providing dense time-series data. We collected 22 cloud-free Sentinel-2 MSI images at the IC site from January to December in 2022. Considering the large data volume and high computational cost of the experiment, only a 750 × 750 sub-area was selected as the test area. After that, the time-series NDVI was generated with the Sentinel-2 MSI images (hereafter, S2 NDVI). In order to test OBSUM's performance in supporting crop mapping, we also collected 22 Sentinel-3 OLCI images on the same dates as the Sentinel-2 MSI images. OBSUM was applied to predict the Sentinel-2 MSI-like images, and then the NDVI maps were calculated using the fused images (hereafter, OBSUM NDVI).

Random forest (RF) was selected as the classifier for crop mapping in this experiment. A more detailed introduction to such a technique can be found in Belgiu and Csillik (2018). The USDA's 30 m CDL, which covers

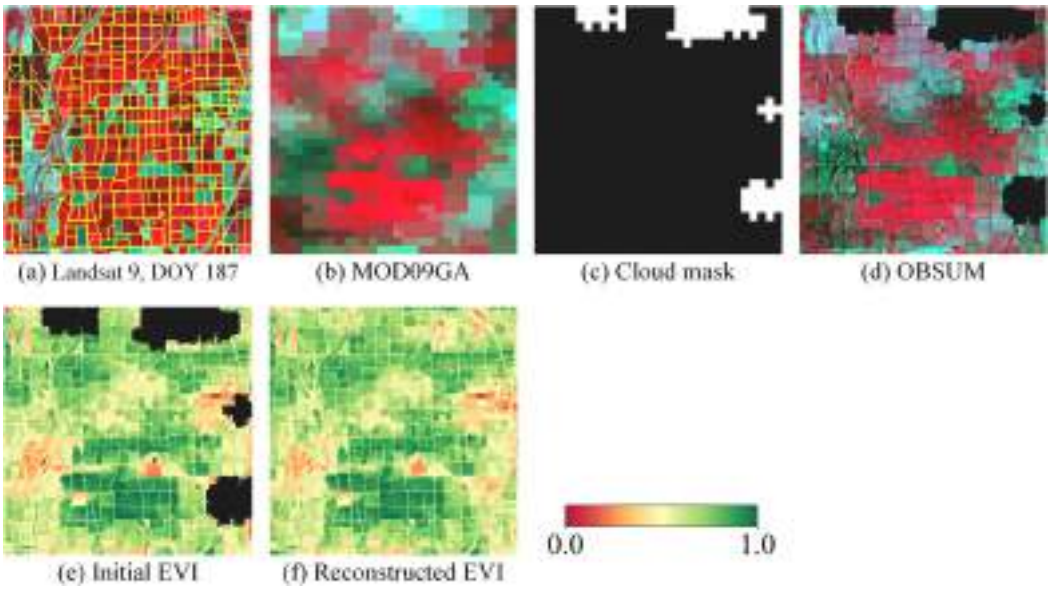

Fig. 21. Input images, intermediate images, and reconstructed gap-free EVI map at DOY 193, 2022. (a) Landsat 9 image at DOY 187 and the segmentation result, (b) MOD09GA image at DOY 193, (c) MOD09GA cloud mask, (d) Landsat-like image predicted by OBSUM, (e) EVI map with gaps, and (f) reconstructed gap-free EVI map.

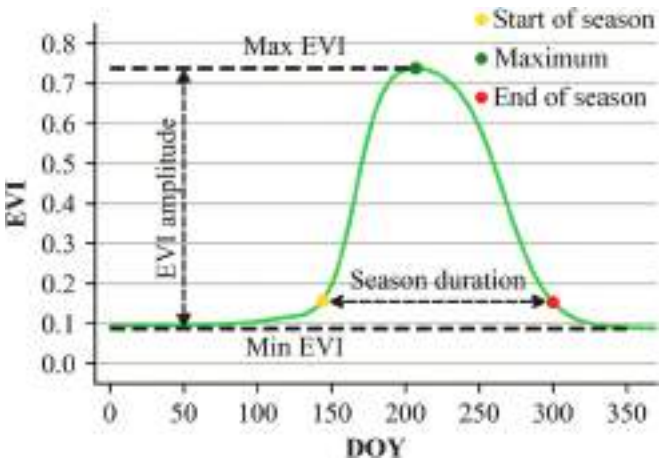

Fig. 22. Smoothed EVI curve and the seasonality parameters extracted using TIMESAT.

Table 7
Seasonality parameters obtained by OBSUM plus TIMESAT compared with the eVIIRS phenology data.

Method	SOS	MAX	EOS	DUR (days)
OBSUM+TIMESAT	144	213	300	156
eVIIRS phenology	141	204	308	167

the test area, was extracted as the reference crop map. Training and validation samples of nine main crop types were randomly generated using the CDL in ENVI software, each covering 20% of the total pixels. Please note that the training and validation samples are disjoint and independent from each other. After that, RF classifiers were trained separately on the S2 NDVI and OBSUM NDVI. Then the final classification process was accomplished.

Fig. 24 shows the USDA's CDL and the crop mapping results by utilizing S2 NDVI and OBSUM NDVI, respectively. One can see that the crop map obtained by OBSUM NDVI contains clear boundaries of cropland parcels. Moreover, it is also similar to the reference CDL and the crop map obtained by S2 NDVI. As shown in Table 8, the overall accuracy (OA) and Kappa coefficient (Kappa) of the crop map obtained by OBSUM NDVI are comparable to those of the S2 NDVI, and even slightly higher than them (which can be explained by the errors of the CDL). The above experiment demonstrates the superior performance of OBSUM in the application of crop mapping.

