This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2023.3308902

Unmixing-based Spatiotemporal Image Fusion Based on the Self-trained Random Forest Regression and Residual Compensation

Xiaodong Li, Yalan Wang, Yihang Zhang, Shuwei Hou, Pu Zhou, Xia Wang, Yun Du, Giles Foody, *IEEE Fellow*

***Abstract*—Spatiotemporal satellite image fusion (STIF) has been widely applied in land surface monitoring to generate high spatial and high temporal reflectance images from satellite sensors. This paper proposed a new unmixing-based spatiotemporal fusion method that is composed of a self-trained random forest machine learning regression (R), low resolution (LR) endmember estimation (E), high resolution (HR) surface reflectance image reconstruction (R), and residual compensation (C), that is, RERC. RERC uses a self-trained random forest to train and predict the relationship between spectra and the corresponding class fractions. This process is flexible without any ancillary training dataset, and does not possess the limitations of linear spectral unmixing, which requires the number of endmembers to be no more than the number of spectral bands. The running time of the random forest regression is about ~1% of the running time of the linear mixture model. In addition, RERC adopts a spectral reflectance residual compensation approach to refine the fused image to make full use of the information from the LR image. RERC was assessed in the fusion of a prediction time MODIS with a Landsat image using two benchmark datasets, and was assessed in fusing images with different numbers of spectral bands by fusing a known time Landsat image (seven bands used) with a known time very-high-resolution PlanetScope image (four spectral bands). RERC was assessed in the fusion of MODIS-Landsat imagery in large areas at the national scale for the Republic of Ireland and France. The code is available at https://www.researchgate.net/profile/Xiao_Li52.**

***Index Terms*—Spatiotemporal image fusion, Landsat, unmixing, self-trained regression, sub-pixel analysis.**

This work was supported by the Natural Science Foundation of China (62071457, 42271400), the International Science and Technology Cooperation Project from Hubei Province, China (2022EHB018), the Key Research Program of Frontier Sciences, Chinese Academy of Sciences (ZDBS-LY-DQC034), Young Top-notch Talent Cultivation Program of Hubei Province, the Hubei Provincial Natural Science Foundation of China for Distinguished Young Scholars (2022CFA045), in part by the Key Research and Development Project of Hubei Province, China (2020BCA074), and in part by the Application Foundation Frontier project of Wuhan (2020020601012283). The authors would like to thank Emelyanova et al. for providing the Landsat and MODIS data used on the CIA and LGC sites in Australia, and would like to thank the Planet Labs Company for providing images for research analysis.

X. Li is with the Key Laboratory for Environment and Disaster Monitoring and Evaluation, Hubei, Innovation Academy for Precision Measurement Science and Technology, Chinese Academy of Sciences, Wuhan 430077, China, and with (Germany) The Offshore International Science and Technology Cooperation Center of Frontier Technology of Geodesy.

Y. Zhang and Y. Du are with the Key Laboratory for Environment and Disaster Monitoring and Evaluation, Hubei, Innovation Academy for Precision Measurement Science and Technology, Chinese Academy of Sciences, Wuhan 430077, China (e-mail: lixiaodong@whigg.ac.cn).

P. Zhou and Y. Wang are with the Key Laboratory for Environment and Disaster Monitoring and Evaluation, Hubei, Innovation Academy for Precision Measurement Science and Technology, Chinese Academy of Sciences, Wuhan 430077, China, and University of Chinese Academy of Sciences, Beijing 100049, China.

S. Hou is with China Academy of Space Technology (Xi'an), Xi'an, China.

X. Wang is with CAS Key Laboratory of Aquatic Botany and Watershed Ecology, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China.

G. M. Foody is with School of Geography, University of Nottingham, Nottingham, NG7 2RD, UK.

I. INTRODUCTION

Accurate monitoring of the Earth's land surface is crucial for understanding the environment and its dynamics [4, 5]. Monitoring of large-area Earth surface dynamics has been greatly facilitated by the development of satellite remote sensing techniques. For instance, the Advanced Very High Resolution Radiometer (AVHRR) enabled monitoring of the Earth at a spatial resolution of approximately 1 km and on a potentially daily basis. Moderate Resolution Imaging Spectroradiometer (MODIS) can also monitor the Earth's surface on a daily basis with a spatial resolution of 250–500 m. High Resolution (HR) imagery provides monitoring capabilities at resolutions typically finer than 100 m, but less frequently than Low Resolution (LR) imagery. The Landsat series data has been monitoring the Earth every 16-18 days at a 15-80 m resolution for approximately 50 years [7]. Sentinel-2 multispectral imagery has provided a revisit frequency of approximately 5 days since 2016 [10]. Although multiple optical satellite remote sensing systems allow the monitoring of the Earth, satellite remote sensing is often limited by the trade-off between spatial and temporal resolutions [11-14].

Spatiotemporal satellite image fusion (STIF) is a technique that fuses low-spatial but high-temporal resolution imagery with high-spatial but low-temporal resolution imagery to generate imagery with not only high-spatial resolution but also high-temporal resolution, utilizing the advantages of each data source [1, 12, 15-19]. Many STIFs have been proposed recently including weighted-function-based STIF [1, 8, 20, 21], Bayesian-based STIF [24, 25], learning-based fusion [26-35], and hybrid methods that combine two or more fusion models [20, 36-45]. These STIFs have been applied not only in the fusion of reflectance images but also in vegetation indices [46-50], surface temperature [51-53], evapotranspiration [54], impervious surfaces [56], land cover [11], snow cover [57], and surface water [58-61].

Different from the aforementioned STIFs, the other group of STIF models is the unmixing-based STIF which reconstructs an image at the spatial resolution of the HR image at a known time while preserving spectral reflectance information from the LR image at the prediction time [6, 62, 63]. The classic

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2023.3308902

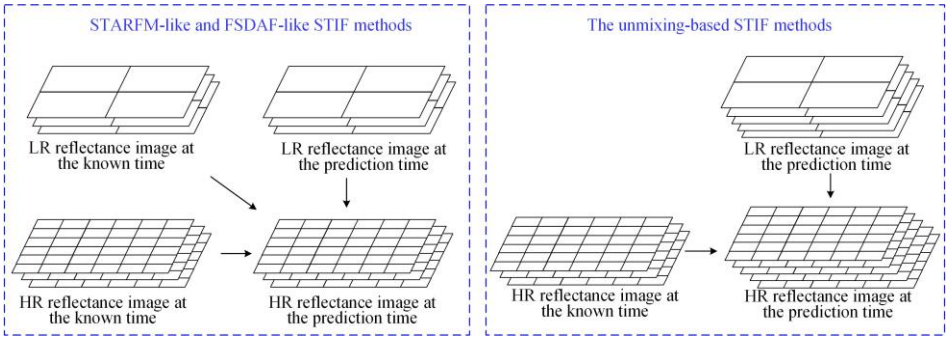


Fig. 1. A flowchart showing the main differences between the STARFM-like and FSDAF-like STIF methods and the unmixing-based STIF methods. The STARFM-like and FSDAF-like methods require the LR image at the known time as input and require the LR and HR images to have similar spectral bands, whereas the unmixing-based STIF does not use the LR image at the known time as input and can fuse LR and HR images with different spectral bands. 'HR' represents high spatial resolution, and 'LR' represents low spatial resolution.

unmixing-based STIF is composed of four steps: (1) classifying the HR image into a multi-class land cover map, (2) calculating the LR class fraction images based on the HR multiclass land cover map, (3) estimating the LR image endmember spectra based on the LR image and the corresponding class fraction images at the prediction time, and (4) assigning the estimated endmember spectra to the HR land cover map according to the corresponding pixel labels. Unlike the aforementioned STIF models such as STARFM [1] and FSDAF [38] that require the HR and LR images to have the same or similar spectral bands, the unmixing-based STIF can fuse HR and LR images with different spectral bands and is more flexible than other fusion methods. For instance, the unmixing-based STIF has been used in the fusion of Landsat TM and MERIS to generate a 30 m resolution multispectral image that preserves MERIS's spectral information of 15 spectral bands [63, 68]. The unmixing-based STIF can fuse HR and LR are acquired at different dates, and is different from the spatial-spectral fusion or pan-sharpening which requires the HR and LR to be acquired from the same sensor or acquired at the same or similar dates. In addition, the unmixing-based STIF requires a minimum number of inputs than other STIF models: a prediction time LR image and a known time (either predates or postdates the prediction time) HR image (Fig. 1), and is thus more flexible than the STARFM-like and FSDAF-like STIFs which require both the HR and LR imagery at the known time to be available.

The first unmixing-based STIF was a multisensor multiresolution image fusion [62], and has been greatly improved in recent years. Zurita-Milla et al. [63] fused Landsat TM with MERIS imagery to generate a 30 m resolution multispectral image that preserves MERIS's spectral information. Amorós-López et al. [68] proposed a model suitable for complex heterogeneous regions by fusing Landsat with MERIS imagery for crop monitoring. Liu et al. [66] applied the linear spectral unmixing model to generate a HR class fraction image, which was used as a substitute for the pixel-based hard classification map to enhance fusion accuracy in regions of heterogeneous land cover. Wang et al. [55] proposed a block-removed unmixing that effectively reduced the blocky effect in the STIF. Unmixing-based STIF has also been applied to generate multispectral or hyperspectral reflectance imagery and the corresponding vegetation indices at high spatiotemporal resolutions. For instance, Busetto et al. [69] estimated the time series sub-pixel NDVI images by the fusion of MODIS and Landsat imagery, and Zurita-Milla et al. [6] estimated vegetation indices to monitor the seasonal changes in vegetation by the fusion of MERIS full-resolution and Landsat imagery. Unmixing-based STIF has also been combined with weighted-function-based fusion to further utilize the advantages of each fusion method. For instance, Gevaert and García-Haro [65] combined the strengths of unmixing-based STIF and STARFM to make the fusion result less sensitive to reflectance changes. Xu et al. [70] modified the unmixing-based STIF by adding a regularization term of the endmember spectra to ensure that the extracted endmember spectra did not differ greatly from the predefined endmember spectra. Jiang and Huang [71] used two spectral unmixing approaches and STARFM to reduce blurring problems. Although the unmixing-based STIF has several advantages over other STIFs, several limitations exist.

First, most unmixing-based fusion methods assume that the HR image pixels are pure and assign each HR pixel to a single class based on clustering or classification algorithms applied to the HR image. If the neighboring HR pixels within the same LR pixel are labeled with the same classes, they are assigned the same spectra in the fused image. As a result, the fusion homogenizes the spectra for the neighboring HR pixels with the same class, that is, the homogenization effect. This phenomenon results in an inability to represent intra-class spectral variability for the neighboring same-class HR pixels for these unmixing-based STIFs [55, 63, 66]. To address the mixed pixel problem that is also common with HR imagery and to reduce the homogenization effect in the unmixing-based STIF, multitemporal fusion (MTF) uses a soft clustering algorithm to map HR class fraction images [68]. However, the MTF unmixing result may be sensitive to the fuzzy parameters used in the soft clustering algorithm. The unmixing would resemble the result of hard classification if the fuzzy parameter is small, and result in similar class fractions for all classes in the HR pixel if the fuzzy parameter is very large [72]. The linear spectral unmixing-based spatiotemporal data fusion model (LSUSDFM) has used the fully constrained least squares linear spectral mixture analysis (FCLS) algorithm to spectrally unmix the HR reflectance image [66]. Anyway, the FCLS is an inversion problem and is ill-posed when the number of clusters is larger than the number of spectral bands in the HR image,

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2023.3308902

and fails to consider the intra-class spectral variabilities in endmembers in the unmixing and cannot deal with the multiple scattering effects using a linear mixture model. Moreover, the FCLS inversion is an optimization approach that is usually time-consuming.

The second limitation of the current unmixing-based STIF model is that it only uses the LR image at the prediction time in endmember extraction and fails to fully use the LR image. In particular, the unmixing-based STIFs estimate LR image endmember spectra based on a set of LR pixels in a local window, with the target LR pixel as the window center, and assign the endmember spectra to the HR scale according to the pixel labels generated from the HR image to generate the fused image. The estimated endmembers do not reflect the information of solely the target LR pixel, but represent the averaged spectral information within the local window. As a result, the estimated LR endmembers may not represent a drastic reflectance change that occurred mainly in the target LR pixel. The using of residual compensation by comparing the fused with the observed LR image at the prediction time could enhance the accuracy of the FSDAF-like STIFs [38] and their derivations [36, 37, 42]. However, the FSDAF-like STIFs require an additional LR image at the known time in the residual compensation, based on the assumption that the HR and LR images have similar spectral bands. This approach is thus not suitable for the unmixing-based STIFs which do not input the LR image at the known time and which may deal with HR and LR images with different spectral bands. To the best of our knowledge, the use of the LR image for residual compensation for the unmixing-based STIF has not yet been reported.

