An object-based spatiotemporal fusion model for remote sensing images

Hua Zhang a, Yue Sun a, Wenzhong Shi b, Dizhou Guo a and Nanshan Zheng a

aSchool of Environmental Science and Spatial Informatics, China University of Mining and Technology Xuzhou, Jiangsu, China; 
bDepartment of Land Surveying and Geo-Informatics, Smart Cities Research Institute, The Hong Kong Polytechnic University, Kowloon, Hong Kong

ABSTRACT

Spatiotemporal fusion technique can combine the advantages of temporal resolution and spatial resolution of different images to achieve continuous monitoring for the Earth’s surface, which is a feasible solution to resolve the trade-off between the temporal and spatial resolutions of remote sensing images. In this paper, an object-based spatiotemporal fusion model (OBSTFM) is proposed to produce spatiotemporally consistent data, especially in areas experiencing non-shape changes (including phenology changes and land cover changes without shape changes). Considering different changes that might occur in different regions, multi-resolution segmentation is first employed to produce segmented objects, and then a linear injection model is introduced to produce preliminary prediction. In addition, a new optimized strategy to select similar pixels is developed to obtain a more accurate prediction. The performance of proposed OBSTFM is validated using two remotely sensed dataset experiencing phenology changes in the heterogeneous area and land cover type changes, experimental results show that the proposed method is advantageous in such areas with non-shape changes, and has satisfactory robustness and reliability in blending large-scale abrupt land cover changes. Consequently, OBSTFM has great potential for monitoring highly dynamic landscapes.

Introduction

In recent years, spatiotemporal data fusion technology has been explored in many studies to promote the applications of multi-source remote sensing data. For example, the high spatial and temporal images generated by spatiotemporal data fusion can be used to monitor crop growth (Gao et al., 2017; Zheng et al., 2016), ecosystem dynamics (Gao et al., 2015; Nagendra et al., 2013), estimate the crop yields (Allen et al., 2011; L. Li et al., 2017), biomass (B. Zhang et al., 2016), and improve land cover classification (Chen et al., 2017; Kong et al., 2016). Many multi-satellite data fusion algorithms are developed to integrate information from different spatial, temporal resolutions to produce data with high spatial and high temporal resolutions simultaneously. According to the techniques used to model the relationship between input two or more data sources, existing spatiotemporal data fusion methods can be classified into five categories: weight function-based, unmixing-based, Bayesian-based, learning-based, and hybrid methods (Li et al., 2020; Zhu et al., 2018).

Among the weight function-based method, the spatial and temporal adaptive reflectance fusion model (STARFM) proposed by Gao et al. is the earliest spatiotemporal fusion model. Most subsequent weight function-based methods are developed by the principle of STARFM. STARFM assumes that the central pixel and its spectrally similar pixels within a moving window have similar temporal increments, and then establishes the relationship between high- and low-resolution images estimates the fused results by weighting the searched similar pixels (Gao et al., 2006). However, this method is not suitable for heterogeneous landscapes or where land cover changes exist. Several weight function-based methods are since developed to overcome the limitations of STARFM. For example, the spatial temporal adaptive algorithm for mapping reflectance changes (STAARCH) was developed to improve the ability to capture disturbance phenomena (Hilker et al., 2009), an enhanced spatial and temporal adaptive reflectance fusion model (ESTARFM) shows better prediction accuracy in heterogeneous landscape than STARFM (Zhu et al., 2010), the spatial and temporal nonlocal filter-based fusion model proposed by Cheng et al. further improves the performance in heterogeneous areas (Cheng et al., 2017), Wang et al. developed an algorithm named Fit-FC to improve the prediction accuracy in different landscapes (Q. Wang & Atkinson, 2018). Furthermore, some research focus on optimizing the selection strategy of similar pixels by integrating initial prediction (Lu et al., 2016; Sun et al., 2019).
and a recently proposed fusion method named the robust adaptive spatial and temporal fusion model (RASTFM) employed the non-local redundancy strategy to search for similar pixels to achieve more robust prediction (Zhao et al., 2018).

With respect to the unmixing-based methods, the first unmixing-based spatiotemporal method might be the multisensor multiresolution technique (MMT) proposed by Zhukov et al. (Zhukov et al., 1999), which faced the challenge in spectral unmixing. Some unmixing-based fusion methods were subsequently proposed to address this challenge by introducing constraints (Zurita-Milla et al., 2008) or incorporating regularization into the unmixing process (Amorós-López et al., 2013), and the object-based spatial and temporal vegetation unmixing model (OB-STVIUM) applied segmentation to better estimate the endmembers (Lu et al., 2016). Some research focused on acquiring the prediction by estimating endmember changes (Wu et al., 2015, 2012), and STDFA was further improved by incorporating sensor differences and spatial variability (Wu et al., 2018). Maselli et al. proposed the recently developed spatial enhancer of vegetation index image series to produce spatially enhanced NDVI data series based on linear spectral mixing theory (Maselli et al., 2019).

Among the Bayesian-based methods, the Bayesian-based method has shown its effectiveness in spatial-spectral fusion of remotely sensed images, for example, the Bayesian maximum entropy method (BME) employs covariance functions to link the images of advanced microwave scanning radiometer for Earth observing system (AMSR-E) and MODIS, blends the synthetic sea surface temperature data effectively (A. Li et al., 2013). The NDVI-BSFM method uses linear mixing model to link MODIS and Landsat NDVI and multi-year average NDVI time series data, as a result, performs strengths and robustness in providing fine NDVI datasets (Liao & Wang, 2016). A Bayesian data fusion approach that incorporates the temporal correlation information in the image time series was proposed by Xue et al. (2017), and showed better performance than STARFM and ESTARFM in blending heterogeneous landscapes.

Recently, learning-based methods have grown considerably (Li et al., 2020; Tan et al., 2018). Huang and Song (2012) proposed the Sparse-representation-based Spatiotemporal reflectance Fusion Model (SPSTFM), which is perhaps the first dictionary-pair learning method in the spatiotemporal fusion field. Wu et al. (2015) proposed a new sparse coding method with dictionary perturbation regularization in sparse-representation-based spatiotemporal reflectance fusion and can provide a more accurate prediction than STARFM. Wei et al. proposed (Wei et al., 2017) a new dictionary-based fusion method using compressed sensing to integrate remote sensing images of different spatial and temporal resolutions and obtained satisfactory reconstruction results. Song et al. (2018) established a novel spatiotemporal fusion using deep convolutional neural network (STFDCNN), which can effectively extract and express information in large-scale remote sensing data. Tan et al. (2018) proposed a new deep convolutional spatiotemporal fusion network (DCSTFN) and achieved satisfactory performance. Sun and Zhang (2019) proposed a two-stage spatiotemporal fusion method, which showed its advantages in capturing phenomenological changes and retrieving land cover changes. H. Zhang et al. (2020) proposed a remote sensing image spatiotemporal fusion method using a generative adversarial network (STFGAN) and showed its satisfactory effectiveness and generalization ability.