5. Discussion and conclusion

This study proposed an object-based spatial unmixing model (OBSUM) for spatiotemporal fusion of remote sensing images. OBSUM predicts the fine image at the prediction date by blending one pair of coarse and fine images at the base date and one coarse image at the prediction date. It includes one preprocessing step and three indispensable fusion steps: object-level unmixing (OL-U), object-level residual compensation (OL-RC), and pixel-level residual compensation (PL-RC). The OL-U produces an initial fusion result by incorporating spatial unmixing and object-based image analysis. After that, the OL-RC calculates and compensates the residual for each object, which can

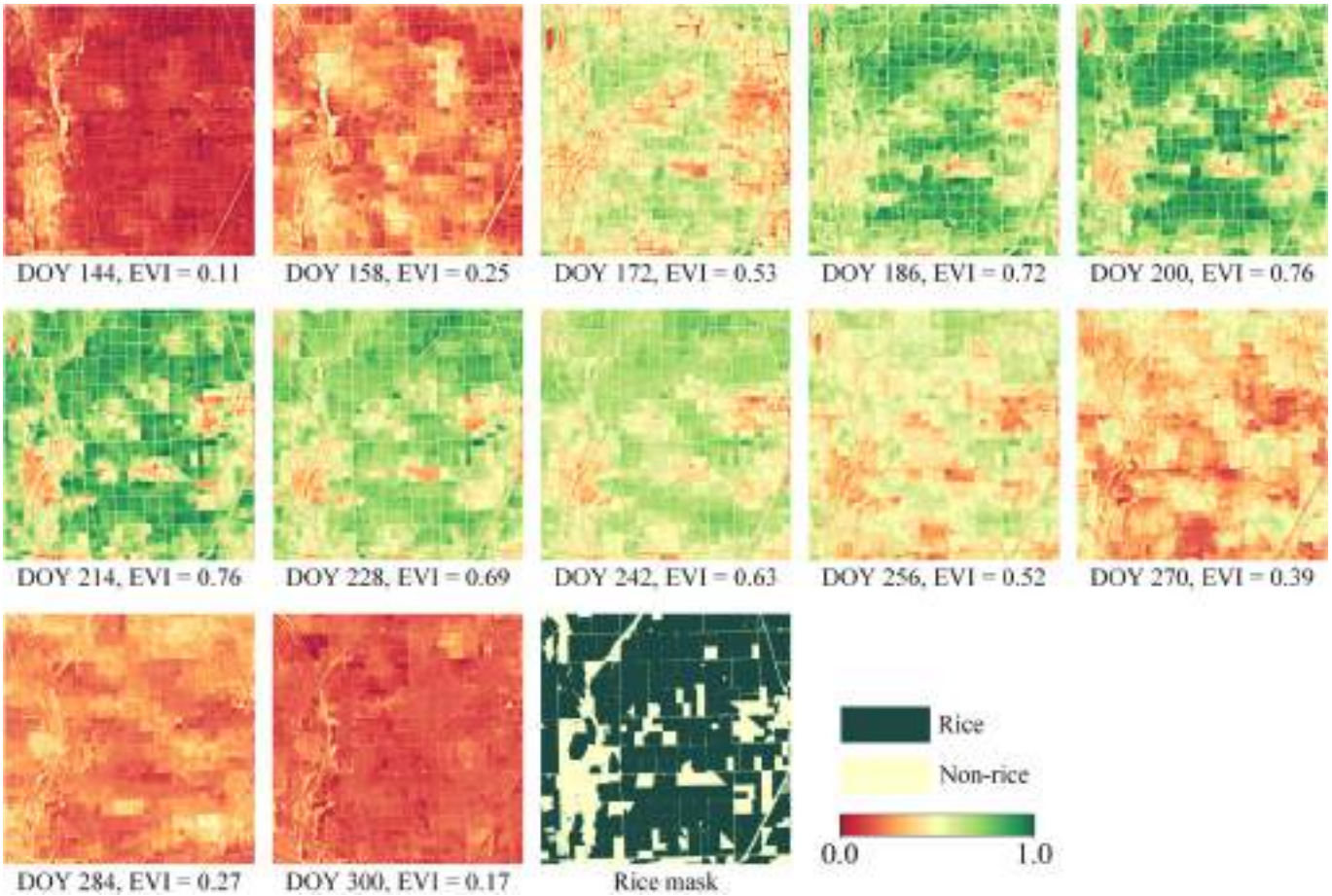

Fig. 23. Reconstructed gap-free EVI time-series at the BC site from DOY 144 to DOY 298 in 2022. The rice mask is extracted from the CDL provided by the USDA.