TABLE I

THE LOCATION AND AREA OF STUDY REGIONS USED IN THE RECENT AND STATE-OF-THE-ART STIFS FOR MULTI- AND HYPER-SPECTRAL REFLECTANCE IMAGE FUSION. THE FUSIONS OF SINGLE-BAND DATA SUCH AS NDVI, TEMPERATURE, AND EVAPOTRANSPIRATION ARE NOT INCLUDED.

Location	Area (km^2)	Reference
Boreal Ecosystem–Atmosphere Study, Canada	1296	Gao et al. [1]
West-central Alberta, Canada	34,225	Hilker et al. [2]
Central British Columbia, Canada	34,225	Hilker et al. [3]
The central part of the Netherlands	2400	Zurita-Milla et al. [6]
Central Virginia, USA	~180	Zhu et al. [8]
Near Jiangsu Province, China	34,225	Wu et al. [9]
Flagstaff, Arizona, USA	13,949	Walker et al. [22]
Southern New South Wales, Australia	2193 and 5540	Emelyanova et al. [23], Zhu et al. [38], Wang et al. [55], Guo et al. [36], Shi et al. [41], Xu et al. [40]
North Central Montana, USA	14,685	Watts et al. [64]
Barrax region near Albacete, Spain	576	Gevaert and Garc á-Haro [65]
In the Henan province and the Longyangxia Reservoir, China	700 and 1300	Liu et al. [66]
Taiyuan, China	3600	Liu et al. [29]
Central Iowa, USA	18,225	Jia et al. [67]
Daxing, Tianjin, and Ar Horqin Banner, China	2421, 3723 and 6250	Li et al. [15]
Nanjing, China	3600	Chen et al. [35]

Finally, although the application of STIF in large areas has been studied, these STIFs are mainly applied for single band data such as vegetation index [3, 47] and land surface temperature band [51, 56], and the STIF study for multispectral reflectance imagery remains challenging because it involves more input bands and the solution is more complicated. For instance, the benchmark datasets [15, 23] and most test imagery used in STIF studies are limited in a spatial span, typically no larger than the range of one Landsat scene of 185 km ×185 km (34,225 km^2), as shown in Table I. In STIF for a very large area, it is necessary to mosaic a series of HR images acquired at different times as the LR image at the known time. Although the composition of large-area cloud-free images has been greatly facilitated by the online cloud computing platform of Google Earth Engine [73], it is very difficult to generate the corresponding LR images in which each pixel should have the same acquisition time as the corresponding HR pixels. This greatly limits the use of the popular STIFs such as STARFM and FSDAF which require the HR and LR imagery before the prediction time to be acquired at the same or similar time in large area image fusion. In contrast, the unmixing-based STIF only inputs the HR image without the corresponding LR image at the known time, and is thus not limited by the same/similar time requirement in the known time HR-LR image pair and is more flexible and has considerable potential for fusion over a very large area unlike the other STIFs. However, to the best of our knowledge, the use of large-area Google Earth Engine-composited imagery in unmixing-based STIF has not been reported.

In this paper, a novel Sub-pixel Unmixing-Based Data Fusion that is composed of a self-trained random forest machine learning regression (R), LR endmember estimation (E), and HR surface reflectance image reconstruction (R), and residual compensation (C), i.e., RERC, is proposed to address the limitations of the current STIFs. The RERC uses the HR class fraction information to reduce the homogenization effect in pixel-based classification. RERC uses a self-trained machine-learning regression model, which is automatic and computationally efficient, to map HR class fractions to overcome the limitations of current linear and soft clustering algorithms. In addition, RERC compares the fused image with the LR image at the prediction time and composites the residuals to refine the prediction image. This residual composition step is different from the FSDAF-like methods because RERC neither requires a LR image at the known time nor requires the HR and LR images to have similar spectral bands. The method has the minimum input in STIF, including a HR reflectance image at a known time and a LR reflectance image at the prediction time, and does not require the LR reflectance image at the known time and is thus more flexible than STARFM and FSDAF. RERC was assessed in three experiments. The first experiment compared the RERC with state-of-the-art unmixing-based STIFs in the fusion of prediction time MODIS images with a known time Landsat using two open-source benchmark datasets. The second experiment assessed the proposed method in fusing a prediction time Landsat with a known time very-high-resolution

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2023.3308902

PlanetScope image (four spectral bands) to assess RERC in the fusion imagery with different spectra bands. The third experiment assessed the proposed method in the fusion of MODIS-Landsat imagery in very large areas at the national scale for the Republic of Ireland (~70,273 km^2) and France (~551,500 km^2) to assess RERC at the national scale.

II. Methodology

The proposed RERC (Fig. 2) has four main steps: (1) generating the HR class fraction images at the known time (t_0) using a self-trained regression; (2) estimating the endmember spectra at the prediction time (t_p) and reconstructing a HR image at t_p according to the linear mixture model; and (3) refining the HR image to generate the final image at t_p based on residual compensation.

A. Unmixing the HR Images at t_0 Based on the Self-trained Regression

The RERC extracts HR class fraction images from the HR image at t_0. First, RERC applies unsupervised algorithms of the *k*-means clustering [74] to the HR image at t_0 to automatically generate a pixel-based HR land cover map with *n* classes. Then, RERC uses a self-trained regression of random forest to train the relationship between the spectral image and the corresponding class fractions at a coarse resolution scale with a pre-defined scale factor *z*. In particular, the inputted HR reflectance image is spatially degraded by averaging the spectra of all the HR pixels within the $z \times z$ window, and the HR land cover map at t_0 is spatially degraded to class fraction images by dividing the total number of HR pixels of a class by z^2. In RERC, the scale factor *z* was set to 10 so that the value of the minimal interval between the two class fraction was 1%. If *z* is too small, the corresponding class fractions may belong to only a limited number of values. For instance, if *z* is set to 2, then the spatially degraded image contains four HR pixels, and the corresponding class fraction for a class is one of the values of 0% (none of the pixels belong to this class), 25% (one pixel belongs to this class), 50% (two pixels belong to this class), 75% (three pixels belong to this class), and 100% (four pixels belong to this class). In contrast, if *z* is too large, the training data would be too complex [75, 76].

With the spatially degraded reflectance image and class fraction image as the training dataset, RERC uses random forest machine learning, which is a supervised ensemble-learning non-linear regression algorithm based on regression trees [77], to construct the regression relationship between the image pixel spectra and the corresponding class fractions. A random forest regression model was constructed for each class separately according to the degraded reflectance and class fraction images. For each class, RERC uses the $B^{HR} \times 1$ size (B^{HR} is the number of HR image bands) spectral vectors from all the degraded pixels as the input of the random forest regression model and uses the corresponding class fraction values for that class from all the corresponding degraded pixels as the output of the regression model. The trained random forest regression model for each class according to the degraded reflectance and class fraction images was then applied to the HR image at t_0 to predict class fractions at the HR scale. The class fraction regression model trained at a LR resolution scale

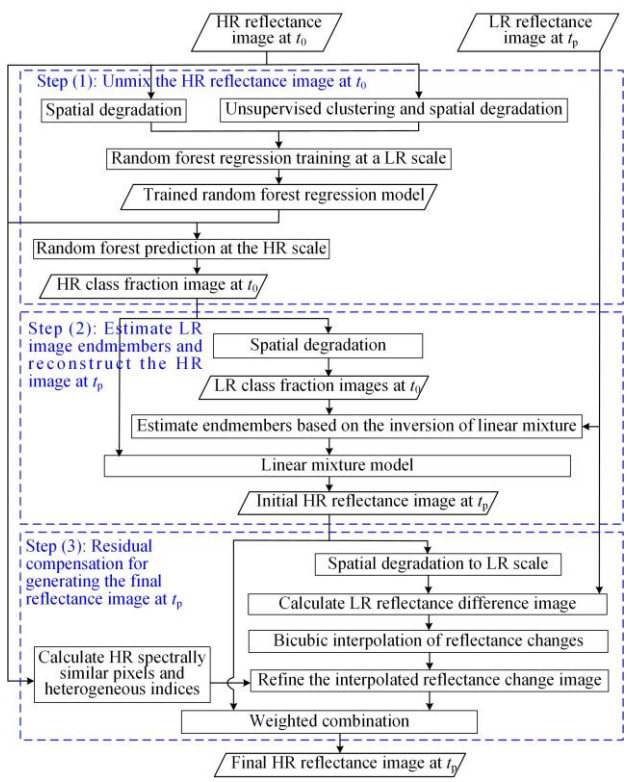


Fig. 2. RERC flowchart. The LR and HR imagery are not required to have the same/similar spectral bands.

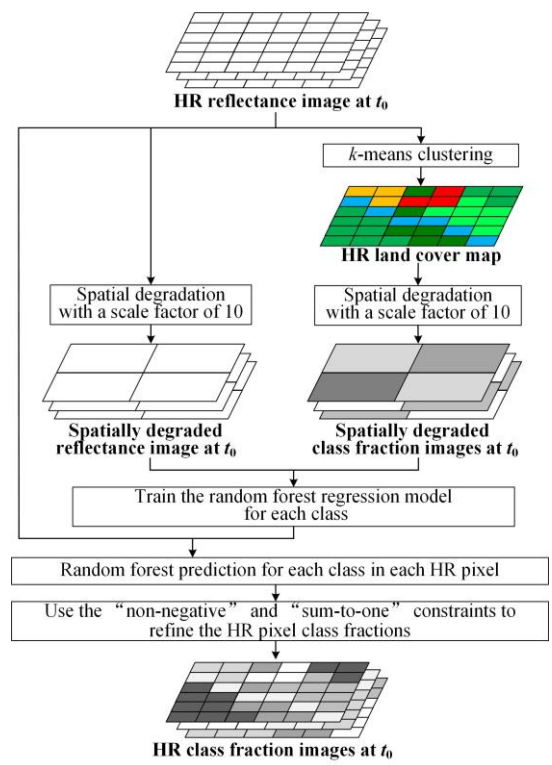


Fig. 3. Flowchart of the self-trained regression model that is used to estimate HR class fraction images at t_0. Scale factor *z* is set to 10 (the minimal interval between two class fraction values is, therefore, 1%). The different colors in the HR land cover map represent different land cover classes, and different gray images in the spatially degraded and HR class fraction images at t_0 represent different class fractions.

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2023.3308902

has been proven to be effective for predicting class fractions at a finer scale [75, 76]. For each HR pixel, the fraction for each class is divided by the sum of all class fractions in that HR pixel, such that all class fractions sum to one. A flowchart of the self-trained regression is shown in Fig. 3.

B. Estimating the LR Endmember Spectra and Reconstructing the HR Image at t_p

This step (step (2) in Fig. 2) is the same as the unmixing-based STIFs that explore the sub-pixel land cover information such as MTF and LSUSDFM. The endmember spectra at t_p were estimated based on the LR reflectance image at t_p and the corresponding class fractions at t_0. First, the HR class fraction image predicted from a random forest was spatially degraded to the LR scale by averaging the HR class fraction within each LR pixel according to s which is the scale factor between the input LR and HR imagery. With the LR image at t_p and the degraded class fractions at t_0, the endmember spectra for each LR pixel at prediction time t_p are estimated based on the inversion of a linear mixture model. To avoid collinearity in the estimation of endmembers, regularization-based linear unmixing is used to estimate the endmember for each LR pixel (i.e., local endmember) [66, 68, 70]. The endmembers used for regularization (i.e., global endmembers) are directly selected from the LR image at t_p to avoid the impact of atmospheric conditions and spectral bias [68]. For a target class, the spectral values from a set of LR pixels at t_p with the highest class abundance value of the target class are averaged as the global endmember of the target class [66]. Finally, for each LR pixel, the local endmember spectra within the LR pixel are estimated according to an inversion of a linear spectral mixture model using a group of LR pixels within a $w\times w$ sliding window, with the target LR pixel as the window center. For the ith LR pixel, the local endmember spectra of different classes in the bth band ($b\in 1, \cdots, B^{LR}$, where B^{LR} is the number of LR image bands.) are estimated based on k^2 neighboring LR pixels by minimizing the least square error:

$$\hat{\boldsymbol{e}}_{b,i} = \min\left\{\sum_{k=1}^{w^2}\left[y_{b,k} - \sum_{c=1}^{n} a_{c,k}\times e_{c,b,k}\right]^2 + \alpha\times\frac{w}{n}\times\sum_{c=1}^{n}\left(\hat{e}_{c,b,i} - \bar{e}_{c,b}\right)^2\right\} \tag{1}$$

where n is the number of classes, k is the number of LR pixels, w is the size of the local window, $\hat{\boldsymbol{e}}_{b,i}$ is the local endmember spectra in the bth band for the ith LR pixel, $y_{b,k}$ is the spectrum in the bth band in the kth LR pixel, $a_{c,k}$ is the class fraction of the cth class (c=1, 2, $\cdots$, n) in the kth LR pixel, $\hat{e}_{c,b,i}$ is the cth class spectrum in the bth band in the ith target LR pixel, $\bar{e}_{c,b}$ is the global endmember spectrum for the cth class in the bth spectral band, and α is the regularization parameter.