In order to take advantage of different algorithms, hybrid spatiotemporal fusion methods are developed to integrate the ideas of different methods listed above. For example, spatial and temporal reflectance unmixing model (STRUM) (Gevaert & García-Haro, 2015) and unmixing-based STARFM (USTARFM) (Xie et al., 2016) combine the strengths of STARFM and unmixing-based method, and the improved STRUM (ISTRUM) further integrates the spectral mixture analysis to better capture spectral variability and spatial details (Ma et al., 2018), and the recently proposed virtual image pair-based spatio-temporal fusion (VIPSTF) approach provides a flexible framework to improve the existing spatial weighting- and spatial unmixing-based methods (Q. Wang et al., 2020). The flexible spatiotemporal data fusion (FSDAF) is a representative spatiotemporal fusion framework integrating linear unmixing, spatial interpolation and weight-based functions. Specifically, FSDAF uses the classifier and the unmixing theory to estimate the temporal change of each class, and then distributes residuals with the help of the thin plate spline interpolator, lastly, obtains the final prediction by introducing the information in neighborhood. As a result, FSDAF can get a more robust and accurate prediction in different landscapes, especially in heterogeneous landscapes. Moreover, FSDAF has the ability to predict both gradual change and land cover type change (Zhu et al., 2016; Liu et al., 2019b). Following FSDAF, the enhanced FSDAF (EFSDAF), an improved FSDAF (IFSDF), an enhanced FSDAF that incorporates sub-pixel class fraction change information (SFSDAF), and FSDAF 2.0 were later developed to overcome the limitations of FSDAF and improve the prediction accuracy (Liu et al., 2019b; Guo et al., 2020; X. Li et al., 2020; Shi et al., 2019).

While above spatiotemporal fusion algorithms have ability to produce effective fusion result to a certain extent, they still face some challenges in...
practice. First, many existing spatiotemporal fusion methods, like ESTARFM, STAARCH and Bayesian-based methods need more than one low- and high-resolution image pair as input data, which greatly limits the application scenarios of spatiotemporal fusion technology (Guo et al., 2020). Specifically, methods like ESTARFM cannot be used to do real-time processing due to it needs the data after the prediction phase. And when we need to blend historical data it may be also difficult to acquire one more low- and high-resolution image pair which is temporally close to the prediction time due to cloud cover in many cases (Song & Huang, 2013; Q. Wang & Atkinson, 2018). Second, the ability of retrieving fast change and the robustness of method need to be further improved, especially when blending heterogeneous sites. For example, the Fit-FC proposed to retrieve strong temporal changes performs more effectively than STARFM and FSDAF in capturing fast phenological changes (Q. Wang & Atkinson, 2018), but shows lower accuracy than FSDAF in retaining image structures when blending heterogeneous sites (Liu et al., 2019a). The accuracy of learning-based methods decreases when spatial heterogeneity is high and spectral scale differences between low- and high-resolution images are large (Zhu et al., 2016). Many of which employ the idea of unmixing like STDFA and FSDAF assume the classification result in two phases is same, as a result, the unmixing step can preserve surface structures well, however, the classification errors can affect the accuracy of unmixing results seriously and lower the robustness of model. Consequently, existing spatiotemporal fusion algorithms need further improvement and optimization.

To address the above-mentioned limitations of aforementioned spatiotemporal fusion methods to some extent, in this paper, an object-based spatiotemporal fusion method (OBSTFM) is proposed to more effectively capturing temporal changes with minimum input data. Considering the actual distribution of surface features, the segmentation method is introduced to produce surface objects with good similarity and uniformity, the computation of fusion process will be optimized in each object. The proposed method consists of three parts including multi-resolution segmentation, linear injection and spatial filtering. Multi-resolution segmentation is firstly applied to generate segmented objects, and then the linear injection model is introduced to fuse input images in each object to produce preliminary prediction, which is finally refined using spatial filtering in each object to derive final prediction. The main contributions of this paper are as follows: 1) based on multi-resolution segmentation result, the local linear interpolation model is implemented in each object, which can ensure the fusion uniformity of objects without being influenced by different variations of other objects, and 2) in the process of spatial filtering, the optimal selection

![Flowchart of the proposed method.](image-url)

Figure 1. Flowchart of the proposed method.
strategy of similar pixels is provided to achieve more robust and accurate prediction.

**Methodology**

The basic idea of OBSTFM is to downscale the low-resolution image to high-resolution object through spatial analysis and the linear injection theory. Firstly, similar to the first step in OB-STVIUM (Lu et al., 2016), multi-resolution segmentation is applied in the input high-resolution image to acquire the high-resolution objects. Secondly, a linear injection model is introduced to fuse intersected high- and low-resolution objects to generate preliminary prediction. However, the pixel-by-pixel computation inevitably produces uncertainties caused by complexity of the formulation. To enhance the stability and robustness of the whole fusion process, spatial filtering is implemented in each object to get the final prediction. The flowchart of the proposed OBSTFM is shown in Figure 1.

**Multi-resolution segmentation**

The features covered by remote sensing images generally show complex distribution characteristics. In order to acquire the proper shape and spectral features and the segmentation scales of surface features, the popular multi-resolution segmentation algorithm (MRSA) of eCognition software is applied to conduct segmentation, which is a bottom-up segmentation technique merging adjacent pixels or objects to create homogenous objects, and pixels in each object are spectrally similar and spatially adjacent (Benz et al., 2004). In the process of multi-resolution segmentation, several parameters including smoothness weight, spectrum weight and scale value need to be set appropriately to obtain segmentation objects, and in which, scale value (SP) is the key parameters to control the internal heterogeneity and the average size of objects (Drăguț et al., 2014). Only by setting appropriate parameters can we get the object which is consistent with the actual ground feature information. The weight of spectrum parameter is generally set to a larger value than shape weight to get a better segmentation accuracy because of the rich spectral information in high-resolution data. The setting of optimal image segmentation parameters is more complicated, in order to get a better segmentation scale, the scale value is set to 150 through performing “trial-and-error” optimization. Smoothness weight and spectrum weight are empirically set to 0.5 and 0.6 respectively as the default values.