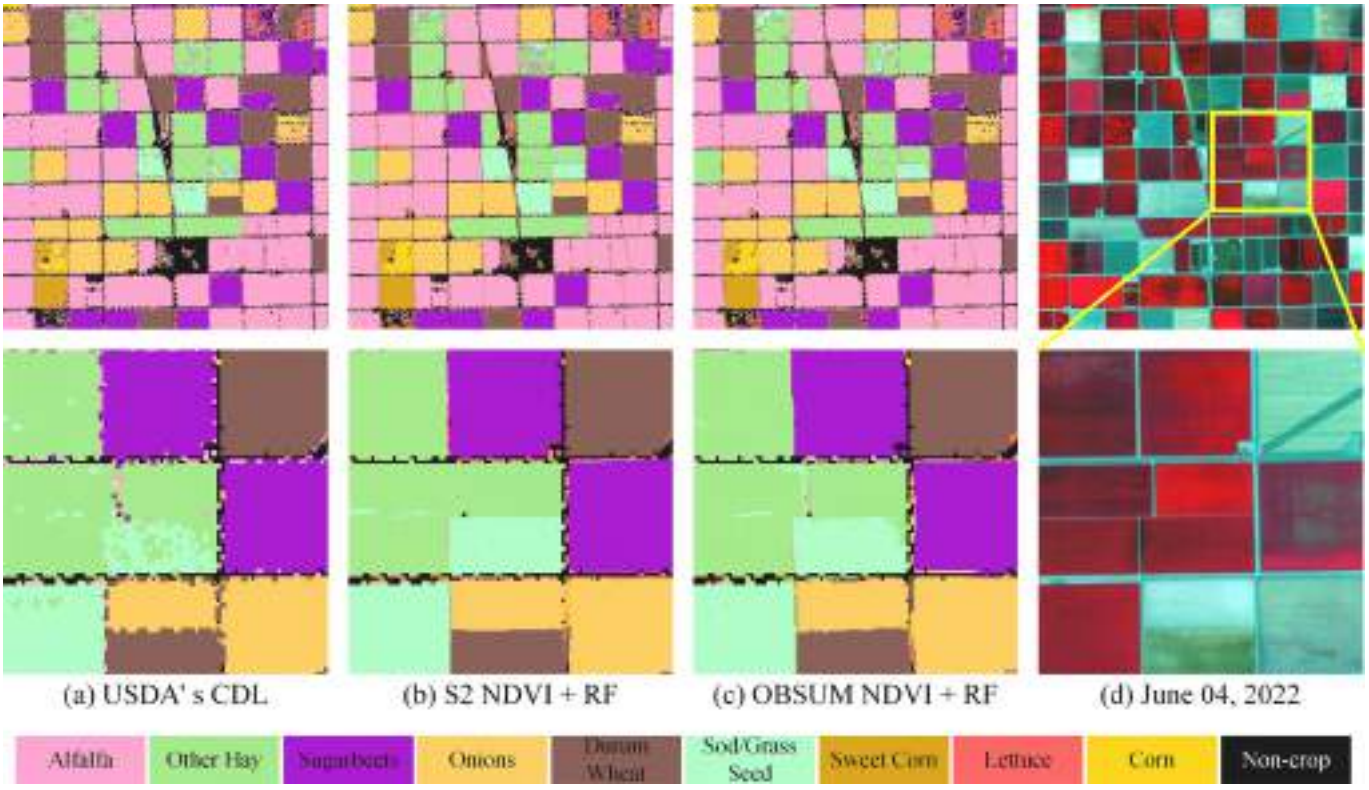

Fig. 24. Crop mapping result at the IC site in 2022. (a) USDA's CDL as reference, (b) crop map obtained by S2 NDVI and RF classification, (c) crop map obtained by OBSUM NDVI and RF classification, and (d) Sentinel-2 MSI image acquired on June 04, 2022. The black pixels represent non-crop areas and unselected minor crops. The second row shows the zoomed-in results of the sub-area marked in the yellow rectangle in the upper right sub-figure. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 8
Accuracy metric values of random forest crop mapping using S2 NDVI and OBSUM NDVI.

Metric	S2 NDVI	OBSUM NDVI
OA	0.8964	0.9111
Kappa	0.8636	0.8820

significantly recover the spectral information. Finally, the PL-RC is applied to retrieve the within-object land-cover changes, thus further improving the fusion accuracy.

The proposed OBSUM was tested at two typical agricultural sites and compared with seven methods, including UBDF, STARFM, Fit-FC, FSDAF, RASDF, VSDF, and OL-HSTFM. The experimental results demonstrate that OBSUM can retrieve both strong phenological changes and land-cover changes, thus outperforming the comparison methods in terms of visual effect and four accuracy indices. Benefiting from its accurate fusion performance, OBSUM has great potential for supporting various remote sensing applications at a fine spatial scale, such as dynamic crop progress monitoring and crop mapping.

OBSUM fully utilizes the spatial information of the fine image at t_b and the spectral information of the coarse image at t_p, thus obtaining an accurate prediction. In the preprocessing step, the fine image at the base date is classified into several land-cover classes to define the endmembers as well as calculate the endmembers' fractions for spatial unmixing. The fine image is also segmented by SAM to define the ground objects, thus guiding the subsequent object-level fusion steps. More importantly, the classification result and the segmented image objects are used together to get the object-level land-cover classification map. This can eliminate the pixel-level classification errors introduced by intra-class spectral variation, thereby providing a classification map that can be applied to the OL-U. In OL-U, the fine image, coarse-scale temporal changes, and image objects are utilized together to obtain an initial fusion result in which no block effect exists. In the following OL-RC, the coarse image at the prediction date is used to calculate the residuals, then downscaled to fine-scale. The object residual index (ORI) map generated from the image objects is applied to guide the residual compensation. In the final PL-RC, the coarse image at t_p is used again to calculate the residuals, and the fine image is used to select similar pixels to strengthen the prediction.