RERC predicts the reflectance image by linearly combining the estimated LR endmember spectra with the HR class fractions based on a linear mixture model:

$$\boldsymbol{y}_j = \boldsymbol{e}_j\boldsymbol{a}_j \tag{2}$$

where $\boldsymbol{e}_j$ is a $B^{LR}\times n$ sized local endmember spectra matrix for the jth HR pixel, $\boldsymbol{a}_j$ is an $n\times 1$ sized class fraction vector of different classes in the jth HR pixel, and $\boldsymbol{y}_j$ is the predicted $B^{LR}\times 1$ sized spectra in the jth HR pixel.

C. Residual Compensation for Generating the Final HR Image at t_p

RERC refines the fused HR image using information from the LR reflectance image based on residual compensation. RERC spatially degrades the fused HR image at t_p to the LR scale by averaging the reflectance values within the LR pixels according to the scale factor s, and compares it with the observed LR image at t_p to generate a spectral difference image. Assume $y_{b,i}$ is the observed spectrum in the bth LR band ($b\in 1, \cdots, B^{LR}$) in the ith LR pixel at t_p, and $\hat{y}_{b,j,i}$ is the estimated spectrum in the bth band in the jth HR pixel of the ith LR pixel. $\xi_{b,i}$ is the spectral difference between the observed spectrum $\hat{y}_{b,i}$ and the synthetic spectrum in the bth band in the ith LR pixel. The spectral residual value for the ith LR pixel in the bth LR image spectral band ($b\in 1, \cdots, B^{LR}$), $\xi_{b,i}$, is calculated as:

$$\xi_{b,i} = y_{b,i} - \frac{1}{s^2}\sum_{j=1}^{s^2}\hat{y}_{b,j,i} \tag{3}$$

$\xi_{b,i}$ was calculated at the LR scale, whereas the fused image spectrum in Eq. (2) is at the HR scale. To match this spatial resolution gap, $\xi_{b,i}$ is spatially interpolated to a HR scale of $\xi_{b,i}^{interpolation}$ using bicubic spatial interpolation with the scale factor s. Direct summation of $\xi_{b,i}^{interpolation}$ with the estimated HR spectrum may cause a blurring effect [38, 55], and the residual image is refined using a weighted sum of spectrally similar pixels within an $m\times m$ sized local square window from the HR image at t_0, assuming that spectrally similar pixels at t_0 would have similar spectral change [1, 38]. The spatial prediction value $SP_{b,j}$ for the jth HR pixel in the bth LR image spectral band is calculated as:

$$SP_{b,j} = \sum_{l=1}^{L} w_l \times \xi_{b,l}^{interpolation} \tag{4}$$

where L is the number of spectrally similar pixels in the inputted HR image. The spectrally similar pixels are selected based on a set of L HR pixels that have the smallest spectral differences in the reflectance image at t_0 [1, 36-38]. The weight of the lth ($l\in L$) HR pixel, w_l, in Eq. (4), is determined according to the geometric distance between the jth target HR pixel and the lth neighborhood pixel, $d_{l,j}$, as

$$w_l = \left(1/d_{l,j}\right)\Big/\sum_{l=1}^{L}\left(1/d_{l,j}\right) \tag{5}$$

The final spectrum in the jth HR pixel in the bth LR image band ($b\in 1, \cdots, B^{LR}$) for the unmixing-based STIF in the fusing of HR and LR imagery with the same spectral bands is calculated by a weighted sum of the fused image and the spatial prediction image as:

$$\hat{y}_{b,j}^{final} = y_{b,j} + HI_{b,j}\times SP_{b,j} \tag{6}$$

where $HI_{b,j}$ denotes the heterogeneous index for the jth HR pixel in the bth band ($B^{LR} = B^{HR}$). The spatial prediction image is generated from a spatial interpolation image at the LR scale, and the fusion may be blurred if the HR pixel is located at the

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2023.3308902

boundary between two land cover classes. The weight $HI_{b,j}$ in Eq. (6) is used to give a low value if the jth target pixel is located in the boundary region to preserve the edge and a high value if the jth target pixel is located in the homogeneous region for smoothing the region. The heterogeneous index $HI_{b,j}$ is calculated as

$$HI_{b,j} = \exp\left(-\left(Std_{b,j} / 0.05\right)^2\right) \tag{7}$$

where $Std_{b,j}$ is the standard deviation of the spectral reflectance ranging from 0 to 1 in the bth band in a 7×7 HR local window (the optimal window size is set through many trials) in the inputted HR image at t_0, with the jth HR pixel as its window center. Note that if the input HR image at t_0 represents the digital number (DN) value instead of the spectral reflectance ranging from 0 to 1, the DN values can be divided by the maximum DN value in the corresponding bands so that the calculated $Std_{b,i}$ ranges from 0 to 1. $HI_{b,j}$ is relatively small if $Std_{b,j}$ is high, which means that the target pixel is located in a heterogeneous region. $HI_{b,j}$ is relatively large if $Std_{b,j}$ is small, which means that the target pixel is located in a homogeneous region.

In the fusion of HR and LR imagery with dissimilar spectral bands, the averaged heterogeneous index in Eq. (8) is used as a substitute for the per-band heterogeneous index in Eq. (7), calculated using all spectral bands in the HR image as follows:

$$\begin{aligned}\hat{y}_{b,j}^{final} &= y_{b,j} + \overline{HI}_j \times SP_{b,j} \\ &= y_{b,j} + \sum_{b}^{B^{HR}} \exp\left(-\left(Std_{b,j} / 0.05\right)^2\right) \Big/ B^{HR} \times SP_{b,j}\end{aligned} \tag{8}$$

III. EXPERIMENTS

A. Experimental Data and Study Site

The proposed RERC was assessed in three experiments. The first experiment compared the RERC with other unmixing-based STIF (the STARFM-like and FSDAF-like methods which require more input data were not compared) in the fusion of a prediction time MODIS images with a known time Landsat in the Coleambally Irrigation Area (CIA) and in the Lower Gwydir Catchment (LGC), Australia, provided by Emelyanova et al. [23]. The second experiment assessed the proposed method in fusing images with different numbers of spectral bands by fusing a prediction time Landsat (30 m resolution, seven bands used) with a known time very-high-resolution PlanetScope image (3 m resolution, four spectral bands). The third experiment assessed the proposed method in the fusion of MODIS and Landsat imagery in very large areas at the national scale for the Republic of Ireland (~70,273 km^2) and France (~551,500 km^2).

1) Simulated and real Image Experiment on CIA and LGC sites

The CIA site is located in a heterogeneous farmland region in Australia. Two subsets of cloud-free Landsat images on November 24, 2001 and February 12, 2002, were used. The study site contained 2000×1280 Landsat pixels with an area of 2,304 km^2. The Landsat image on November 24, 2001 (Fig. 4(a)) was used as the t_0 time HR image. The Landsat image on February 12, 2002 was used as the t_p time reference image (Fig. 4(b)). In the simulated image experiment, the Landsat image on February 12, 2002, shown in Fig. 4(b), was degraded spatially to the t_p time MODIS-like image by averaging the spectral reflectance values within each LR pixel with a scale factor s=16 (Fig. 4(c)). In the real image experiment, the MODIS image representing February 13, 2002 was geometrically transformed and co-registered with the corresponding Landsat image with sub-pixel accuracy and was used as the LR at t_p (Fig. 4(d)).

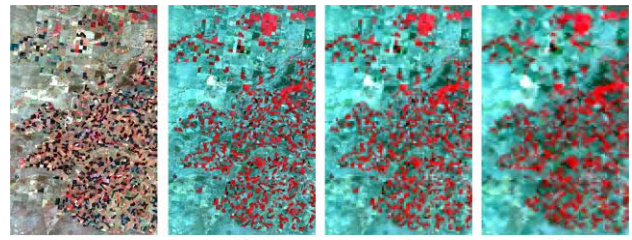


Fig. 4. Experiment imagery used on CIA. (a) Landsat image at t_0 (November 24, 2001), (b) Landsat image at t_p (February 12, 2002), (c) Degraded MODIS-like image at t_p (February 12, 2002), and (d) Real MODIS image representing at t_p (February 13, 2002). The Landsat images contain 2000×1280 pixels. The false color imagery are composited with NIR-red-green as RGB.

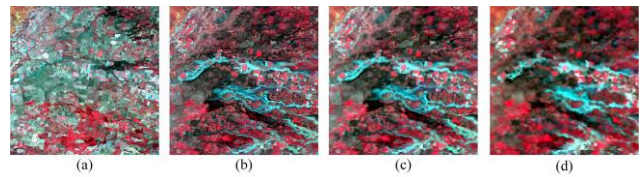


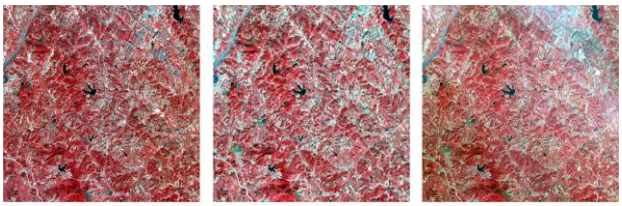
Fig. 5. Experiment imagery used on LGC. (a) Landsat image at t_0 (November 26, 2004), (b) Landsat image at t_p (December 12, 2004), (c) Degraded MODIS-like image at t_p (December 12, 2004), and (d) Real MODI image at t_p (December 12, 2004). The Landsat images contain 2400×2400 pixels. The false color imagery are composited with NIR-red-green as RGB.

At the LGC site, two subsets of cloud-free Landsat images on November 26, 2004, and December 12, 2004, were used. The study site contains 2400×2400 Landsat pixels with an area of 5,184 km^2. The Landsat image on November 26, 2004 (Fig. 5(a)) was used as the t_0 time image. The Landsat image on December 12, 200 was used as the t_p time reference image (Fig. 5(b)). In the simulated image experiment, the Landsat image on December 12, 2004, in Fig. 5(b), was spatially degraded to the t_p time MODIS-like image in Fig. 5(c) with a scale factor s=16. In the real image experiment, the MODIS image representing December 12, 2004, was geometrically transformed and co-registered with the corresponding Landsat image with sub-pixel accuracy, and was used as the LR at t_p (Fig. 5(c)). It is clear that a flood was present on December 12, 2004, and this dataset was used to assess the unmixing-based STIF when dealing with abrupt land cover change associated with a flood event.

2) PlanetScope and Landsat Imagery Experiment

In this experiment, the very-high-resolution PlanetScope image and Landsat 8 image were adopted. The PlanetScope has a 3 m spatial resolution and four spectral bands of blue, green, red, and near-infrared (NIR). The multispectral Landsat 8 image, with seven bands including the coastal aerosol, blue, green, red, NIR, and two shortwave infrared (SWIR) bands,

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2023.3308902



(a) (b) (c)

Fig. 6. PlanetScope and Landsat experiment imagery. (a) PlanetScope image used as the HR image at t_0 (September 19, 2019), (b) Landsat image used as the LR image at t_p (October 20, 2019), and (c) PlanetScope image used as the HR image at t_p (October 21, 2019). The PlanetScope images contain 4000×4000 pixels. The false color imagery are composited with NIR-red-green as RGB.

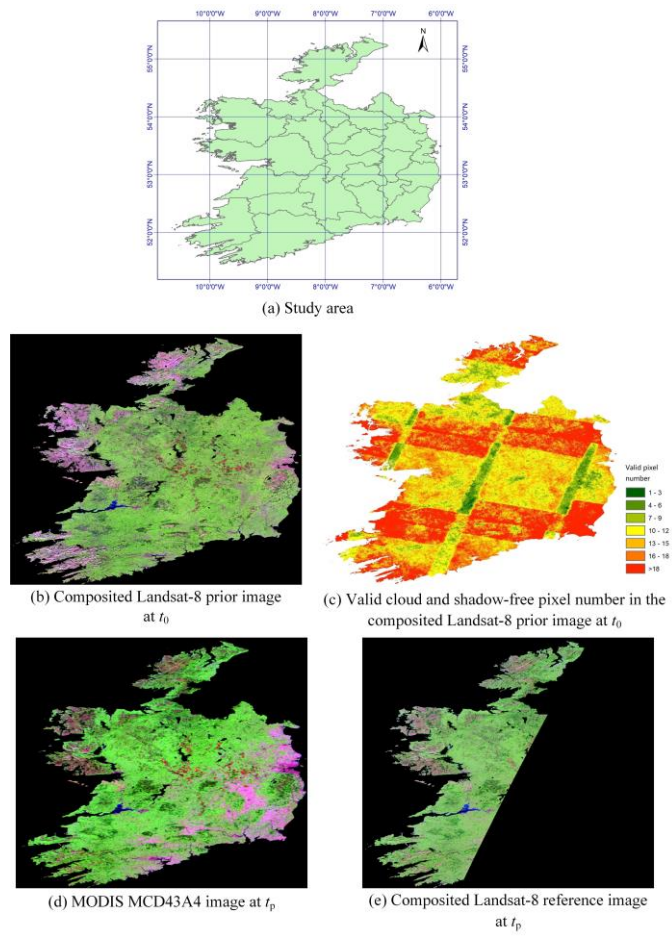
were selected, and the panchromatic, cirrus and two thermal infrared bands were excluded in the experiment. RERC was used to fuse a 3 m resolution image with Landsat spectral bands (seven spectral bands).