**Linear injection model**

To fuse the input high- and low-resolution images in each object, a simple linear interpolation model is considered in the second step, which is commonly used for pansharpening. A general formulation of the pansharpening scheme (Vivone et al., 2015) is given as:

\[ M_{b}^{HR} = M_{b} + g_{b}(P - P_{low}) \]  

where the subscript \( b \) indicates the \( b \)th spectral band, \( M_{b} \) is the \( b \)th band of the interpolated multispectral image, \( M_{b}^{HR} \) is the reconstructed high-resolution of \( M_{b} \), \( P \) represents the panchromatic image, and \( P_{low} \) is obtained by low-pass filtering \( P \), \( g_{b} \) is the injection gain.

In spatiotemporal fusion, the auxiliary high-resolution data at the base date are used to enhance the resolution of low-resolution data at prediction phase. It is reasonable to apply relevant spatial detail from auxiliary high-resolution data to enhance the resolution of low-resolution data at prediction time while preserving necessary spectral information. Therefore, the spatiotemporal fusion can be taken as the pansharpening problem, it is described as follows:

\[ \hat{H}_{2}(b) = L_{2}(b) + g_{b}\left(H_{1}(b) - L_{1}(b)\right) \]  

here \( \hat{H}_{2}(b) \) is the fused result of band \( b \) for the high-resolution image at prediction phase, \( H_{1}(b) \) is the high-resolution image at base date of band \( b L_{1}(b) \) and \( L_{2}(b) \) are the resampled low-resolution images at base phase and prediction phase of band \( b \) with the scale of the high-resolution image, \( g_{b} \) is the injection gain which affects the accuracy of the model significantly, an algorithm that relies upon the optimization of injection coefficients by a least squares method is applied (Aiazzi et al., 2007; Vivone et al., 2015) to obtain \( g_{b} \):

\[ g_{b} = \frac{\text{cov}(L_{1}(b), L_{2}(b))}{\text{var}(L_{1}(b))} \]  

where \( \text{cov}(L_{1}(b), L_{2}(b)) \) denotes the covariance between \( L_{1}(b) \) and \( L_{2}(b) \), and \( \text{var}(L_{1}(b)) \) means the variance of \( L_{1}(b) \).

This model was introduced for spatiotemporal fusion recently because of its simplicity and effectiveness (Sun et al., 2019). In the fusion method proposed by Sun et al. (Sun et al., 2019), the linear injection model is applied in the entire image (simplified as global linear injection mode here), from the fusion results, we can see that the boundary of some features in the fused results is deblurred to some extent and some spatial details are weakened. In this step, the low-resolution images were downscaling to the same resolution as the high-resolution image by a bicubic interpolation operator. The fused process can be formulated as follows,
\[ H_2^{\text{LLIM}}(O_i, b) = L'_2(O_i, b) + g_b \left( H_1(O_i, b) - L'_1(O_i, b) \right) \]  
\[ (4) \]

in which \( H_1(O_i, b) \) represents band \( b \) value in the \( i \)th segmented object \( O_i \) of high-resolution image at \( t_1 \), \( L'_1(O_i, b) \) and \( L'_2(O_i, b) \) have the same meaning with \( H_1(O_i, b) \).

But for the resampled low-resolution image at \( t_1 \) and \( t_2 \), \( H_2^{\text{LLIM}}(O_i, b) \) is the preliminary fused objects \( O_i \) at \( t_2 \) using this linear injection model. Accordingly, \( g_b \) is defined as follows:

\[ g_b = \frac{\text{cov}(L'_1(O_i, b), L'_2(O_i, b))}{\text{var}(L'_1(O_i, b))} \]
\[ (5) \]

Although this process of fusing images in objects ensures better spatial fidelity, some abnormal values might occur in the fused result because of the noise contained in all input images. To mitigate these abnormal values, a global threshold \( \text{thre} \) is set to filter noisy pixels,

\[ \text{thre} = 3 \times \text{std}(H_2^{\text{GLIM}}(O_i, b)) \]
\[ (6) \]

where \( H_2^{\text{GLIM}}(O_i, b) \) denotes the value of object \( O_i \) calculated using the global linear interpolation model in Sun et al. (2019), which can be obtained in Equation (2). Therefore, in Equation (4), the values of fused objects less than or equal to \( \text{thre} \) are regarded as preliminary fused objects, whereas the values greater than \( \text{thre} \) are substituted using the estimation computed from the global linear interpolation model in Sun et al. (2019).

**Spatial filtering**

After acquiring the preliminary prediction from the previous steps, considering the inevitably uncertainties caused by pixel-by-pixel computation, the neighborhood information is used to reduce this kind of uncertainties and further enhance the robustness in blending areas that land cover changed. Some research (Gao et al., 2017; Zhu et al., 2016) just selected similar pixels from input high-resolution data in the base date for further computation without considering possible complex changes between different dates. Chen et al. (2018) and Sun et al. (2019) optimized the selection strategy by using the intersection of selected pixels in the base data and the initially predicted result in the prediction date. For the method of selecting similar pixels based on the moving window, it is necessary to select the appropriate window size to optimize the selection of similar pixels through trial-and-error optimization. However, the object-oriented method can directly select similar pixels from the homogeneous segmentation object, avoiding the window size parameter setting. To facilitate the selection of similar pixels and avoid the uncertainties introduced by the setting of moving window size, the local information in each object is applied for spatial enhancement.

After implementing multi-resolution segmentation to input high-resolution image, the segmented objects are often homogenous with similar spatial and spectral characteristics. Considering that complex changes might occur for different objects in different time phases, the intersection of selected pixels in the object input base image \( H_1(O_i, b) \) and preliminary fused results \( H_2^{\text{LLIM}}(O_i, b) \) are defined as final similar pixels, and two times of the standard deviation of each object are set as the threshold to select similar pixels. The selected strategy is formulated as

\[ \left| H_1(O_i, b) - H_1(P_i^t, b) \right| \leq 2 \times \text{std}(H_1(O_i, b)) \]
\[ (7) \]

\[ \left| H_2^{\text{LLIM}}(O_i, b) - H_2^{\text{LLIM}}(P_i^t, b) \right| \leq 2 \times \text{std}(H_2^{\text{LLIM}}(O_i, b)) \]
\[ (8) \]

where, \( H_1(P_i^t, b) \) and \( H_2^{\text{LLIM}}(P_i^t, b) \) represent the value of \( t \)th target pixel \( P \) in \( i \)th object \( O_i \) input base image and preliminary estimation, respectively. The schematic diagram of optimal selection strategy of similar pixels is shown in Figure 2. We define the final selected similar pixels in each object of two phases as \( H_1(P_k^t, b) \) and \( H_2^{\text{LLIM}}(P_k^t, b) \), respectively, where \( k \) means \( k \)th similar pixel. Then, the selected most similar \( N \) pixels are used to spatially filter the preliminary fused result