It is noteworthy that the recently proposed OL-HSTFM (Guo and Shi, 2023), which is also an object-level hybrid fusion method, outperformed other comparison methods (except OBSUM) in the experiments in Section 3.4 as well. Therefore, the experiments conducted in this paper can also demonstrate that object-level fusion methods are superior to traditional pixel-level fusion methods at agricultural sites. In order to improve both the spectral and spatial accuracies of the predicted image, OL-HSTFM incorporates the OL-STARFM and OL-Fit-FC methods by adaptively weighting their predictions according to the spatial details in the base fine image. However, OL-HSTFM simply adds the coarse-scale residuals to the preliminary prediction, which ignores the uncertainties caused by the scale difference between the coarse and fine images. By contrast, the proposed object-level residual compensation (OL-RC) only selects the most reliable residual pixels to compute and compensate for the residual of each object. By considering the shape information of image objects and the uncertainties in the residual map, OL-RC can effectively capture strong temporal changes and improve the prediction accuracy, as indicated in Sections 3.4 and 3.5. Moreover, although the pixel-level weighting scheme in OL-HSTFM can help to combine the advantages of both OL-STARFM and OL-Fit-FC, it also introduced many salt-and-pepper noise-like spectral distortions into the predicted images. We believe more object-level fusion methods will be developed in the future, providing better solutions to improve both the

Table 9
Accuracy metric values of OL-HSTFM and OBSUM with different image segmentation techniques.

Site	t_p	Metric	Method			
			OL-HSTFM + MRS	OBSUM + MRS	OL-HSTFM + SAM	OBSUM + SAM
BC	July 23, 2022	AD	−0.00036	**0.00010**	−0.00038	0.00029
		RMSE	0.04611	0.04254	0.04368	**0.04069**
		r	0.62246	0.70518	0.66038	**0.73439**
		SSIM	0.66615	0.69478	0.68456	**0.71634**
IC	June 04, 2022	AD	−0.00066	**0.00005**	−0.00081	0.00007
		RMSE	0.04444	0.04358	0.04162	**0.03882**
		r	0.80603	0.81543	0.83598	**0.85595**
		SSIM	0.76661	0.79058	0.79028	**0.83085**

object-level spectral accuracy and the pixel-level spatial details in the predicted images.

In this paper, the state-of-the-art SAM was adopted to segment the fine image into different ground objects. As shown in sub-figure (a) in Figs. 7 and 8, SAM performed well for the two experimental sites, with all of the ground objects segmented in regular shapes. However, the performance of SAM can become unstable in areas with heterogeneous land-cover and images with low spatial resolution (Osco et al., 2023). Nonetheless, it should be noticed that any other segmentation methods, such as the multiresolution segmentation (MRS) (Baatz, 2000), can be used to get a segmentation result for OBSUM. Table 9 shows the accuracy of OL-HSTFM and OBSUM with SAM and MRS as the segmentation techniques, respectively. One can see from Tables 9 and 5 that OBSUM outperformed seven comparison methods, regardless of the segmentation method. Moreover, for both OL-HSTFM and OBSUM, using SAM for image segmentation produced more accurate fusion results than using MRS. In the future, the modified version of SAM can be integrated into OBSUM to improve the fusion accuracy under complex scenarios.

In the experiment, the performance of OBSUM was validated on four different prediction dates, while only one base fine image was used. Such an experimental design ensures the validation of OBSUM's performance in retrieving strong temporal changes as well as its accuracy over a long time-series. A feasible strategy in practical application scenarios is to select the fine image that is nearest in time to the prediction date as the input fine image (Wen et al., 2023). Fig. 25 shows the fusion results of OBSUM at the BC site on July 23, 2022 with different base fine images, and Table 10 presents the fusion accuracy. One can observe from Fig. 25 (a1) that the base image on Feb. 13, 2022 has different image characteristics from the reference image at both the object-level (upper right ellipse) and pixel-level (bottom left ellipse). As a result, the fused image in Fig. 25 (b1) presents relatively low similarity to the reference image. By contrast, the base image on July 13, 2022 has a short time-interval from the reference image as well as more similar image characteristics (see pixels marked in blue ellipses). Therefore, it presents a better visual effect and much higher accuracy, as indicated in Fig. 25 (b2) and Table 10. Actually, there are many factors that affect the fusion accuracy (Chen et al., 2020), such as the spectral similarity, radiometric and geometric consistencies between the images at the base date and the prediction date. Shu et al. (2022) compared the fusion performance of FSDAF with different time intervals and concluded that the accuracy decreases with a longer time interval. A short time interval may ensure a high spectral similarity between the input images. However, it cannot guarantee the similarity and consistency of other factors that can affect the fusion performance, especially when abrupt land-cover changes occur. In the future, we will investigate the optimal strategy of generating dense time-series observations with OBSUM in practical applications.

It is noteworthy that there are also some limitations that motivate our future improvement on OBSUM. First, the object-based fusion strategy is only effective in areas with clear object boundaries, e.g., croplands, since both the reflectance and temporal changes within a cropland parcel are more likely to be consistent. In other words, OBSUM

Table 10
Accuracy metric values of OBSUM predictions at the BC site on July 23, 2022 with different base fine images.

Metric	Base date	
	Feb. 13, 2022	July 13, 2022
AD	0.00029	**0.00005**
RMSE	0.04069	**0.02955**
r	0.73439	**0.88723**
SSIM	0.71634	**0.83799**