The study area (114.45° E, 31.23° N) is located near Wuhan, China with an area of 144 km^2 in Fig. 6. The Landsat subset image acquired on October 20, 2019, was used as the LR image at the prediction time t_p (Fig. 6(b)). The PlanetScope subset image acquired on October 21, 2019, was used as the HR image for validation. The PlanetScope subset image acquired on September 19, 2019, was used as the HR image at the known time t_0. There is no cloud-free Landsat image acquired near the known time t_0; only the unmixing-based STIF can be used in the STIF, whereas the STARFM-like and FSDAF-like methods which require the LR image at the known time as input can not be used. The Landsat image contains 400×400 pixels, and the PlanetScope image contains 4000 × 4000 pixels. The scale factor is set to 10. The PlanetScope images on September 19, 2019 (Fig. 6(a)) and the Landsat image on October 20, 2019 (Fig. 6(b)) were used as the HR image at t_0 and LR image at t_p, respectively. RERC was used to generate a 3 m resolution seven bands image on October 20, 2019.

3) Experiment in Very Large Areas for the Republic of Ireland and France

In this experiment, Landsat and real MODIS imagery for the entire Republic of Ireland were adopted to assess the RERC for national-scale image fusion. The Republic of Ireland is located in the North Atlantic Ocean (Fig. 7(a)). Ireland has a temperate oceanic climate with cloudy and wet weather, and is cloudy and rainy. Cloud- and shadow-free Landsat imagery is rare for this region. STIF was used to generate imagery at the spatial resolution of Landsat with MODIS spectral reflectance.

In this study, the 463 m MODIS MCD43A4 daily surface reflectance image acquired on August 12, 2022, was adopted as the LR image at t_p. The MODIS image was cloud-free, and was re-projected to the WGS 1984 projection and resized to a resolution of 480 m (Fig. 7(d)), and the scale factor s was 16 between the MODIS and Landsat imagery. The Landsat 8 OLI products were used as known and validation time data. The mosaiced Landsat 8 OLI images that were mostly cloud- and shadow-free, acquired on August 12, 2022, were used for validation (Fig. 7(e)). The prevailing climatic conditions often result in variable cloud cover which makes it difficult to form a Landsat image for the entire nation at the same or similar time, the known time HR image used in the unmixing-based STIF is



(a) Study area

(b) Composited Landsat-8 prior image at t_0

(c) Valid cloud and shadow-free pixel number in the composited Landsat-8 prior image at t_0

(d) MODIS MCD43A4 image at t_p

(e) Composited Landsat-8 reference image at t_p

Fig. 7. Test data used for the Republic of Ireland experiment. (a) Location of the study site, (b) Composited and mosaic Landsat 8 OLI image representing the time of t_0 (14912×17344 Landsat pixels). (c) Valid cloud- and shadow-free pixel number in generating the composited and mosaic Landsat-8 image. (d) MODIS MCD43A4 image representing the time of t_p, (e) Composited and mosaic Landsat 8 OLI images representing the time of t_p for validation. Black indicates pixels that do not fall into the area of the Ireland region or no Landsat image that was covered at t_p, or cloud pixels in the Landsat image. The false color images are composited with SWIR2-red-green as RGB.

a synthetic composited and mosaic Landsat 8 OLI image acquired at different times from the Google Earth Engine in Fig. 7(b). The composite and mosaic image was generated from all Landsat 8 surface reflectance images that were atmospherically corrected. The cloud and cloud-shadow pixels in all Landsat 8 images were masked using the quality assessment band in the Landsat 8 level 2 product. The composited Landsat imagery was generated using the median values from all cloud- and shadow-free values for each pixel between January 1, 2021, and August 1, 2022, from a total of 341 Landsat imagery. Using the median value has the benefit of removing clouds (which have a high value) and shadows (which have a low value) that are not masked by the quality assessment band for the Landsat imagery [78]. Almost all Landsat pixels in the composited image are cloud- and shadow-free. All the composited images were then mosaicked in Fig. 7(b). The number of valid cloud- and shadow-free pixels during this period in the corresponding composite Landsat 8 OLI image is shown in Fig. 7(c). Although most areas with yellow and red colors in Fig. 7(c) have more than ten valid cloud- and shadow-free observations from

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2023.3308902

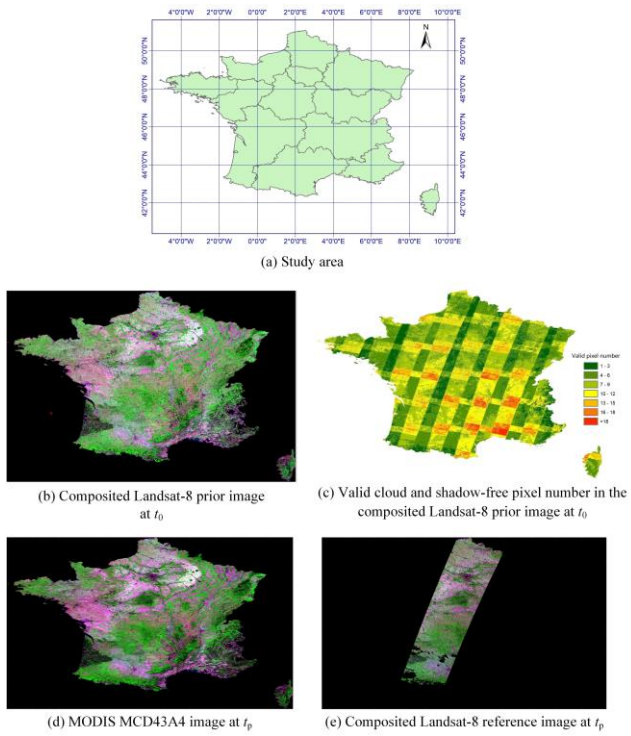


Fig. 8. Test data used in the France experiment. (a) Location of the study site, (b) Composited and mosaic Landsat 8 OLI image representing the time of t_0 (36192×54560 Landsat pixels). (c) Valid cloud- and shadow-free pixel number in generating the composited and mosaic Landsat-8 image. (d) MODIS MCD43A4 image representing the time of t_p, (e) Composited and mosaic Landsat 8 OLI images representing the time of t_p for validation. Black indicates pixels that do not fall into the area of the Ireland region or no Landsat image that was covered at t_p, or cloud pixels in the Landsat image. The false color images are composited with SWIR2-red-green as RGB.

Landsat-8 during this period, about 13.52% of the regions have less than ten cloud- and shadow-free Landsat-8 observations with green color in Fig. 7(c), including 0.25% regions with no more than three cloud- and shadow-free Landsat-8 observations.

Landsat and real MODIS imagery for France were also used to assess the RERC. France is a country that has historically been one of the world's major agricultural centers (Fig. 8(a)). France has various natural land cover types, including forests, croplands, moorlands, and grasses, which may present different reflectance features throughout the year. The STIF, which generates imagery at the spatial resolution of Landsat with MODIS spectral reflectance, is helpful in the understanding of phenology and variations of the land covers, especially agricultural land.

The 463 m MODIS MCD43A4 image acquired on August 13, 2022 (Fig. 8(d)) was reprojected onto the WGS 1984 projection and resized to a resolution of 480 m, and was used as the LR image at the prediction time. The composited and mosaic Landsat 8 OLI images that were mostly cloud- and shadow-free, acquired on August 13, 2022, were used for validation (Fig. 8(e)). Considering the large area of France and the impact of clouds, it is difficult to form a mosaic of Landsat images covering the entire country in one time period. The known time HR image used in the unmixing-based STIF is a synthetic composited and mosaic Landsat 8 OLI image acquired during the period between April 1, 2022, and August 1, 2022, generated from Google Earth Engine using the median values from all cloud- and shadow-free values during this period for each pixel based on a total of 653 Landsat imagery. The number of valid cloud- and shadow-free pixels between April 1, 2022, and August 1, 2022, in the corresponding composited and mosaic Landsat 8 OLI image is shown in Fig. 8(c). Approximately 70.21% of regions have less than ten cloud- and shadow-free Landsat-8 observations during the period with green color in Fig. 8(c), including 9.64% of regions with no more than three cloud- and shadow-free Landsat-8 observations.

B. Model Comparison and Parameter Settings

The proposed RERC was compared to three established unmixing-based STIF methods in the CIA and LGC experiments. Since only the HR at t_0 and LR image at t_p are available, the STARFM-like and FSDAF-like methods which require a LR image at t_0 as input were not compared.

The first comparator method was the unmixing-based data fusion (UBDF) proposed by Zurita-Milla et al. [63]. The UBDF assumes that the HR pixels at t_0 are pure and directly assigns the estimated endmember spectra from the LR image to the corresponding HR image pixel. The second comparator method was the MTF proposed by Amorós-López et al. [68]. The MTF uses a soft clustering algorithm to generate the HR class fractions using the Mahalanobis distance between a HR pixel spectra and the cluster centroid. The third comparator method is the LSUSDFM proposed by Liu et al. [66]. The LSUSDFM uses FCLS to generate HR class fraction images.

The performance of the proposed RERC depends on several parameters. For all methods, the LR sliding window size used for endmember estimation was set to k=11 [68]. The regularization parameter was set to α=0.1 according to the previous studies [66, 68]. In the RERC, the fusion accuracy and computational efficiency are related to the number of clusters n. A larger cluster number is more suitable for dealing with heterogeneous landscapes, but it increases the computation time. The cluster number is usually set to a relatively large value (usually 30) for unmixing-based STIF to reduce the impact of the homogenization effect, which indicates the predicted reflectances are the same for neighboring HR pixels that contain the same class located within the same LR pixel, in the fused image. The optimal number of clusters was set to n=16 for MTF [68]. The optical cluster number in the LSUSDFM is dependent on the landscape complexity of the study site. The cluster number was set to n=10 for both the LSUSDFM and RERC, considering landscape heterogeneity and computational efficiency. In the RERC, the size of the window used for selecting spectral-similar pixels was set to m=16, which equals the scale factor s, and the spectrally similar pixel number was set to L=20 [37, 38]. In the experiment in very large areas for the Republic of Ireland and France, the mosaiced Landsat image was divided into 2400×2400 pixel patches considering the computational efficiency and computer memory, and all the pixels were used for training the regression model.

C. Accuracy Assessment

Many quantitative metrics have been used to assess the

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2023.3308902

different unmixing-based STIFs. The metrics included the root mean square error (RMSE), average absolute deviation (AAD), correlation coefficient (CC), and structure similarity (SSIM) [8, 37, 38]. A smaller RMSE and AAD and larger CC and SSIM indicate a better match between the fused and reference images.

IV. Results

A. Experiment on the CIA and LGC sites

1) Simulated MODIS-like Image Experiment on the CIA site

The study site, which is located in a region of heterogeneous cropland cover, experienced a drastic change in the spectral reflectance values from time t_0 in Fig. 9(a) to time t_p in Fig. 9(b)–(c). The UBDF, which assigns each HR pixel to a unique endmember spectrum, usually homogenizes the spectral reflectance values for spatially adjacent pixels, such as in regions I and II in the zoomed areas (Fig. 9). In contrast, the MTF, LSUSDFM, and RERC generated more variable spectral reflectance values in these regions. The sub-pixel MTF, LSUSDFM, and RERC significantly reduced the homogenization effect to a great extent. In the regions highlighted with circles in the zoomed areas A and B, the RERC in Fig. 9(g) is more similar to the reference images in Fig. 9(c) than UBDF, MTF, and LSUSDFM. For instance, in region V in zoomed area B, the patch has a darker yellow color in the Landsat image at t_0 which is similar to its adjacent patches in Fig. 9(a), which have a relatively darker color than the surrounding patches in the reference image in Fig. 9(c). UBDF, MTF, and LSUSDFM predicted this patch with a light cyan color in region V, which is similar to the surrounding patches. In contrast, the RERC in Fig. 9(g) predicted a darker color for this patch in region V, which is similar to that in the reference image in Fig. 9(c).