![Figure 2](Image-Link)
conducted from the previous steps. This strategy avoids the error caused by the incorrect selection of similar pixels to some extent. The weight of each similar pixel $W_k$ is computed based on spatial dependence between the selected similar pixels and the target pixel. The spatial distance $d_k$ is defined as Equation (9) and normalized to an appropriate range using Equation (10):

$$d_k = \sqrt{(x_k - x_t)^2 + (y_k - y_t)^2}$$

$$DN_k = (d_k - d_{min}) / (d_{max} - d_{min}) + 1$$

where $(x_k, y_k)$ and $(x_t, y_t)$ denote the location of similar pixels and the target pixel, $d_{min}$ and $d_{max}$ indicate the minimum value and maximum value, respectively. Then, the weight of $k$th similar pixel is computed as:

$$W_k = (1/DN_k) / \sum_{k=1}^{N} (1/DN_k)$$

To obtain the change information in different time phases and produce a more robust prediction, the preliminary fusion result produce from linear injection model is applied here. All selected similar pixels from preliminary prediction $H_{1}^{LIM}(P_t^i, b)$ and the high-resolution images at base date $H_1(P_t^i, b)$ are weighted to compute the change information. Thus, the final estimation $H_2(P_t^i, b)$ for $k$th target pixel in band $b$ is calculated as:

$$H_2(P_t^i, b) = H_1(P_t^i, b) + \sum_{k=1}^{N} W_k * (H_2^{LIM}(P_t^i, b) - H_1(P_t^i, b))$$

Experiments and results

The proposed OBSTFM was tested by real satellite images in two commonly used sites (Emelyanova et al., 2013; Zhu et al., 2016). Landsat and MODIS data are considered as the high- and low-resolution test data. To eliminate the inconsistencies between Landsat and MODIS images and directly test the performance of fusion methods, simulated MODIS-like images are aggregated from corresponding Landsat images. Two study sites experiencing different landscape changes are selected for experiments. The first site is located in Coleambally Irrigation Area, Southern New South Wales, Australia. Landsat images covering 800 × 800 pixels on the base date and prediction date were acquired on 3 December 2001 and 4 January 2002, respectively. The second site we chose is also covering 800 × 800 pixels with land cover change as main type, which is located in northern New South Wales, Australia. Most of land cover in this site changed from croplands to water because of a large flood. In this experiment, we chose Landsat and MODIS-like image pair on 26 November 2004 and MODIS-like image aggregated from Landsat image on 12 December 2004 as input data to predict the Landsat image on 12 December 2004. The false color composites of these two datasets are shown in Figure 3 and Figure 4.

In the practical application of spatiotemporal fusion technique, the performance of the spatiotemporal fusion algorithm is significantly affected by registration error, differences in sensor characteristics and other factors (Shen et al., 2013; Tang et al., 2020). In order to further explore the performance of the proposed algorithm in fusing real MODIS data and Landsat data, another experiment with images experiencing fast phenology change was added. Similarly, from the dataset in the northern New South Wales provided by Emelyanova et al. (Emelyanova et al., 2013), two Landsat images acquired on 16 April 2004 and 5 July 2004 and their corresponding MODIS images are selected for this experiment. MODIS images have been reprojected and resampled to the same resolution with Landsat images. Landsat and MODIS images are geo-referenced by selecting corresponding control points. All these images are resized to cover the same area with 800 × 800 pixels, and Figure 5 shows the false color composites of these images.

Before the fusion experiments, the multi-resolution segmentation algorithm (MRSA) of eCognition software is applied to segment the Landsat images of the first phase, Figure 6 shows the segmentation results of three sites.

In these three experiments, STARFM and FSDAF are selected as comparison methods to test the performance of OBSTFM. To quantitatively evaluate the accuracy of three fusion methods, six indices in Sun et al. are used: the root mean square error (RMSE), structure similarity (SSIM) (Z. Wang et al., 2004), correlation coefficient (r), universal image quality index (UIQI) (Z. Wang & Bovik, 2002), erreur relative globale adimensionnelle de synthèse (ERGAS) (Du et al., 2007) and spectral angle mapping (SAM) (Dennison et al., 2004). The predictions are more accurate when the values of RMSE, SAM and ERGAS are smaller and the values of UIQI, SSIM and r are higher, in which smaller SAM indicates better spectral fidelity, and higher SSIM means more similar spatial structure between predicted image and true image.

Experiment 1

In this experiment, Landsat and MODIS-like image pair on 3 December 2001 and MODIS-like image on 4 January 2002 are chosen as the input data to predict Landsat image on 4 January 2002. There was no distinct type change in this heterogeneous image but rapid phenological changes
between two time periods. The phenological changes of farmlands have no significant shape change, it can be considered that the segmentation results have no significant difference in the two phases. Theoretically, OBSTFM is very suitable for the fusion of this site. The performance of OBSTFM is compared with STARFM and FSDAF. The original Landsat data and fused results of three fusion algorithms are shown in Figure 7, and partial area of them are enlarged in Figure 8 to clearly compare the performance of three methods, and the difference areas between proposed and comparison methods have been marked with yellow circles. As we can see from Figure 7, the fusion results from FSDAF and OBSTFM show better spectral and spatial similarity to original Landsat data than that of STARFM. From the more detailed information demonstrated in Figure 8, the ground objects in prediction from STARFM are characterized as fragmented phenomenon without the consistency of spectral and spatial textures. The performance of FSDAF is better than that of STARFM in object homogeneity. However, OBSTFM shows best performance because of its best spectral similarity and spatial texture in objects to original data.