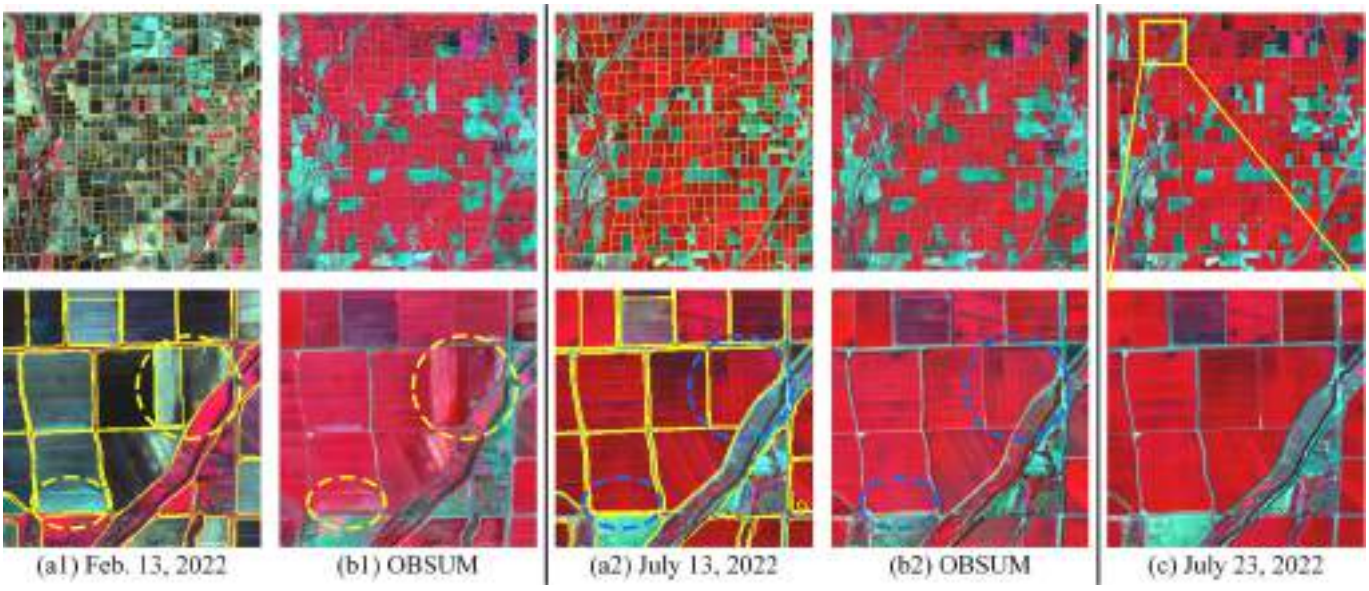

Fig. 25. Fusion results of OBSUM at the BC site on July 23, 2022 with different base fine images. (a1) and (b1) show the base image acquired on Feb. 13, 2022 and the fusion result, (a2) and (b2) present the base image acquired on July 13, 2022 and the fusion result, and (c) is the reference image. The second row shows the zoomed-in results of the sub-area marked in the yellow rectangle in the upper right sub-figure. All images use NIR-red-green as RGB false colour composition. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table A1
RMSE of the three fusion steps of OBSUM with different numbers of land-cover classes (n_c). The bold values indicate the highest accuracy in each column.

Site	n_c	OL-U	OL-RC	OBSUM
BC	3	0.06319	0.04365	0.04079
	4	0.06164	0.04357	0.04070
	5	0.06111	0.04355	0.04069
	6	0.06112	0.04355	0.04069
	7	**0.06025**	**0.04350**	**0.04066**
IC	3	0.06939	0.03726	**0.03876**
	4	0.06601	0.03745	0.03896
	5	0.06294	**0.03723**	0.03882
	6	0.06189	0.03728	0.03885
	7	**0.06133**	0.03759	0.03916

has limitations to predict the fine image accurately in forest and build-up areas, where the pixels cannot be segmented into objects properly. Therefore, OBSUM is only recommended for applications in which pixels can be properly segmented into objects, such as agricultural applications for crop progress monitoring and crop mapping. Moreover, OBSUM may also fail to predict the reflectance of small objects accurately. The reason is that OBSUM predicts the reflectance of the small objects and their neighboring objects separately. Therefore, it is sensitive to the large scale factor between the coarse and fine images.

Second, among the three fusion steps of OBSUM, only the PL-RC is designed for recovering within-object land-cover changes. Therefore, OBSUM may fail to capture abrupt land-cover changes that would break the object boundaries segmented from the fine image at t_b, such as floods, forest degradation, wildfires, etc. The main difficulties in such fusion circumstances are to select the unchanged coarse pixels for spatial unmixing and to identify the precise extent of land-cover changes (Zhu et al., 2018). Moreover, the similar pixel selection scheme in PL-RC assumes that the similar pixels at t_b still remain similar at t_p, which is empirical and not robust enough because it ignores the land-cover dynamics. As a result, the ability of PL-RC to predict within-object land-cover changes is limited. In addition, as illustrated in Section 3.3, the PL-RC is also sensitive to spatial inconsistencies between the coarse and fine images. In our further research, the reliability index (Shi et al., 2022), land-cover change detection (Jiang and Huang, 2022), and thin plate spline (TPS) interpolation (Zhu et al., 2016) techniques will be integrated into the OBSUM framework to address the abovementioned problems. The reliability index can filter out the changed and unreliable coarse pixels, thus improving the OBSUM's robustness to land-cover changes. Moreover, it can also guide the PL-RC to obtain a more accurate prediction. Change detection and TPS interpolation can help to retrieve the abrupt land-cover changes and predict the changed object boundaries, respectively.

The Python code of OBSUM and the experimental dataset are available at https://github.com/HoucaiGuo/OBSUM-code.

CRediT authorship contribution statement

Houcai Guo: Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization, Writing – original draft. **Dingqi Ye:** Validation, Methodology, Conceptualization, Writing – original draft. **Hanzeyu Xu:** Data curation. **Lorenzo Bruzzone:** Supervision, Methodology, Conceptualization, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The Python code of OBSUM and the experimental dataset are available at https://github.com/HoucaiGuo/OBSUM-code.