The quantitative accuracy metrics obtained for the outputs from the different STIFs at the CIA site are listed in Table II. The pixel-based UBDF method typically generated the highest RMSE and AAD and the lowest CC and SSIM in different spectral bands, whereas the sub-pixel-based methods improved the accuracy of these metrics. The proposed RERC outperformed the comparator STIFs with a decrease in the RMSE and AAD and an increase in the CC and SSIM in all spectral bands.

2) Real MODIS Image Experiment on the CIA site

In zoomed area A in Fig. 10, the region highlighted with the circle I experienced a drastic spectral reflectance change due to land cover changes such as crop rotation and is represented as a white color in the reference Landsat image at t_p. The unmixing-based STIF of UBDF, MTF, and LSUSDFM predicted the region with cyan color highlighted with the circle I, and RERC predicted this region with white color, which is

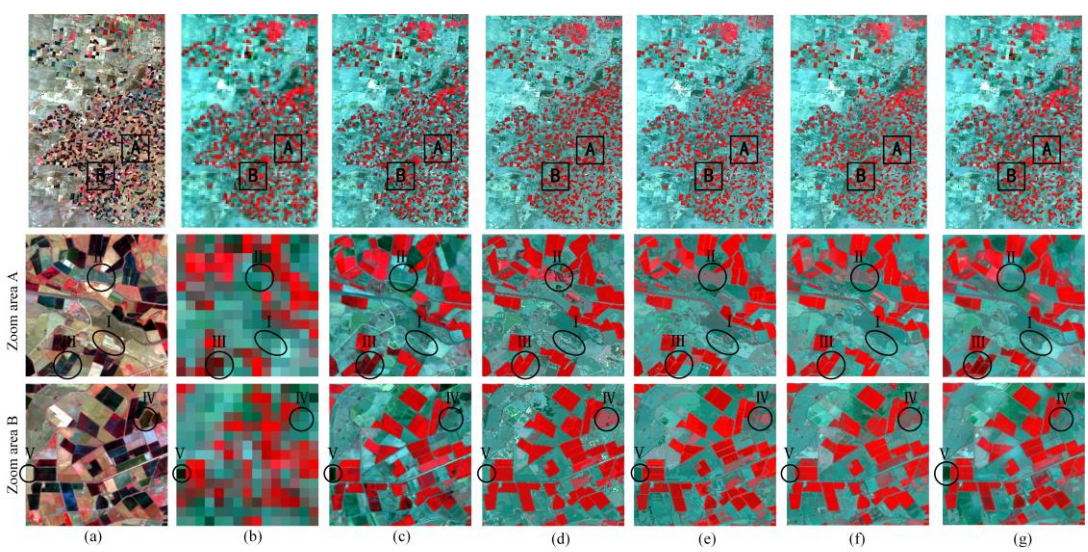


Fig. 9. The input and result images in the CIA simulated image experiment. (a) Landsat image on November 24, 2001, used as the input HR image at t_0, (b) Degraded MODIS-like image on February 12, 2002, used as the LR image at t_p, (c) Landsat image on February 12, 2002, used as the reference, (d) output from UBDF, (e) output from MTF, (f) output from LSUSDFM, and (g) output from RERC.

TABLE II
ACCURACY METRICS IN THE CIA SIMULATED IMAGE EXPERIMENT.

band	RMSE				AAD				CC				SSIM			
	UBDF	MTF	LSUS DFM	RERC	UBDF	MTF	LSUS DFM	RERC	UBDF	MTF	LSUS DFM	RERC	UBDF	MTF	LSUS DFM	RERC
Blue	0.0162	0.0140	0.0136	**0.0104**	0.0116	0.0100	0.0098	**0.0074**	0.7655	0.8154	0.8259	**0.9004**	0.7655	0.8134	0.8225	**0.8995**
Green	0.0219	0.0191	0.0183	**0.0140**	0.0154	0.0133	0.0128	**0.0097**	0.7601	0.8105	0.8235	**0.8999**	0.7600	0.8086	0.8198	**0.8989**
Red	0.0333	0.0292	0.0278	**0.0214**	0.0228	0.0200	0.0192	**0.0145**	0.7911	0.8334	0.8459	**0.9116**	0.7911	0.8327	0.8436	**0.9109**
NIR	0.0609	0.0534	0.0528	**0.0390**	0.0400	0.0353	0.0356	**0.0256**	0.6067	0.6635	0.6654	**0.8298**	0.6065	0.6537	0.6508	**0.8239**
SWIR1	0.0527	0.0471	0.0454	**0.0378**	0.0362	0.0327	0.0317	**0.0254**	0.8454	0.8729	0.8805	**0.9185**	0.8454	0.8725	0.8797	**0.9182**
SWIR2	0.0492	0.0439	0.0424	**0.0348**	0.0337	0.0304	0.0295	**0.0237**	0.8429	0.8710	0.8782	**0.9193**	0.8428	0.8706	0.8770	**0.9189**

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2023.3308902

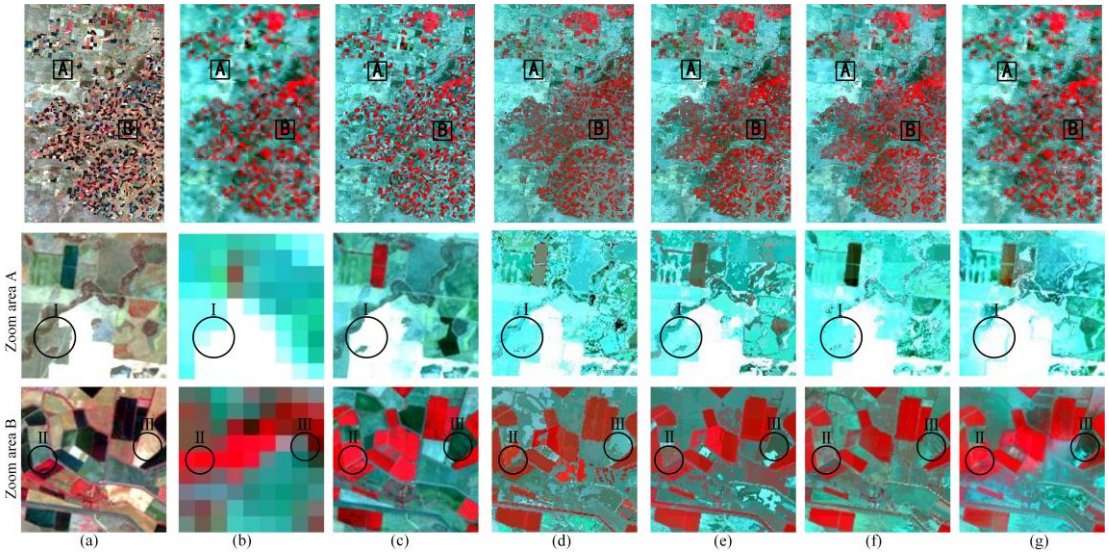


Fig. 10. The input and result images in the CIA real MODIS image experiment. (a) Landsat image on November 24, 2001, used as the HR image at t_0, (b) Real MODIS image on February 13, 2002, used as the LR image at t_p, (c) Landsat image on February 12, 2002, used as the reference, (d) output from UBDF, (e) output from MTF, (f) output from LSUSDFM, and (g) output from RERC.

TABLE III
ACCURACY METRICS IN THE CIA REAL MODIS IMAGE EXPERIMENT.

band	RMSE				AAD				CC				SSIM			
	UBDF	MTF	LSUS DFM	RERC	UBDF	MTF	LSUS DFM	RERC	UBDF	MTF	LSUS DFM	RERC	UBDF	MTF	LSUS DFM	RERC
Blue	0.0175	0.0163	0.0158	**0.0154**	0.0134	0.0126	0.0122	**0.0116**	0.6894	0.7396	0.7624	**0.7745**	0.6557	0.6914	0.7014	**0.7303**
Green	0.0238	0.0220	0.0215	**0.0207**	0.0180	0.0166	0.0163	**0.0155**	0.6801	0.7449	0.7615	**0.7773**	0.6388	0.6835	0.6869	**0.7222**
Red	0.0366	0.0336	0.0328	**0.0317**	0.0276	0.0254	0.0248	**0.0237**	0.7058	0.7615	0.7810	**0.7943**	0.6736	0.7148	0.7204	**0.7487**
NIR	0.0629	0.0594	0.0575	**0.0524**	0.0440	0.0415	0.0404	**0.0372**	0.4593	0.5203	0.5667	**0.6648**	0.4152	0.4315	0.4516	**0.5679**
SWIR1	0.0662	0.0614	0.0598	**0.0587**	0.0510	0.0473	0.0462	**0.0451**	0.7604	0.8036	0.8164	**0.8245**	0.7466	0.7820	0.7890	**0.8019**
SWIR2	0.0590	0.0540	0.0524	**0.0515**	0.0453	0.0415	0.0405	**0.0395**	0.7569	0.8002	0.8125	**0.8198**	0.7475	0.7853	0.7907	**0.8034**

most similar to the reference. In zoomed area B, the subregion highlighted in circle II is represented by a dark red color, and the subregion highlighted in circle III is represented as a dark green color reference Landsat image at t_p. The UBDF and LSUSDFM predicted dissimilar reflectance to the reference image in circles II and III, and the MTF predicted dissimilar reflectance to the reference image in circle III. In contrast, the proposed RERC predicted an image that is the most similar to the reference in circles II and III.

The quantitative accuracy metrics obtained for the outputs from the different STIFs at the CIA site are listed in Table III. The pixel-based UBDF method typically generated the highest RMSE and AAD and the lowest CC and SSIM in different spectral bands, whereas the sub-pixel-based methods improved the accuracy of these metrics. RERC outperformed the comparator STIFs with a decrease in the RMSE and AAD and an increase in the CC and SSIM in all spectral bands.

3) Simulated MODIS-like Image Experiment on the LGC site

The study site experienced a flood event which is observable in the MODIS and reference Landsat images at t_p in Fig. 11. MTF, UBDF, and LSUSDFM predicted only parts of the flood, as shown in zoomed areas A and B, whereas RERC mapped almost all the regions covered by the flood. In particular, in region I in zoomed area A, UBDF, MTF, and LSUSDFM predicted a flood with a discontinuous color, whereas RERC mapped the flood with a more continuous color. In region II in zoomed area A, UBDF, MTF, and LSUSDFM predicted a flood with a disconnected shape, whereas RERC predicted a flood that was spatially connected and continuous. In the zoomed area, B, UBDF, MTF, and LSUSDFM failed to map the flood or only partly mapped the flood in regions III and IV, whereas RERC successfully mapped the flood in both regions. In general, compared with the UBDF and LSUSDFM images, the RERC images in the entire and zoomed areas in Fig. 11(g) are more similar to the reference. Similar to the results for the CIA site, the sub-pixel scale MTF, LSUSDFM, and RERC decreased the RMSE and increased the CC and SSIM in comparison with UBDF. RERC generated the lowest RMSE and AAD and the highest CC and SSIM in Table IV.

4) Real MODIS Image Experiment on the LGC site

The study site experienced a flood event which is observable in the MODIS and reference Landsat images at t_p in Fig. 12. Both zoomed areas A and B experienced floods according to the known (Fig. 12(a)) and prediction time (Fig. 12(c)) Landsat images. In zoomed area A, only the proposed RERC predicted a flood, as highlighted by circle I. In zoomed area B, UBDF failed to predict the flood, as highlighted in circle II, and MTF and LSUSDFM predicted part of the flood, as highlighted in

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2023.3308902

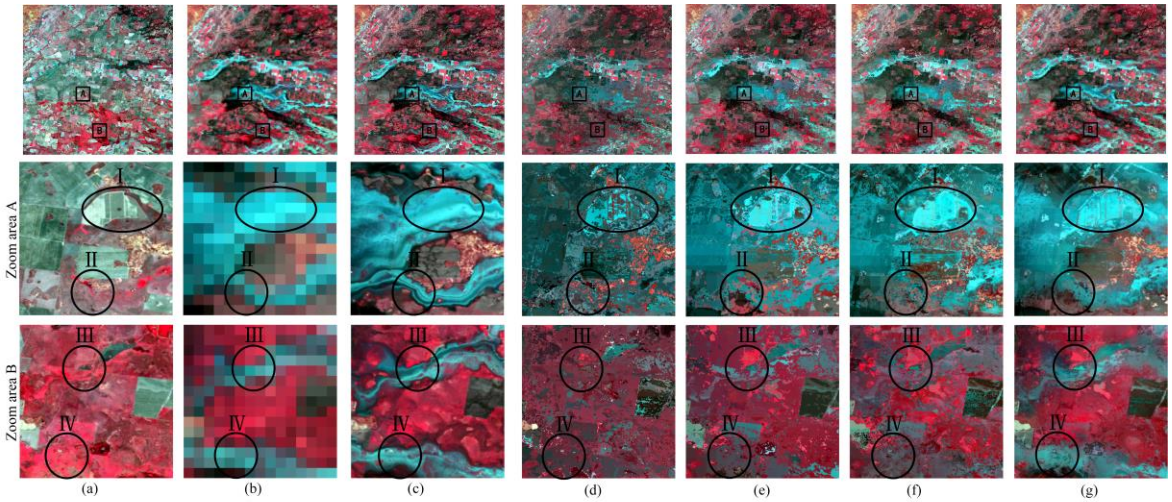


Fig. 11. The input and result images in the LGC simulated image experiment. (a) Landsat image on November 26, 2004, used as the HR image at t_0, (b) Degraded MODIS-like image on December 12, 2004, used as the LR image at t_p, (c) Landsat image on December 12, 2004, used as the reference, (d) output from UBDF, (e) output from MTF, (f) output from LSUSDFM, and (g) output from RERC.