To quantitatively evaluate the performance of three fusion methods, six indices including RMSE, SSIM, r, QI, ERGAS and SAM are applied to access their performance in different aspects. The computation results of six indices are demonstrated in Table 1, in which the optimal accuracy results have been shown in bold. From Table 1 we can see that, among three methods, FSDAF provides better prediction accuracy than that of STARFM for all six bands, OBSTFM shows more powerful prediction ability that the other two algorithms because it achieves best performance in all six indices with the smallest RMSE, ERGAS and SAM and highest SSIM, QI and r. Smallest RMSE and ERGAS represent the best prediction accuracy, and smallest SAM indicates its better spectral fidelity than the other two methods. Furthermore, higher SSIM, QI and r of results fused by OBSTFM demonstrate its best ability in capturing spatial information and predicting pixels, and the results of quantitative assessment are consistent to that of visual evaluation.
Experiment 2

In this site, most of land cover changed from croplands to water because of a large flood. The purpose of Experiment 2 is to show the performance of the proposed algorithm in retrieving land cover changes with shape changes, and test the robustness of the strategy that incorporating segmentation technology, linear injection model and spatial filtering in retrieving large-scale abrupt land cover changes. Figure 9 presents original Landsat image of 12 December 2004 and its corresponding predictions fused by the three methods. It is apparent that the image OBSTFM predicted retains the most spatial details and has a satisfactory performance in retrieving land cover changes. To show more details, the results for the sub-area (marked in yellow) are shown in Figure 10, in which, the difference areas between proposed and comparison methods have been marked with yellow circles. It is apparent that the result predicted by OBSTFM seems to be more similar to original Landsat image than those predicted by STARFM and FSDAF. The sub-area shows that STARFM and FSDAF tend to deblurred some objects with less clear boundaries. In contrast, OBSTFM effectively predicts continuous spatial and spectral information in objects as well as captures better land cover reflectance changes in pixels than the other two methods. The quantitative indices calculated from predictions and original Landsat image shown in Table 2 demonstrated that three fusion methods have certain ability to capture changed information in different time phases. For all six bands, OBSTFM provided best prediction accuracy with smallest RMSE, ERGAS and SAM and highest SSIM, QI and $r$, which indicates that OBSTFM has better ability to capture spectral and spatial information between base date and prediction date than STARFM and FSDAF. FSDAF has lower accuracy than OBSTFM, but it performs much better than STARFM because of smaller RMSE, ERGAS and SAM and higher SSIM and $r$ for all six bands.

Experiment 3

The prediction results generated using three different spatiotemporal methods in Experiment 3 are shown in Figure 11. The sub-area (marked in
yellow) in Figure 11 is shown in Figure 12 to show more details. Due to the strong temporally changes between base date and prediction date, the prediction fused by STARFM shows obvious spectral difference in comparison to the real image. FSDAF reproduced more accurate prediction that STARFM with more similar spectral information. However, the phenology change causes the assumption that temporal increments of each class are the same among all pixels is not strictly correct, this problem introduces significant errors in the step of unmixing in FSDAF. As a result, the fused results

Figure 5. Test data in Experiment 3: Landsat data on (a) 16 April 2004 and (b) 5 July 2004, (c) and (d) are corresponding resampled real MODIS images to (a) and (b).

Figure 6. The segmentation results of Landsat data at the base phase in (a) Experiment 1, (b) Experiment 2 and (c) Experiment 3.
Figure 7. (a) Original Landsat image on 4 January 2002 and fused images by (b) STARFM, (c) FSDAF, and (d) OBSTFM.

Figure 8. Enlarged marked area in Figure 7(a) Original Landsat image on 4 January 2002 and in fused images by (b) STARFM, (c) FSDAF, and (d) OBSTFM.

Table 1. Accuracy evaluation of fusion results of STARFM, FSDAF and the proposed algorithm in Experiment 1.

|       | RMSE  | SSIM  | r    | QI   | ERGAS | SAM  |
|-------|-------|-------|------|------|-------|------|
| STARFM |       |       |      |      |       |      |
| Band1 | 0.0164| 0.9039| 0.8607| 0.7909| 1.2730| 5.8174|
| Band2 | 0.0244| 0.8598| 0.8535| 0.7910|       |      |
| Band3 | 0.0339| 0.8207| 0.9058| 0.8519|       |      |
| Band4 | 0.0650| 0.5861| 0.6075| 0.4906|       |      |
| Band5 | 0.0437| 0.7726| 0.9150| 0.8438|       |      |
| Band6 | 0.0347| 0.8204| 0.9245| 0.8610|       |      |
| Band7 | 0.0321| 0.8833| 0.8517| 0.7944|       |      |
|       |       |       |      |      |       |      |
| FSDAF |       |       |      |      |       |      |
| Band1 | 0.0162| 0.9039| 0.8607| 0.7909| 1.2687| 5.0208|
| Band2 | 0.0243| 0.8598| 0.8535| 0.7910|       |      |
| Band3 | 0.0339| 0.8207| 0.9058| 0.8519|       |      |
| Band4 | 0.0650| 0.5861| 0.6075| 0.4906|       |      |
| Band5 | 0.0437| 0.7726| 0.9150| 0.8438|       |      |
| Band6 | 0.0347| 0.8204| 0.9245| 0.8610|       |      |
| Band7 | 0.0321| 0.8833| 0.8517| 0.7944|       |      |
|       |       |       |      |      |       |      |
| OBSTFM|       |       |      |      |       |      |
| Band1 | 0.0163| 0.9155| 0.8606| 0.8060| 1.2155| 4.8161|
| Band2 | 0.0236| 0.8993| 0.8619| 0.8184|       |      |
| Band3 | 0.0327| 0.8764| 0.9113| 0.8750|       |      |
| Band4 | 0.0559| 0.7227| 0.6909| 0.5677|       |      |
| Band5 | 0.0429| 0.8408| 0.9172| 0.8619|       |      |
| Band7 | 0.0332| 0.8727| 0.9304| 0.8771|       |      |
produced by FSDAF are blurry, and exists obvious errors in Nir band. The fusion results generated by OBSTFM seem to capture more changed information in prediction date and show more similar spatial information to real image. The quantitative evaluation data of the experimental results are shown in Table 3. For this site, the fused results produced by OBSTFM show smaller SAM and higher r than that of STARFM and FSDAF, which means OBSTFM can preserve more spectral information and be more relevant to the reference image. As for RMSE, SSIM and QI, the proposed

Table 2. Accuracy evaluation of fusion results of STARFM, FSDAF and the proposed algorithm in Experiment 2.