Acknowledgement

This study was supported by the China Scholarship Council under Grant 202306860011. The authors would like to thank Dr. Feng Gao, Prof. Qunming Wang, Prof. Xiaolin Zhu, Mr. Chen Xu, Mr. Dizhou Guo, Prof. Jin Chen, and Ms. Nikolina Mileva for providing the source code of STARFM, Fit-FC, FSDAF, VSDF, RASDF, OL-HSTFM, the vegetation index Savitzky–Golay filter, and the Python implementation of STARFM. The authors would also like to thank Ms. Yijie Tang and Ms. Zhuoning Gu for their helpful comments with preprocessing the Sentinel-3 OLCI data. The authors would also like to thank the Editor in Chief and three anonymous reviewers for their professional, constructive, and insightful comments, which helped us improve our method as well as the quality of this manuscript.

Appendix A

This Appendix analyzes the impact of different number of land-cover classes (n_c) on the fusion results. Table A1 shows the RMSE of the three fusion steps of OBSUM with different n_c values when predicting the fine image at the BC site on July 23, 2022, and at the IC site on June 04, 2022, respectively. The other parameters remain the same as the settings in Section 3.2. From Table A1, one can see that the RMSE values of all three steps decrease gradually by increasing n_c at the BC site. To be specific, when n_c increases from 3 to 7, the RMSE provided by OL-U, OL-RC, and OBSUM decreases by 0.00294, 0.00015, and 0.00013, respectively. However, the accuracy improvements on OL-RC and OBSUM are not as noticeable as that on OL-U, demonstrating that increasing n_c does not improve the accuracy of OBSUM effectively. One can also observe from Table A1 that when predicting the fine image at the IC site, the increase of n_c can effectively improve the accuracy of OL-U. However, it also decreases the accuracy of both OL-RC and the final OBSUM predictions. The above discussion indicates that the number of land-cover classes is not a key parameter of OBSUM. The reason is that the proposed object-level residual compensation (OL-RC) can effectively retrieve strong temporal changes between the base date and the prediction date, thus compensating for the inaccurate OL-U prediction caused by a lower n_c. Therefore, the users are recommended to set the n_c parameter according to their visual interpretation of the fine image at the base date instead of simply using a large value.

References

Baatz, M., 2000. Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. In: Angewandte Geographische Informationsverarbeitung, pp. 12–23. https://cir.nii.ac.jp/crid/1572261550679971840.

Belgiu, M., Csillik, O., 2018. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. Remote Sens. Environ. 204, 509–523. https://doi.org/10.1016/j.rse.2017.10.005.

Blaschke, T., 2010. Object based image analysis for remote sensing. ISPRS J. Photogramm. Remote Sens. 65 (1), 2–16. https://doi.org/10.1016/j.isprsjprs.2009.06.004.

Brodu, N., 2017. Super-resolving multiresolution images with band-independent geometry of multispectral pixels. IEEE Trans. Geosci. Remote Sens. 55 (8), 4610–4617. https://doi.org/10.1109/TGRS.2017.2694881.

Chen, J., Jönsson, P., Tamura, M., Gu, Z., 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 (3), 332–344. https://doi.org/10.1016/j.rse.2004.03.014.

Chen, Y., Cao, R., Chen, J., Zhu, X., Zhou, J., Wang, G., Shen, M., Chen, X., Yang, W., 2020. A new cross-fusion method to automatically determine the optimal input image pairs for NDVI spatiotemporal data fusion. IEEE Trans. Geosci. Remote Sens. 58 (7), 5179–5194. https://doi.org/10.1109/TGRS.2020.2973762.

Chen, S., Wang, J., Gong, P., 2023. ROBOT: a spatiotemporal fusion model toward seamless data cube for global remote sensing applications. Remote Sens. Environ. 294, 113616. https://doi.org/10.1016/j.rse.2023.113616.

Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.-C., Skakun, S.V., Justice, C., 2018. The harmonized Landsat and Sentinel-2 surface reflectance data set. Remote Sens. Environ. 219, 145–161. https://doi.org/10.1016/j.rse.2018.09.002.

Donlon, C., Berruti, B., Buongiorno, A., Ferreira, M.H., Féménias, P., Frerick, J., Goryl, P., Klein, U., Laur, H., Mavrocordatos, C., Nieke, J., Rebhan, H., Seitz, B., Stroede, J., Sciarra, R., 2012. The global monitoring for environment and security (GMES) Sentinel-3 mission. Remote Sens. Environ. 120, 37–57. https://doi.org/10.1016/j.rse.2011.07.024.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote Sens. Environ. 120, 25–36. https://doi.org/10.1016/j.rse.2011.11.026.

Erdem, F., Avdan, U., 2023. STFRDN: a residual dense network for remote sensing image spatiotemporal fusion. Int. J. Remote Sens. 44 (10), 3259–3277. https://doi.org/10.1080/01431161.2023.2221800.

Feng, G., 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 (8), 2207–2218. https://doi.org/10.1109/TGRS.2006.872081.

Gao, F., Hilker, T., Zhu, X., Anderson, M., Masek, J., Wang, P., Yang, Y., 2015. Fusing Landsat and MODIS data for vegetation monitoring. IEEE Geosci. Remote Sens. Mag. 3 (3), 47–60. https://doi.org/10.1109/MGRS.2015.2434351.

Gao, F., Anderson, M.C., Zhang, X., Yang, Z., Alfieri, J.G., Kustas, W.P., Mueller, R., Johnson, D.M., Prueger, J.H., 2017. Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. Remote Sens. Environ. 188, 9–25. https://doi.org/10.1016/j.rse.2016.11.004.

Gevaert, C.M., García-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. https://doi.org/10.1016/j.rse.2014.09.012.