TABLE IV
ACCURACY METRICS IN THE LGC SIMULATED MODIS IMAGE EXPERIMENT.

band	RMSE				AAD				CC				SSIM			
	UBDF	MTF	LSUS DFM	RERC	UBDF	MTF	LSUS DFM	RERC	UBDF	MTF	LSUS DFM	RERC	UBDF	MTF	LSUS DFM	RERC
Blue	0.0189	0.0149	0.0142	**0.0107**	0.0133	0.0105	0.0100	**0.0074**	0.5541	0.6864	0.7197	**0.8504**	0.5540	0.6761	0.7114	**0.8470**
Green	0.0277	0.0217	0.0209	**0.0156**	0.0193	0.0151	0.0145	**0.0107**	0.5394	0.6797	0.7082	**0.8467**	0.5393	0.6671	0.6989	**0.8429**
Red	0.0343	0.0270	0.0258	**0.0194**	0.0236	0.0186	0.0178	**0.0131**	0.5465	0.6816	0.7126	**0.8461**	0.5464	0.6690	0.7027	**0.8419**
NIR	0.0545	0.0464	0.0397	**0.0316**	0.0398	0.0338	0.0289	**0.0224**	0.6475	0.7253	0.7946	**0.8735**	0.6473	0.7240	0.7910	**0.8720**
SWIR1	0.0766	0.0615	0.0593	**0.0454**	0.0571	0.0458	0.0441	**0.0326**	0.5706	0.7014	0.7252	**0.8465**	0.5694	0.6877	0.7125	**0.8383**
SWIR2	0.0545	0.0437	0.0426	**0.0330**	0.0401	0.0322	0.0314	**0.0233**	0.5825	0.7116	0.7286	**0.8456**	0.5809	0.6950	0.7143	**0.8363**

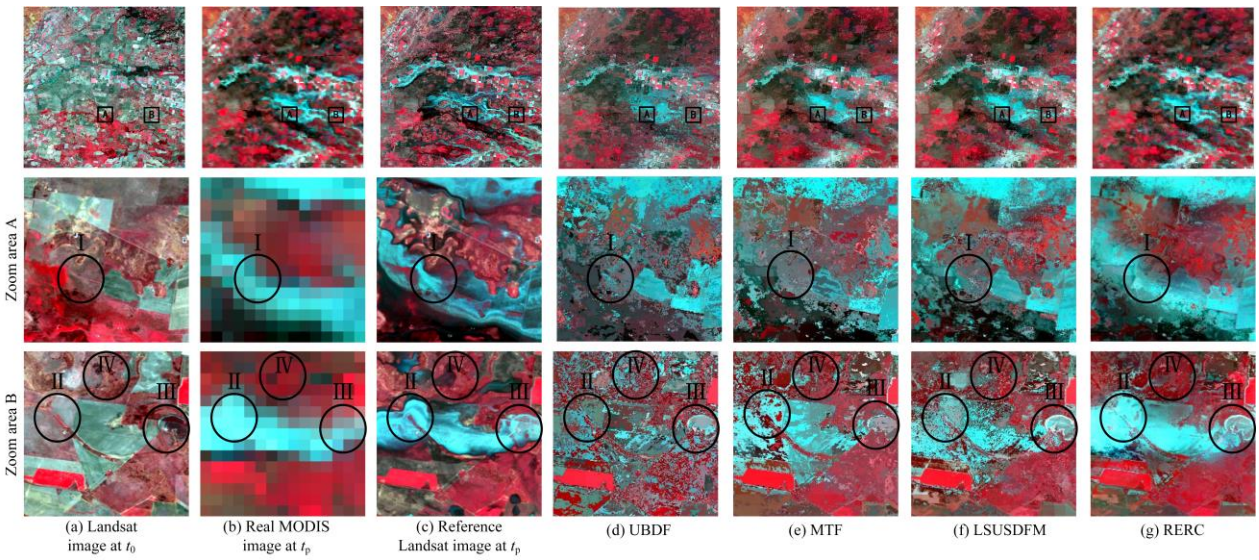


Fig. 12. The input and result images in the LGC real MODIS image experiment. (a) Landsat image on November 26, 2004, used as the HR image at t_0, (b) MODIS image on December 12, 2004, used as the LR image at t_p, (c) Landsat image on December 12, 2004, used as the reference, (d) output from UBDF, (e) output from MTF, (f) output from LSUSDFM, and (g) output from RERC.

circles II and III. In contrast, the RERC predicted a flood that was more similar to that in the reference Landsat image. The region highlighted in Circle IV did not experience floods. The UBDF, MTF, and LSUSDFM incorrectly predicted this region with some floods with light blue color, as shown in Fig. 12, whereas the RERC prediction is more similar to the Landsat image at t_p. This experiment shows that RERC outperformed the comparison method in the prediction of reflectance changes due to both gradual phenological changes and land-cover changes. Similar to the results for the CIA site, the sub-pixel scale MTF, LSUSDFM, and RERC decreased the RMSE and increased the CC and SSIM in comparison with UBDF. RERC

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2023.3308902

TABLE V
ACCURACY METRICS IN THE LGC REAL MODIS IMAGE EXPERIMENT.

band	RMSE				AAD				CC				SSIM			
	UBDF	MTF	LSUS DFM	RERC	UBDF	MTF	LSUS DFM	RERC	UBDF	MTF	LSUS DFM	RERC	UBDF	MTF	LSUS DFM	RERC
Blue	0.0209	0.0180	0.0177	**0.0169**	0.0154	0.0132	0.0130	**0.0125**	0.4794	0.5938	0.6185	**0.6729**	0.4727	0.5728	0.5976	**0.6600**
Green	0.0290	0.0244	0.0239	**0.0225**	0.0207	0.0176	0.0171	**0.0160**	0.4709	0.5883	0.6093	**0.6714**	0.4698	0.5754	0.5980	**0.6669**
Red	0.0358	0.0302	0.0296	**0.0279**	0.0253	0.0216	0.0210	**0.0195**	0.4798	0.5929	0.6130	**0.6744**	0.4785	0.5796	0.6014	**0.6695**
NIR	0.0536	0.0481	0.0444	**0.0420**	0.0402	0.0363	0.0334	**0.0313**	0.6047	0.6791	0.7389	**0.7726**	0.5893	0.6458	0.6974	**0.7407**
SWIR1	0.0736	0.0658	0.0638	**0.0586**	0.0578	0.0518	0.0503	**0.0451**	0.5393	0.6386	0.6652	**0.7296**	0.5113	0.5861	0.6105	**0.6811**
SWIR2	0.0536	0.0485	0.0475	**0.0443**	0.0407	0.0367	0.0359	**0.0328**	0.5424	0.6433	0.6635	**0.7229**	0.4975	0.5674	0.5847	**0.6493**

generated the lowest RMSE and AAD and the highest CC and SSIM in Table V.

B. PlanetScope and Landsat Image Experiment

The input and RERC result images are shown in Fig. 13. The UBDF, MTF, and LSUSDFM were not compared because they are computationally inefficient. In particular, the running time of the random forest regression used in RERC is only ~1% of the running time of the FCLS linear mixture model (the running time is 506 seconds for random forest regression, longer than 7,000 seconds for the non-linear model, and longer than 50,000 seconds for FCLS on the CIA site). In Fig. 13 (c), the Landsat images were relatively coarse to represent the detailed spatial distribution of land covers. The boundaries are jagged and the settlements were blurred in the Landsat image at t_p such as highlighted in circle I in zoom area A, and are clear in the RERC prediction image in Fig. 13(d). In zoom area B, the bare land region highlighted with circle II had a surface reflectance change from t_0 to t_p, and RERC had predicted this reflectance change in Fig. 13(d). In zoom area B, the region highlighted with circle III had experienced a drastic surface reflectance change from t_0 to t_p, and the RERC prediction in Fig. 13(d) is similar to the reference PlanetScope image at t_p in Fig. 13(b).

The Landsat image contains more spectral bands, such as the coastal aerosol band and the SWIR bands, than the four bands PlanetScpoe image (NIR, red, green, and blue bands). In this study, RERC predicted the image with Landsat spectral information including the coastal aerosol and the SWIR bands and at the PlanetScpoe spatial resolution in Fig. 13(e). Since RERC can fuse imagery with different spectral bands in Fig. 13(f), it is more flexible than the STARFM-like and FSDAF-like STIFs which fuse imagery with the same spectral bands.

Table VI shows the quantitative metrics for RERC. RERC generated RMSE and AD of approximately lower than 0.040 and 0.032, respectively, in all four spectral bands. The quantitative metrics were not assessed in the coastal aerosol and SWIR bands which the PlanetScope does not contain. The RMSE and AAD values for RERC usually increase with an increase in the standard deviation in the spectral band of the reference Landsat image at t_p. In other words, if the standard deviation in the reflectance in a spectral band is high, the reflectance is more variable in this band, and it is more difficult for the RERC to accurately predict the reflectance. In particular, RERC generated the highest RMSE and AAD values in the NIR band, which had the highest standard deviation in reflectance in the reference Landsat image at t_p, and generated the lowest RMSE and AAD in the blue band, which has the lowest standard deviation in reflectance in the reference Landsat image at t_p. It is also evident that there is a

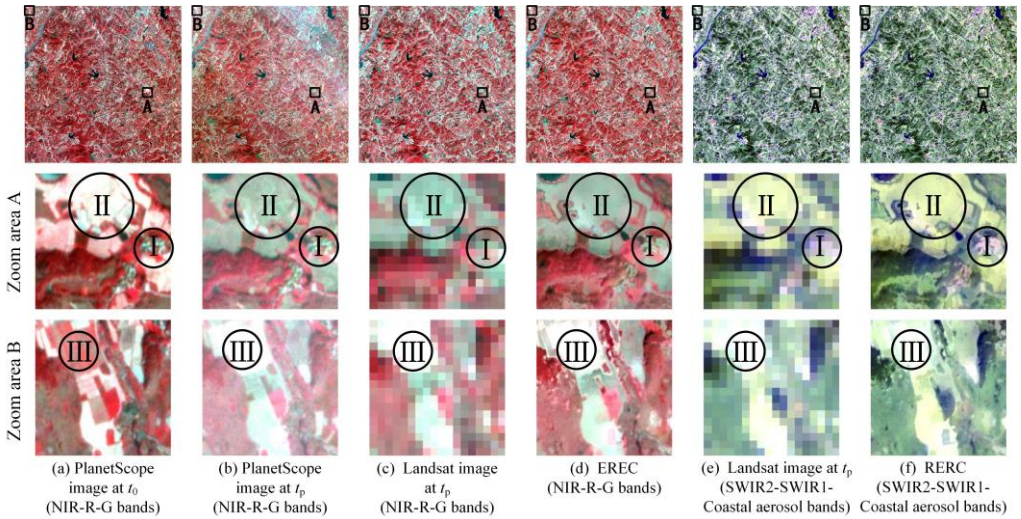


Fig. 13. The input, reference, and result images in the PlanetScpoe and Landsat image experiment. (a) PlanetScope image on September 19, 2019, used as the HR image at t_0, (b) Landsat image on October 20, 2019, used as the LR image at t_p, (c) PlanetScope image on October 21, 2019, used as the reference, (d) output from RERC, (e) PlanetScope image on October 21, 2019, used as the reference, (f) output from RERC. The NIR-red-green bands are composited as RGB in (a)-(d), and the SWIR2-SWIR1-coastal aerosol bands are composited as RGB in (e) and (f).