|          | RMSE  | SSIM  | r     | QI    | ERGAS | SAM   |
|----------|-------|-------|-------|-------|-------|-------|
| **STARFM** |       |       |       |       |       |       |
| Band1    | 0.0177| 0.8666| 0.7745| 0.4374| 2.1665| 11.7267|
| Band2    | 0.0247| 0.8200| 0.7336| 0.4339|       |       |
| Band3    | 0.0295| 0.7893| 0.7548| 0.4535|       |       |
| Band4    | 0.0388| 0.7008| 0.8282| 0.4596|       |       |
| Band5    | 0.0608| 0.5226| 0.7521| 0.3118|       |       |
| Band7    | 0.0473| 0.5695| 0.7336| 0.3290|       |       |
| **FSDAF** |       |       |       |       |       |       |
| Band1    | 0.0179| 0.8689| 0.7677| 0.4091| 2.1065| 10.0408|
| Band2    | 0.0241| 0.8372| 0.7782| 0.4185|       |       |
| Band3    | 0.0281| 0.8175| 0.7728| 0.4486|       |       |
| Band4    | 0.0358| 0.7410| 0.8508| 0.4942|       |       |
| Band5    | 0.0607| 0.5115| 0.7436| 0.2725|       |       |
| Band7    | 0.0450| 0.6047| 0.7466| 0.3206|       |       |
| **OBSTFM** |       |       |       |       |       |       |
| Band1    | 0.0171| 0.8851| 0.7892| 0.4715| 2.0098| 9.6191|
| Band2    | 0.0235| 0.8584| 0.7924| 0.4780|       |       |
| Band3    | 0.0275| 0.8427| 0.7858| 0.5042|       |       |
| Band4    | 0.0339| 0.7982| 0.8680| 0.5465|       |       |
| Band5    | 0.0570| 0.5957| 0.7791| 0.3406|       |       |
| Band7    | 0.0426| 0.6645| 0.7770| 0.3793|       |       |

Figure 9. (a) Original Landsat image on 12 December 2004 and fused images by (b) STARFM, (c) FSDAF, and (d) OBSTFM.
Figure 10. Enlarged marked area in Figure 9(a) Original Landsat image on 12 December 2004 and in fused images by (b) STARFM, (c) FSDAF, and (d) OBSTFM.

Figure 11. (a) Original Landsat image on 16 March 2004 and fused images by (b) STARFM, (c) FSDAF, and (d) OBSTFM.
OBSTFM shows better performance than STARFM and FSDAF in most bands. Moreover, Figure 13 shows the scatter plots of predicted values for NIR band compared with actual values by the three methods, which also demonstrated that OBSTFM produced predicted values closer to the actual values than the other two methods and shown best consistency to the actual image. Generally, compared to STARFM and FSDAF, the proposed OBSTFM method shows advantages in preserving spatial information and produce prediction result temporally close to the information in prediction date.

**Conclusion**

This paper presents a new method named OBSTFM, including multi-resolution segmentation, linear injection model and spatial filtering. Multi-resolution segmentation aims to acquire segmented objects with homogeneous and similar information, linear
injection model is employed to fuse input data in each object, and then spatial filtering is implemented in each object to more accurately capture change information in different time phases and further produce more robust prediction. The proposed algorithm was tested using two datasets with simulated MODIS images experiencing phenology change and land cover type changes in heterogeneous landscape and one dataset with real MODIS images. Minimum input data are required to implement this method, and only the parameters in multi-resolution segmentation needed to be set without setting other parameters which might influence the sensitivity of the method in following steps. In comparison with STARFM and FSDAF, OBSTFM achieves better performance in predicting images with complex temporal changes, especially in non-shape changes. In this framework, it is worth mentioning that better alternative processing strategy can be developed to improve the performance of this algorithm.

Acknowledgments

The authors would like to thank the Editor and all reviewers whose insightful suggestions have significantly improved this paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by the National Natural Science Foundation of China under Grant 41971400, 41974039 and in part by the Fundamental Research Funds for the Central Universities under Grant 2019ZDPY09

ORCID

Hua Zhang http://orcid.org/0000-0003-0945-6613
Wenzhou Shi http://orcid.org/0000-0002-3886-7027
Dizhou Guo http://orcid.org/0000-0001-5325-5080

References

Aiazzì, B., Baronti, S., & Selva, M. (2007). Improving component substitution pansharpening through multivariate regression of MS+Pan data. *IEEE Transactions on Geoscience and Remote Sensing*, 45(10), 3230–3239. https://doi.org/10.1109/TGRS.2007.901007
Allen, R., Irmak, A., Trezza, R., Hendrickx, J. M. H., Bastiaanssen, W., & Kjaersgaard, J. (2011). Satellite-based ET estimation in agriculture using SEBAL and METRIC. *Hydrological Processes*, 25(26), 4011–4027. https://doi.org/10.1002/hyp.8408
Amorós-López, J., Gómez-Chova, L., Alonso, L., Guanter, L., Zurita-Milla, R., Moreno, J., & Camps-Valls, G. (2013). Multitemporal fusion of landsat/TM and ENVISAT/MERIS for crop monitoring. *International Journal of Applied Earth Observation and Geoinformation*, 23(1), 132–141. https://doi.org/10.1016/j.jag.2012.12.004
Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I., & Heynen, M. (2004). Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(3–4), 239–258. https://doi.org/10.1016/j.isprsjprs.2003.10.002
Chen, B., Chen, L., Huang, B., Michishita, R., & Xu, B. (2018). Dynamic monitoring of the poyang lake wetland by integrating Landsat and MODIS observations. *ISPRS Journal of Photogrammetry and Remote Sensing*, 139, 75–87. https://doi.org/10.1016/j.isprsjprs.2018.02.021
Chen, B., Huang, B., & Xu, B. (2017). Multi-source remotely sensed data fusion for improving land cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 124, 27–39. https://doi.org/10.1016/j.isprsjprs.2016.12.008
Cheng, Q., Liu, H., Shen, H., Wu, P., & Zhang, L. (2017). A spatial and temporal nonlocal filter-based data fusion method. *IEEE Transactions on Geoscience and Remote Sensing*, 55(8), 4476–4488. https://doi.org/10.1109/TGRS.2017.2692802
Dennison, P. E., Halligan, K. Q., & Roberts, D. A. (2004). A comparison of error metrics and constraints for multiple endmember spectral mixture analysis and spectral angle mapper. *Remote Sensing of Environment*, 93(3), 359–367. https://doi.org/10.1016/j.rse.2004.07.013
Drăguț, L., Csillik, O., Elisank, C., & Tiede, D. (2014). Automated parameterisation for multi-scale image segmentation on multiple layers. *ISPRS Journal of Photogrammetry and Remote Sensing*, 88, 119–127. https://doi.org/10.1016/j.isprsjprs.2013.11.018
Du, Q., Younan, N. H., King, R., & Shah, V. P. (2007). On the performance evaluation of pan-sharpening techniques. *IEEE Geoscience and Remote Sensing Letters*, 4(4), 518–522. https://doi.org/10.1109/LGRS.2007.896328
Emelyanova, I. V., McVicar, T. R., Van Niel, T. G., Li, L. T., & van Dijk, A. I. J. M. (2013). Assessing the accuracy of blending Landsat-MODIS surface reflectances in two landscapes with contrasting spatial and temporal dynamics: A framework for algorithm selection. *Remote Sensing of Environment*, 133, 193–209. https://doi.org/10.1016/j.rse.2013.02.007
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 Sensing of Environment*, 188, 9–25. https://doi.org/10.1016/j.rse.2016.11.004
Gao, F., Hilker, T., Zhu, X., Anderson, M., Masek, J., Wang, P., & Yang, Y. (2015). Fusing landsat and MODIS data for vegetation monitoring. *IEEE Geoscience and Remote Sensing Magazine*, 3(3), 47–60. https://doi.org/10.1109/MGRS.2015.2434351
Gao, F., Masek, J., Schwaller, M., & Hall, F. (2006). On the blending of the landsat and MODIS surface reflectance: Predicting daily landsat surface reflectance. *IEEE Transactions on Geoscience and Remote Sensing*, 44(8), 2207–2218. https://doi.org/10.1109/TGRS.2006.1772081
Gevaert, C. M., & García-Haro, F. J. (2015). A comparison of STARFM and an unmixing-based algorithm for landsat and MODIS data fusion. *Remote Sensing of Environment*, 156, 34–44. https://doi.org/10.1016/j.rse.2014.09.012
Guo, D., Shi, W., Hao, M., & Zhu, X. (2020). FSDAF 2.0: Improving the performance of retrieving land cover changes and preserving spatial details. Remote Sensing of Environment, 248, 111973. https://doi.org/10.1016/j.rse.2020.111973