Ghamisi, P., Rasti, B., Yokoya, N., Wang, Q., Hofle, B., Bruzzone, L., Bovolo, F., Chi, M., Anders, K., Gloaguen, R., Atkinson, P.M., Benediktsson, J.A., 2019. Multisource and multitemporal data fusion in remote sensing: a comprehensive review of the state of the art. IEEE Geosci. Remote Sens. Mag. 7 (1), 6–39. https://doi.org/10.1109/MGRS.2018.2890023.

Gu, Z., Chen, J., Chen, Y., Qiu, Y., Zhu, X., Chen, X., 2023. Agri-fuse: a novel spatiotemporal fusion method designed for agricultural scenarios with diverse phenological changes. Remote Sens. Environ. 299, 113874. https://doi.org/10.1016/j.rse.2023.113874.

Guan, H., Su, Y., Hu, T., Chen, J., Guo, Q., 2019. An object-based strategy for improving the accuracy of spatiotemporal satellite imagery fusion for vegetation-mapping applications. Remote Sens. (Basel) 11 (24).

Guo, D., Shi, W., 2023. Object-level hybrid spatiotemporal fusion: reaching a better tradeoff among spectral accuracy, spatial accuracy, and efficiency. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 16, 8007–8021. https://doi.org/10.1109/JSTARS.2023.3310195.

Guo, D., Shi, W., Qian, F., Wang, S., Cai, C., 2022a. Monitoring the spatiotemporal change of Dongting Lake wetland by integrating Landsat and MODIS images, from 2001 to 2020. Eco. Inform. 72, 101848. https://doi.org/10.1016/j.ecoinf.2022.101848.

Guo, D., Shi, W., Zhang, H., Hao, M., 2022b. A flexible object-level processing strategy to enhance the weight function-based spatiotemporal fusion method. IEEE Trans. Geosci. Remote Sens. 60, 1–11. https://doi.org/10.1109/TGRS.2022.3212474.

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 Sens. Environ. 113 (8), 1613–1627. https://doi.org/10.1016/j.rse.2009.03.007.

Hossain, M.D., Chen, D., 2019. Segmentation for object-based image analysis (OBIA): a review of algorithms and challenges from remote sensing perspective. ISPRS J. Photogramm. Remote Sens. 150, 115–134. https://doi.org/10.1016/j.isprsjprs.2019.02.009.

Houborg, R., McCabe, M.F., 2018. A Cubesat enabled spatio-temporal enhancement method (CESTEM) utilizing planet, Landsat and MODIS data. Remote Sens. Environ. 209, 211–226. https://doi.org/10.1016/j.rse.2018.02.067.

Huang, B., Song, H., 2012. Spatiotemporal reflectance fusion via sparse representation. IEEE Trans. Geosci. Remote Sens. 50 (10), 3707–3716. https://doi.org/10.1109/TGRS.2012.2186638.

Jiang, X., Huang, B., 2022. Unmixing-based spatiotemporal image fusion accounting for complex land cover changes. IEEE Trans. Geosci. Remote Sens. 60, 1–10. https://doi.org/10.1109/TGRS.2022.3173172.

Jonsson, P., Eklundh, L., 2002. Seasonality extraction by function fitting to time-series of satellite sensor data. IEEE Trans. Geosci. Remote Sens. 40 (8), 1824–1832. https://doi.org/10.1109/TGRS.2002.802519.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A., Lo, W.-Y., Dollár, P., Girshick, R., 2023. Segment Anything. https://doi.org/10.48550/arXiv.2304.02643.

Kwan, C., Zhu, X., Gao, F., Chou, B., Perez, D., Li, J., Shen, Y., Koperski, K., Marchisio, G., 2018. Assessment of spatiotemporal fusion algorithms for planet and worldview images. Sensors 18 (4).

Li, A., Bo, Y., Zhu, Y., Guo, P., Bi, J., He, Y., 2013. Blending multi-resolution satellite sea surface temperature (SST) products using Bayesian maximum entropy method. Remote Sens. Environ. 135, 52–63. https://doi.org/10.1016/j.rse.2013.03.021.

Liu, X., Deng, C., Chanussot, J., Hong, D., Zhao, B., 2019. StfNet: a two-stream convolutional neural network for spatiotemporal image fusion. IEEE Trans. Geosci. Remote Sens. 57 (9), 6552–6564. https://doi.org/10.1109/TGRS.2019.2907310.

Lloyd, S., 1982. Least squares quantization in PCM. IEEE Trans. Inf. Theory 28 (2), 129–137. https://doi.org/10.1109/TIT.1982.1056489.

Mingquan, W., Zheng, N., Changyao, W., Chaoyang, W., Li, W., 2012. Use of MODIS and Landsat time series data to generate high-resolution temporal synthetic Landsat data using a spatial and temporal reflectance fusion model. J. Appl. Remote. Sens. 6 (1) https://doi.org/10.1117/1.JRS.6.063507, 063507.

Osco, L.P., Wu, Q., de Lemos, E.L., Gonçalves, W.N., Ramos, A.P.M., Li, J., Marcato, J., 2023. The segment anything model (SAM) for remote sensing applications: from zero to one shot. Int. J. Appl. Earth Obs. Geoinf. 124, 103540. https://doi.org/10.1016/j.jag.2023.103540.

Shen, H., Meng, X., Zhang, L., 2016. An integrated framework for the spatio–temporal–spectral fusion of remote sensing images. IEEE Trans. Geosci. Remote Sens. 54 (12), 7135–7148. https://doi.org/10.1109/TGRS.2016.2596290.

Shi, W., Guo, D., Zhang, H., 2022. A reliable and adaptive spatiotemporal data fusion method for blending multi-spatiotemporal-resolution satellite images. Remote Sens. Environ. 268, 112770. https://doi.org/10.1016/j.rse.2021.112770.