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2023.3308902

TABLE VI
ACCURACY METRICS OF THE PROPOSED RERC FOR THE PLANETSCOPE AND LANDSAT EXPERIMENT.

band	Accuracy metrics for RERC				Mean (±Standard deviation) in the reflectance		
	RMSE	AAD	CC	SSIM	Reference PlanetScope at t_p	Landsat at t_p	RERC prediction at t_p
Coastal aerosol	No PlanetScope band available					0.0415 (±0.0127)	0.0416 (±0.0136)
Blue	0.0239	0.0200	0.6475	0.6034	0.0651 (±0.0174)	0.0457 (±0.0141)	0.0457 (±0.0152)
Green	0.0307	0.0268	0.7296	0.6865	0.0937 (±0.0219)	0.0627 (±0.0181)	0.0672 (±0.0195)
Red	0.0329	0.0277	0.7901	0.7493	0.0973 (±0.0293)	0.0702 (±0.0263)	0.0702 (±0.0283)
NIR	0.0403	0.0323	0.7675	0.7598	0.2508 (±0.0370)	0.2201 (±0.0370)	0.2201 (±0.0392)
SWIR1	No PlanetScope band available					0.1963 (±0.0580)	0.1965 (±0.0617)
SWIR2	No PlanetScope band available					0.1227 (±0.0496)	0.1228 (±0.0526)

band difference (including the central wavelength and bandwidth) between the reference Planetscpoe image and the input Landsat image at t_p, and the spectral band difference would also impact the quantitative analysis of the fusion result. The RERC generates similar mean values in each spectral band with the Landsat image at t_p, showing that the RERC can fuse HR image at the spatial resolution of PlanetScope while maintaining the spectral information from the Landsat image. The RERC prediction image enlarged the standard deviation values in each spectral band compared with the Landsat image at t_p, showing that RERC can enhance the detail and variation in the pixel spectral reflectance values compared with the Landsat image.

C. Experiment in Very Large Areas for the Republic of Ireland and France

1) Experiment in the Republic of Ireland

The MODIS and RERC prediction images for the Republic of Ireland site are shown in Fig. 14. The UBDF, MTF, and LSUSDFM which are computationally inefficient were not compared. The RERC prediction image and the corresponding zoomed-area image are shown in Fig. 14. The zoomed area highlighted with the ellipse in the known Landsat image is represented in pink in Fig. 14(c), and this region is represented in green in the reference Landsat image in Fig. 14(e) because the vegetation is in the growing season, showing a reflectance change due to phenological factors. This region, highlighted with the ellipse, is represented as green in the zoomed area in the MODIS image in Fig. 14(d); however, the spatial details are blurred in the MODIS image. In the RERC prediction image, the zoomed area highlighted with the ellipse is represented in green in Fig. 14(b), which is similar to that in the reference Landsat image in Fig. 14(e), showing the ability to map surface reflectance changes due to phenological changes. Compared with the MODIS image, the fused RERC represents more of the spatial details of land cover, such as the vegetation highlighted with the ellipse and the lakes highlighted with the circle in Fig. 14(b), showing the necessity of the fusion of MODIS and Landsat in monitoring land cover.

The quantitative metrics for RERC are listed in Table VII. In the blue, green, red, and SWIR bands, the RMSE and AAD values range from 0.0090 to 0.0273 and from 0.0062 to 0.0195, respectively. The RMSE and AAD values are relatively larger

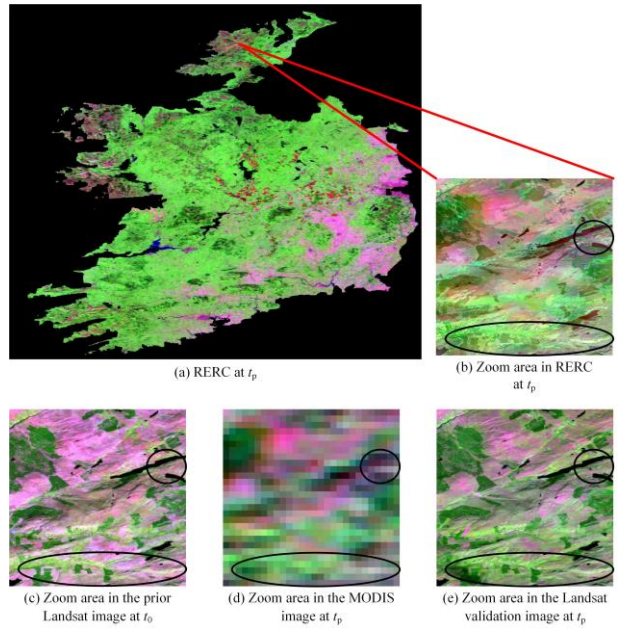


Fig. 14. The RERC prediction image (14912×17344 Landsat pixels) and the zoomed areas (448×448 Landsat pixels) in the Republic of Ireland experiment. (a) The entire RERC fused image, (b) the zoomed RERC fused image, (c) the zoomed Landsat image at t_0, (d) the zoomed MODIS image at t_p, and (e) the zoomed Landsat image at t_p used for validation.

TABLE VII
ACCURACY METRICS OF THE PROPOSED RERC FOR THE REPUBLIC OF IRELAND. UBDF, MTF, AND LSUSDFM WERE NOT COMPARED BECAUSE THEY ARE RELATIVELY COMPUTATIONALLY INEFFICIENT WHEN THE STUDY SITE IS VERY LARGE.

band	Accuracy metrics for RERC				Mean (±Standard deviation) in the reflectance		
	RMSE	AAD	CC	SSIM	Reference Landsat at t_p	MODIS at t_p	RERC prediction at t_p
Blue	0.0090	0.0062	0.5666	0.5307	0.0257 (±0.0104)	0.0233 (±0.0066)	0.0233 (±0.0073)
Green	0.0135	0.0106	0.7138	0.7021	0.0538 (±0.0157)	0.0612 (±0.0124)	0.0611 (±0.0138)
Red	0.0161	0.0115	0.5814	0.5235	0.0421 (±0.0193)	0.0454 (±0.0104)	0.0454 (±0.0122)
NIR	0.0583	0.0444	0.8243	0.8024	0.3594 (±0.0977)	0.3777 (±0.0719)	0.3776 (±0.0778)
SWIR2	0.0273	0.0195	0.6129	0.5499	0.0891 (±0.0345)	0.0876 (±0.0197)	0.0876 (±0.0215)

in the NIR band than those in other bands. The main reason is that the reflectance values have the largest variance in the NIR band with a standard deviation of 0.0977. RERC predicts a CC value of 0.8243 and SSIM value of 0.8024 in the NIR band, showing that the RERC prediction image has a high correlation with the reference Landsat image for validation.

2) Experiment in France

The results of the RERC and zoomed areas in the input, result, and reference imagery are shown in Fig. 15. It is evident that the zoomed area has experienced drastic surface reflectance changes by comparing the known time image (Fig. 15(c)) with the reference image (Fig. 15(e)), as highlighted by the circles. The drastic surface reflectance changes are obvious in the MODIS image, as highlighted by the circles in Fig. 15 (d).

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2023.3308902

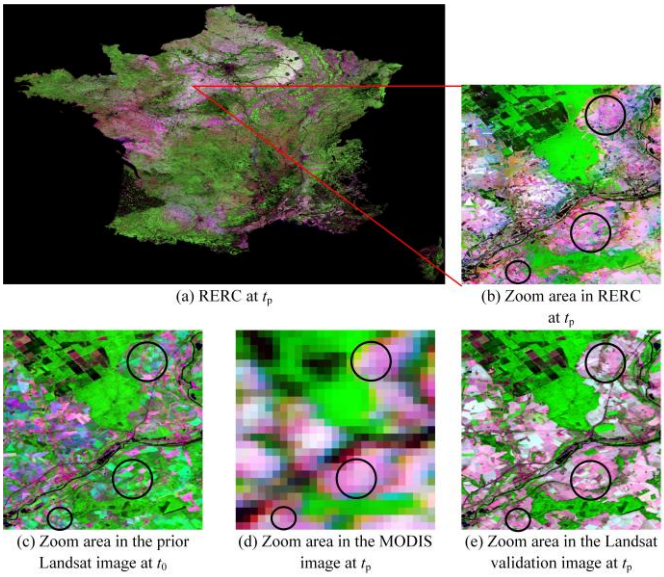


Fig. 15. The RERC prediction image (36192×54560 Landsat pixels) and the zoomed areas (448×448 Landsat pixels) for the France experiment. (a) The entire RERC fused image, (b) the zoomed RERC fused image, (c) the zoomed Landsat image at t_0, (d) the zoomed MODIS image at t_p, and (e) the zoomed Landsat image at t_p used for validation.

TABLE VIII
ACCURACY METRICS OF THE PROPOSED RERC FOR FRANCE. UBDF, MTF, AND LSUSDFM WERE NOT COMPARED BECAUSE THEY ARE COMPUTATIONALLY INEFFICIENCY WHEN THE STUDY SITE IS VERY LARGE.

band	Accuracy metrics for RERC				Mean (±Standard deviation) in the reflectance		
	RMSE	AAD	CC	SSIM	Reference Landsat at t_p	MODIS at t_p	RERC prediction at t_p
Blue	0.0213	0.0156	0.7439	0.7035	0.0559 (±0.0309)	0.0506 (±0.0199)	0.0506 (±0.0224)
Green	0.0280	0.0207	0.7587	0.7385	0.0928 (±0.0429)	0.0915 (±0.0302)	0.0915 (±0.0340)
Red	0.0438	0.0324	0.7543	0.7294	0.1092 (±0.0665)	0.1053 (±0.0466)	0.1053 (±0.0514)
NIR	0.0459	0.0338	0.6346	0.5911	0.3220 (±0.0592)	0.3233 (±0.0345)	0.3234 (±0.0404)
SWIR2	0.0579	0.0435	0.7279	0.7069	0.1646 (±0.0839)	0.1683 (±0.0594)	0.1683 (±0.0659)

However, the MODIS image is blurred and the relatively coarse resolution fails to represent the spatial detail of the reflectance changes. In comparison with the MODIS image, the RERC image is generated at 30 m resolution and better demonstrates the spatial detail of the reflectance at the prediction time such as highlighted in the circles in Fig. 15 (b). The quantitative metrics of the RERC for the experiment focused on France are listed in Table VIII. Similar to the Republic of Ireland experiment, RERC predicted the highest RMSE and AAD values in the SWIR2 band, which had the largest standard deviation value in reflectance, and predicted the lowest RMSE and AAD values in the blue band, which had the smallest standard deviation value in reflectance. The mean values generated from the RERC were very similar to those in the MODIS image at tp in each spectral band. The standard deviations in the MODIS image at t_p are smaller than those in the reference Landsat image at t_p, which are enlarged by the RERC, showing the ability of the RERC to enhance pixel spectral reflectance variance via fusion.

V. Discussion

The comparison of different unmixing-based algorithms in generating the HR class fraction images, the performance of different unmixing-based STIFs in dealing with blocky effects in the output, and the limitations and future work for RERC are discussed here.

A. Comparison of Different Unmixing Algorithms used in Generating the HR Class Fraction Images

RERC and the two comparator methods of MTF and LSUSDFM used different unmixing algorithms to generate HR class fraction images. The random forest regression used in RERC has several advantages compared with the soft-clustering algorithm used in MTF and the FCLS unmixing used in LSUSDFM.

First, in MTF, the soft-clustering algorithm used is dependent on the softness parameter [68]. A very small softness parameter results in class fractions from different classes close to 0 or 1, and the result is similar to a hard classification map. A very large softness parameter would result in class fractions from different classes similar to $1/n$ which is rare in real scenarios [72]. Therefore, the optimal softness parameter value is dependent on the heterogeneity of land cover and the corresponding existence of mixed pixels in the HR image. In comparison with the soft-clustering algorithm, the random forest regression used in RERC is more flexible and robust in use.

Second, in the LSUSDFM, the FCLS spectral unmixing is ill-posed if the number of endmembers is larger than the number of HR image bands, and the ill-posed problem is more severe if the HR image (such as Landsat, PlanetScope, and Chinese GF) used in the unmixing has very few spectral bands. In contrast, the random forest regression can generate accurate fractions of multiple classes when the image has a limited number of spectral bands [79, 80], and has been used to generate class fraction images for the PlanetScope image which has only four spectral bands. In addition, the FCLS is an inversion problem and is especially time-consuming when the spectral bands and the number of endmembers are large, which greatly limits its application to large areas. In contrast, the self-trained regression is not based on the inversion approach and the regression model is more computationally efficient; its unmixing running time is only ~1% of the running time of the FCLS linear mixture model. Third, the FCLS linear mixture model is dependent on the endmembers, and may fail to consider the intra-class spectral variability information in the unmixing, whereas self-trained regression does not require information from endmembers. Lastly, the FCLS linear mixture model is not suitable for non-linear mixture models, whereas the self-trained regression is more flexible.