Hilker, T., Wulder, M. A., Coops, N. C., Linke, J., McDermid, G., Masek, J. G., Gao, F., & White, J. C. (2009). A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on Landsat and MODIS. Remote Sensing of Environment, 113(8), 1613–1627. https://doi.org/10.1016/j.rse.2009.03.007

Huang, B., & Song, H. (2012). Spatiotemporal reflectance fusion via sparse representation. IEEE Transactions on Geoscience and Remote Sensing, 50(10 PART1), 3707–3716. https://doi.org/10.1109/TGRS.2012.2186638

Kong, F., Li, X., Wang, H., Xie, D., Li, X., & Bai, Y. (2016). Land cover classification based on fused data from GF-1 and MODIS NDVI time series. Remote Sensing, 8(9), 741. https://doi.org/10.3390/rs8090741

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

Li, J., Li, Y., He, L., He, L., Chen, J., & Plaza, A. (2020). Spatio-temporal fusion for remote sensing data: An overview and new benchmark. Information Sciences, 63 (140301), 1–140301. https://doi.org/10.1016/s1423-0191-2020-2785-5

Li, D., Zhao, Y., Fu, Y., Pan, Y., Yu, L., & Xin, Q. (2017). High resolution mapping of cropping cycles by fusion of landsat and MODIS data. Remote Sensing, 9(12), 1–19. https://doi.org/10.3390/rs9121232

Li, X., Foody, G. M., Boyd, D. S., Ge, Y., Zhang, Y., Du, Y., & Ling, F. (2020). SFSDAF: An enhanced FSDAF that incorporates sub-pixel class fraction change information for spatio-temporal image fusion. Remote Sensing of Environment, 237(July 2019), 111537. https://doi.org/10.1016/j.rse.2019.111537

Liao, C., & Wang, J. (2016). Evaluation of spatio-temporal data fusion methods for generating NDVI time series in cropland areas. 2016. IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2570–2573. https://doi.org/10.1109/IGARSS.2016.7729664

Liu, M., Ke, Y., Yin, Q., Chen, X., & Im, J. (2019a). Comparison of five spatio-temporal satellite image fusion models over landscapes with various spatial heterogeneity and temporal variation. Remote Sensing, 11(22), 2612. https://doi.org/10.3390/rs11222612

Liu, M., Yang, W., Zhu, X., Chen, J., Chen, X., Yang, L., & Helmer, E. H. (2019b). An improved flexible spatiotemporal data fusion (IFSDAF) method for producing high spatiotemporal resolution normalized difference vegetation index time series. Remote Sensing of Environment, 227, 74–89. https://doi.org/10.1016/j.rse.2019.03.012

Lu, M., Chen, J., Tang, H., Rao, Y., Yang, P., & Wu, W. (2016). Land cover change detection by integrating object-based data blending model of landsat and MODIS. Remote Sensing of Environment, 184, 374–386. https://doi.org/10.1016/j.rse.2016.07.028

Ma, J., Zhang, W., Marinoni, A., Gao, L., & Zhang, B. (2018). An improved spatial and temporal reflectance unmixing model to synthesize time series of landsat-like images. Remote Sensing, 10(9), 1388. https://doi.org/10.3390/rs10091388

Maselli, F., Chiesi, M., & Pieri, M. (2019). A new method to enhance the spatial features of multitemporal NDVI image series. IEEE Transactions on Geoscience and Remote Sensing, 57(7), 4967–4979. https://doi.org/10.1109/TGRS.2019.2894850

Nagendra, H., Lucas, R., Honrado, J. P., Jongman, R. H. G., Tarantino, C., Adamo, M., & Mairota, F. (2013). Remote sensing for conservation monitoring: Assessing protected areas, habitat extent, habitat condition, species diversity, and threats. Ecological Indicators, 33, 45–59. https://doi.org/10.1016/j.ecolind.2012.09.014

Shen, H., Wu, P., Liu, Y., Ai, T., & Wang, Y. (2013). A spatial and temporal reflectance fusion model considering sensor observation differences. International Journal of Remote Sensing, 34(12), 4367–4383. 1161. https://doi.org/10.1080/01431161.2013.777488

Shi, C., Wang, X., Zhang, M., Liang, X., Niu, L., Han, H., & Zhu, X. (2019). A comprehensive and automated fusion method: The enhanced flexible spatiotemporal data fusion model for monitoring dynamic changes of land surface. Applied Sciences, 9(18), 3693. https://doi.org/10.3390/app9183693

Song, H., & Huang, B. (2013). Spatiotemporal satellite image fusion through one-pair image learning. IEEE Transactions on Geoscience and Remote Sensing, 51(4), 1883–1896. https://doi.org/10.1109/TGRS.2012.2213095