Shu, H., Jiang, S., Zhu, X., Xu, S., Tan, X., Tian, J., Xu, Y.N., Chen, J., 2022. Fusing or filling: which strategy can better reconstruct high-quality fine-resolution satellite time series? Sci. Remote Sens. 5, 100046. https://doi.org/10.1016/j.srs.2022.100046.

Stefan, A., Sindy, S., Keukelaere, L.D., Kerchove, R.V.D., Knaeps, E., 2018. Atmospheric correction Icor and integration in operational workflows. In: IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, 22–27 July 2018.

Tobler, W.R., 1970. A computer movie simulating urban growth in the Detroit region. Econ. Geogr 46 (sup1), 234–240. https://doi.org/10.2307/143141.

USDA, 2022a. 2022 Crop Progress and Conditions (California). Retrieved May 18 from. https://www.nass.usda.gov/Charts_and_Maps/Crop_Progress_&_Condition/2022/CA_2022.pdf.

USDA, 2022b. Crop Progress. Retrieved May 27 from. https://usda.library.cornell.edu/concern/publications/8336h188j?locale=en&page=3#release-items.

USDA, 2022c. Cropland Data Layer. Retrieved May 19 from. https://croplandcros.scinet.usda.gov/.

USGS, 2022. -NPP Western U.S. 375 m eVIIRS Remote Sensing Phenology Data. Retrieved December 18 from. https://doi.org/10.5066/F7PC30G1.

Wang, Q., Atkinson, P.M., 2018. Spatio-temporal fusion for daily Sentinel-2 images. Remote Sens. Environ. 204, 31–42. https://doi.org/10.1016/j.rse.2017.10.046.

Wang, Q., Peng, K., Tang, Y., Tong, X., Atkinson, P.M., 2021. Blocks-removed spatial unmixing for downscaling MODIS images. Remote Sens. Environ. 256 https://doi.org/10.1016/j.rse.2021.112325, 112325.

Wang, Q., Tang, Y., Ge, Y., Xie, H., Tong, X., Atkinson, P.M., 2023. A comprehensive review of spatial-temporal-spectral information reconstruction techniques. Sci. Remote Sens. 8, 100102. https://doi.org/10.1016/j.srs.2023.100102.

Wen, Y., Yang, J., Liao, W., Xiao, J., Yan, S., 2023. Refined assessment of space-time changes, influencing factors and socio-economic impacts of the terrestrial ecosystem quality: a case study of the GBA. J. Environ. Manage. 345, 118869. https://doi.org/10.1016/j.jenvman.2023.118869.

Xu, C., Du, X., Yan, Z., Zhu, J., Xu, S., Fan, X., 2022. VSDF: a variation-based spatiotemporal data fusion method. Remote Sens. Environ. 283, 113309. https://doi.org/10.1016/j.rse.2022.113309.

Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., Hodges, J.C.F., Gao, F., Reed, B.C., Huete, A., 2003. Monitoring vegetation phenology using MODIS. Remote Sens. Environ. 84 (3), 471–475. https://doi.org/10.1016/S0034-4257(02)00135-9.

Zhang, F., Zhu, X., Liu, D., 2014. Blending MODIS and Landsat images for urban flood mapping. Int. J. Remote Sens. 35 (9), 3237–3253. https://doi.org/10.1080/01431161.2014.903351.

Zhang, H., Song, Y., Han, C., Zhang, L., 2021a. Remote sensing image spatiotemporal fusion using a generative adversarial network. IEEE Trans. Geosci. Remote Sens. 59 (5), 4273–4286. https://doi.org/10.1109/TGRS.2020.3010530.

Zhang, H., Sun, Y., Shi, W., Guo, D., Zheng, N., 2021b. An object-based spatiotemporal fusion model for remote sensing images. Eur. J. Remote Sens. 54 (1), 86–101. https://doi.org/10.1080/22797254.2021.1879683.

Zhao, X., Nishina, K., Akitsu, T.K., Jiang, L., Masutomi, Y., Nasahara, K.N., 2023. Feature-based algorithm for large-scale rice phenology detection based on satellite images. Agric. For. Meteorol. 329, 109283. https://doi.org/10.1016/j.agrformet.2022.109283.

Zhou, W., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13 (4), 600–612. https://doi.org/10.1109/TIP.2003.819861.

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 (11), 2610–2623. https://doi.org/10.1016/j.rse.2010.05.032.

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 Sens. Environ. 172, 165–177. https://doi.org/10.1016/j.rse.2015.11.016.

Zhu, X., Cai, F., Tian, J., Williams, T.K., 2018. Spatiotemporal fusion of multisource remote sensing data: literature survey, taxonomy, principles, applications, and future directions. Remote Sens. (Basel) 10 (4).

Zhu, X., Zhan, W., Zhou, J., Chen, X., Liang, Z., Xu, S., Chen, J., 2022. A novel framework to assess all-round performances of spatiotemporal fusion models. Remote Sens. Environ. 274, 113002. https://doi.org/10.1016/j.rse.2022.113002.

Zhukov, B., Oertel, D., Lanzl, F., Reinhackel, G., 1999. Unmixing-based multisensor multiresolution image fusion. IEEE Trans. Geosci. Remote Sens. 37 (3), 1212–1226. https://doi.org/10.1109/36.763276.

Zurita-Milla, R., Clevers, J.G.P.W., Schaepman, M.E., 2008. Unmixing-based Landsat TM and MERIS FR data fusion. IEEE Geosci. Remote Sens. Lett. 5 (3), 453–457. https://doi.org/10.1109/LGRS.2008.919685.