Song, H., Liu, Q., Wang, G., Hang, R., & Huang, B. (2018). Spatiotemporal Satellite Image Fusion Using Deep Convolutional Neural Networks. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11(3), 821–829. https://doi.org/10.1109/JSTARS.2018.2797894

Sun, Y., & Zhang, H. (2019). A two-stage spatiotemporal fusion method for remote sensing images. Photogrammetric Engineering & Remote Sensing, 85(12), 907–914. https://doi.org/10.14358/PERS.85.12.907

Sun, Y., Zhang, H., & Shi, W. (2019). A spatio-temporal fusion method for remote sensing data using a linear injection model and local neighbourhood information. International Journal of Remote Sensing, 40(8), 2965–2985. https://doi.org/10.1080/01431161.2018.1538585

Tan, Z., Peng, Y., Di, L., & Tang, J. (2018). Deriving high spatiotemporal remote sensing images using deep convolutional network. Remote Sensing, 10(7), 1066. https://doi.org/10.3390/rs10071066

Tang, Y., Wang, Q., Zhang, K., & Atkinson, P. M. (2020). Quantifying the effect of registration error on spatio-temporal fusion. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 487–503. https://doi.org/10.1109/JSTARS.2020.2965390

Vivone, G., Alparone, L., Chanussot, J., Mura, M. D., Garzelli, A., Liccariadi, G. A., Restaino, R., & Wald, L. (2015). A critical comparison among pansharpening algorithms. IEEE Transactions on Geoscience and Remote Sensing, 53(5), 2565–2586. https://doi.org/10.1109/TGRS.2014.2361734

Wang, Q., & Atkinson, P. M. (2018). Spatio-temporal fusion for daily sentinel-2 images. Remote Sensing of Environment, 204, 31–42. https://doi.org/10.1016/j.rse.2017.10.046

Wang, Q., Tang, Y., Tong, X., & Atkinson, P. M. (2020). Remote sensing of environment virtual image pair-based spatio-temporal fusion. Remote Sensing of Environment, 249(July), 112009. https://doi.org/10.1016/j.rse.2020.112009
Wang, Z., & Bovik, A. C. (2002). A universal image quality index. IEEE Signal Processing Letters, 9(3), 81–84. https://doi.org/10.1109/79.995823

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4), 600–612. https://doi.org/10.1109/tip.2003.819861

Wei, J., Wang, L., Liu, P., Chen, X., Li, W., & Zomaya, A. Y. (2017). Spatiotemporal fusion of MODIS and Landsat-7 reflectance images via compressed sensing. IEEE Transactions on Geoscience and Remote Sensing, 55(12), 6791–6803. https://doi.org/10.1109/TGRS.2015.2448100

Wu, B., Huang, B., & Zhang, L. (2015). An error-bound-regularized sparse coding for spatiotemporal reflectance fusion. IEEE Transactions on Geoscience and Remote Sensing, 53(12), 6791–6803. https://doi.org/10.1109/TGRS.2015.2448100

Wu, M., Huang, W., Niu, Z., & Wang, C. (2015). Generating daily synthetic landsat imagery by combining landsat and MODIS data. Sensors, 15(9), 24002–24025. https://doi.org/10.3390/s150924002

Wu, M., Huang, W., Niu, Z., Wang, C., Li, W., & Yu, B. (2018). Validation of synthetic daily Landsat NDVI time series data generated by the improved spatial and temporal data fusion approach. Information Fusion, 40, 34–44. https://doi.org/10.1016/j.inffus.2017.06.005

Wu, M., Niu, Z., Wang, C., Wu, C., & Wang, L. (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. Journal of Applied Remote Sensing, 6(1), 063507. https://doi.org/10.1117/1.jrs.6.063507

Xie, D., Zhang, J., Zhu, X., Pan, Y., Liu, H., Yuan, Z., & Yun, Y. (2016). An improved STARFM with help of an unmixing-based method to generate high spatial and temporal resolution remote sensing data in complex heterogeneous regions. Sensors, 16(2), 207. https://doi.org/10.3390/s16020207

Xue, J., Leung, Y., & Fung, T. (2017). A bayesian data fusion approach to spatio-temporal fusion of remotely sensed images. Remote Sensing. 9(12), 1310. https://doi.org/10.3390/rs9121310

Zhang, B., Zhang, L., Xie, D., Yin, X., Liu, C., & Liu, G. (2016). Application of synthetic NDVI time series blended from landsat and MODIS data for grassland biomass estimation. Remote Sensing. 8(1), 1–21. https://doi.org/10.3390/rs810010

Zhang, H., Song, Y., Han, C., & Zhang, L. (2020). Remote sensing image spatiotemporal fusion using a generative adversarial network. IEEE Transactions on Geoscience and Remote Sensing, 99, 1–14. https://doi.org/10.1109/tgrs.2020.3010530

Zhao, Y., Huang, B., & Song, H. (2018). A robust adaptive spatial and temporal image fusion model for complex land surface changes. Remote Sensing of Environment, 208, 42–62. https://doi.org/10.1016/j.rse.2018.02.009

Zheng, Y., Wu, B., Zhang, M., & Zeng, H. (2016). Crop phenology detection using high spatio-temporal resolution data fused from SPOT5 and MODIS products. Sensors, 16(12), 2099. https://doi.org/10.3390/s16122099

Zhu, X., Cai, F., Tian, J., & Williams, T. K. A. (2018). Spatiotemporal fusion of multisource remote sensing data: Literature survey, taxonomy, principles, applications, and future directions. Remote Sensing, 10(4), 527. https://doi.org/10.3390/rs10040527

Zhu, X., Chen, J., Gao, F., Chen, X., & Masek, J. G. (2010). An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. Remote Sensing of Environment, 114(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 Sensing of Environment, 172, 165–177. https://doi.org/10.1016/j.rse.2015.11.016

Zhuok, B., Oertel, D., Lanlz, F., & Reinhäckel, G. (1999). Unmixing-based multisensor multiresolution image fusion. IEEE Transactions on Geoscience and Remote Sensing, 37(3), 1212–1226. https://doi.org/10.1109/36.763276

Zurita-Milla, R., Clevers, J. G. P. W., & Schaeppman, M. E. (2008). Unmixing-based landsat TM and MERIS FR data fusion. IEEE Geoscience and Remote Sensing Letters, 5, 453–457. https://doi.org/10.1109/LGRS.2008.919685