B. Performance of Different STIFs in Dealing with Blocky Effects

The blocky effect indicates the discontinuity in reflectance between HR pixels of the same class crossing the boundaries of two neighboring LR pixels. The blocky effect results from the fact that the endmembers in the two neighboring LR pixels are estimated using different sets of LR pixels, and the same class in the two neighboring LR pixels may be assigned different

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2023.3308902

reflectance. The blocky effect arises because each LR pixel is unmixed independently in the estimation of the local endmember. Fig. 16 compares the impact of the blocky effect highlighted with circles in the figures for different unmixing-based fusion methods on the CIA and LGC sites using real MODIS imagery. The blocky effect is the most obvious in the pixel-scale UBDF result and the MTF result because MTF with a relatively small softness parameter of 1.1 would generate class fractions close to 0 or 1. RERC generated the minimal blocks. The advantage of reducing the blocky effect accounts for two reasons for RERC. First, for the pixel-based STIF, which directly assigns the estimated endmember spectra to the corresponding HR pixel, the resultant reflectances for the same class pixels located in the same LR pixel are homogenized, and the same class pixels located at neighboring LR pixels would result in a blocky effect in reflectance if the estimated endmembers for neighboring LR pixels are different. In contrast, the sub-pixel-based STIFs linearly combine the local endmembers with the HR class fractions to generate the fused image, and the resultant reflectance is dependent on both the endmembers and HR class fractions. The effectiveness of this strategy, that is, the sub-pixel strategy, is also demonstrated in the MTF and LSUSDFM results in comparison with the UBDF results in Fig. 16. Second, RERC selects similar pixels to post-process and averages the pixel spectral values in the predicted image. Since similar pixels may be selected for HR pixels located in different LR pixels, the blocky effect that occurs for similar pixels located at neighboring LR pixels could be reduced.

C. Limitations and Future Works

RERC spatially interpolates the LR image at t_p to a HR scale based on bicubic interpolation, and then downscales the LR image at t_p to a HR scale based on the spectrally similar pixels in the HR image at t_0. However, if a land-cover change occurs, the HR land-cover spatial pattern changes accordingly, and the HR image at t_0 may not represent the real land-cover spatial pattern at t_p. Results show that, like other STIFs, the proposed RERC better mapped the reflectance change in homogeneous regions but may fail to predict the texture in land cover changed areas. This is because RERC uses bilinear interpolation in the residual compensation approach in which the spatial details are not reconstructed. Future studies can focus on the use of deep learning to map regions where land cover has changed [81].

Although RERC could reduce the blocky effect to some extent, it does not involve incorporating new constraints in unmixing. The blocky effect is mainly because different LR pixels are involved in unmixing spatially adjacent pixels, resulting in dissimilar spectral values even for the same class located in the spatially adjacent LR pixels. Thus, an effective and direct way to reduce the blocky effect for the unmixing-based STIF is by minimizing the reflectance difference of the same class in spatially adjacent LR pixels. Wang et al. [55] proposed a novel block-removal method that minimizes the residual error to ensure the spatial continuity of the endmember reflectance in unmixing, which is an effective and general solution for UBDF and other STIFs. This constrained-function strategy is applied directly to the same class pixels at the pixel scale, and its application in sub-pixel scale block removal should be explored in future studies. Thirdly, this paper fused images using only one HR image at a known time. If both HR images that pre-date and post-date the prediction time are available, it is suggested to fuse the prediction time image separately using different HR images, and then combine the fused image to generate the final predicted image to further improve the fusion accuracy.

In addition, although many machine learning models such as the artificial neural network and support vector machine regression can be used in the training and predicting of the sub-pixel class fractions, the random forest has been adopted in RERC for its simplicity and high precision [82, 83]. Future works can explore various simple machine-learning regressions and deep learning regression in exploring sub-pixel class fraction information from the remote sensing imagery in the unmixing-based STIF.

Lastly, the unmixing-based STIF requires less input than the state-of-the-art STARFM-like and FSDAF-like fusions and is thus more flexible in use, especially in image fusion in very large areas. For instance, the unmixing-based STIF can be used to generate high spatiotemporal imagery based on the LR imagery such as MODIS and Sentinel-3 which have a large span and the mosaicked medium resolution image such as Landsat and Sentinel-2 acquired at different dates to generate Landsat-like or Sentinel-2-like imagery with very high temporal repetition rates. Future studies could focus on this while continuously increasing the efficiency of the

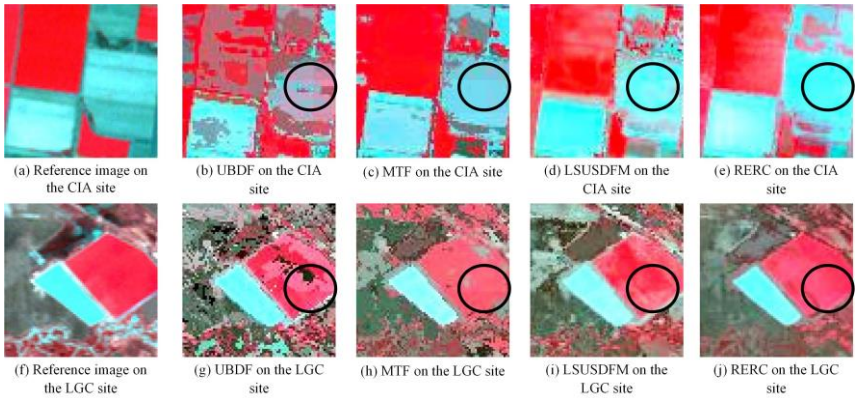


Fig. 16. Comparison of the impact of the blocky effect on different unmixing based fusion methods in the CIA and LGC experiments using real MODIS imagery. (a) reference image on the CIA site. (b) UBDF on the CIA site, (c) MTF on the CIA site, (d) LSUSDFM on the CIA site, (e) RERC on the CIA site, (f) reference image on the LGC site. (g) UBDF on the LGC site, (h) MTF on the LGC site, (i) LSUSDFM on the LGC site, and (j) RERC on the LGC site.

unmixing-based STIF.

VI. Conclusion

This study proposes a new unmixing-based STIF of RERC based on a self-trained machine-learning regression, LR endmember estimation, HR image reconstruction, and residual compensation. The self-trained regression trains the relationship between the reflectance image and the corresponding class fractions at a coarse resolution scale, and then uses this relationship in unmixing the HR class fraction images. In comparison with the FCLS linear spectral unmixing and the soft-clustering, the self-trained regression does not require any information about endmembers, and is flexible in use. In addition, the self-trained regression does not have physical constraints like the FCLS linear spectral unmixing which requires the number of the number endmembers to be no more than the number of spectral bands to generate reliable results, and does not require the information about endmember distribution that is used in the soft-clustering. Lastly, self-trained regression is computationally efficient, and its computation time is approximately 1% and 10% of that for the FCLS linear spectral unmixing and the soft-clustering, respectively. The proposed RERC incorporates the LR image at the prediction time and better predicts the reflectance in regions that experienced drastic reflectance changes than the comparator of unmixing-based STIFs, owing to the residual compensation term to make the full use of the LR image at the prediction time. The experimental results also show that RERC not only reduced the homogenization effect compared with UBDF, but also reduced the blocky effect to a great extent. RERC has been applied to fuse a 3 m PlanetScope image of four bands image at the known time with a Landsat image of seven bands at the prediction time to generate a 3 m seven bands multispectral image and is more flexible than the STARFM-like and FSDAF-like STIFs which require additional LR image at the known time and requires the LR and HR to have similar spectral bands. RERC has been applied to fuse 30 m imagery with MODIS spectral reflectance at the national scale for the Republic of Ireland (~70,273 km^2) and France (~551,500 km^2), showing its potential for mapping daily multispectral imagery at Landsat spatial resolution for large-area land surface monitoring.

References

[1] F. Gao, J. Masek, M. Schwaller, and F. Hall, "On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance," *IEEE T. Geosci. Remote.*, vol. 44, no. 8, pp. 2207-2218, 2006.

[2] T. Hilker *et al.*, "A new data fusion model for high spatial-and temporal-resolution mapping of forest disturbance based on Landsat and MODIS," *Remote Sens. Environ.*, vol. 113, no. 8, pp. 1613-1627, 2009.

[3] T. Hilker *et al.*, "Generation of dense time series synthetic Landsat data through data blending with MODIS using a spatial and temporal adaptive reflectance fusion model," *Remote Sens. Environ.*, vol. 113, no. 9, pp. 1988-1999, 2009.

[4] P. H. Verburg, K. Neumann, and L. Nol, "Challenges in using land use and land cover data for global change studies," *Global change biology*, vol. 17, no. 2, pp. 974-989, 2011.

[5] J. A. Foley *et al.*, "Global consequences of land use," *Science*, vol. 309, no. 5734, pp. 570-574, 2005.

[6] R. Zurita-Milla, G. Kaiser, J. Clevers, W. Schneider, and M. E. Schaepman, "Downscaling time series of MERIS full resolution data to monitor vegetation seasonal dynamics," *Remote Sens. Environ.*, vol. 113, no. 9, pp. 1874-1885, 2009.

[7] M. A. Wulder *et al.*, "Current status of Landsat program, science, and applications," *Remote Sens. Environ.*, vol. 225, pp. 127-147, 2019.

[8] X. Zhu, J. Chen, F. Gao, X. Chen, and J. G. Masek, "An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions," *Remote Sens. Environ.*, vol. 114, no. 11, pp. 2610-2623, 2010.

[9] M. Wu, Z. Niu, C. Wang, C. Wu, and L. Wang, "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.*, vol. 6, no. 1, pp. 063507-063507, 2012.

[10] M. Drusch *et al.*, "Sentinel-2: ESA's optical high-resolution mission for GMES operational services," *Remote Sens. Environ.*, vol. 120, pp. 25-36, 2012.

[11] X. Li, F. Ling, G. M. Foody, Y. Ge, Y. Zhang, and Y. Du, "Generating a series of fine spatial and temporal resolution land cover maps by fusing coarse spatial resolution remotely sensed images and fine spatial resolution land cover maps," *Remote Sens. Environ.*, vol. 196, pp. 293-311, 2017.

[12] X. Zhu, F. Cai, J. Tian, and T. K.-A. Williams, "Spatiotemporal fusion of multisource remote sensing data: Literature survey, taxonomy, principles, applications, and future directions," *Remote Sens-Basel.*, vol. 10, no. 4, p. 527, 2018.

[13] C. Zhang, Q. Wang, H. Xie, Y. Ge, and P. M. Atkinson, "Spatio-temporal subpixel mapping with cloudy images," *Sci. Remote Sens.*, vol. 6, p. 100068, 2022.

[14] H. Shu *et al.*, "Fusing or filling: Which strategy can better reconstruct high-quality fine-resolution satellite time series?," *Sci. Remote Sens.*, vol. 5, p. 100046, 2022.

[15] J. Li, Y. Li, L. He, J. Chen, and A. Plaza, "Spatio-temporal fusion for remote sensing data: An overview and new benchmark," *Science China Information Sciences*, vol. 63, pp. 1-17, 2020.

[16] M. Belgiu and A. Stein, "Spatiotemporal image fusion in remote sensing," *Remote Sens-Basel.*, vol. 11, no. 7, p. 818, 2019.

[17] P. Ghamisi *et al.*, "Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 1, pp. 6-39, 2019.

[18] X. Zhu *et al.*, "A novel framework to assess all-round performances of spatiotemporal fusion models," *Remote Sens. Environ.*, vol. 274, p. 113002, 2022.

[19] Z. Wang, Y. Ma, and Y. Zhang, "Review of pixel-level remote sensing image fusion based on deep learning," *Information Fusion*, 2022.

[20] Y. Luo, K. Guan, and J. Peng, "STAIR: A generic and fully-automated method to fuse multiple sources of optical satellite data to generate a high-resolution, daily and cloud-/gap-free surface reflectance product," *Remote Sens. Environ.*, vol. 214, pp. 87-99, 2018.

[21] Q. Wang and P. M. Atkinson, "Spatio-temporal fusion for daily Sentinel-2 images," *Remote Sens. Environ.*, vol. 204, pp. 31-42, 2018.

[22] J. Walker, K. De Beurs, R. Wynne, and F. Gao, "Evaluation of Landsat and MODIS data fusion products for analysis of dryland forest phenology," *Remote Sens. Environ.*, vol. 117, pp. 381-393, 2012.

[23] I. V. Emelyanova, T. R. McVicar, T. G. Van Niel, L. T. Li, and A. I. Van Dijk, "Assessing the accuracy of blending Landsat–MODIS surface reflectances in two landscapes with contrasting spatial and temporal dynamics: A framework for algorithm selection," *Remote Sens. Environ.*, vol. 133, pp. 193-209, 2013.

[24] H. Shen, X. Meng, and L. Zhang, "An integrated framework for the spatio–temporal–spectral fusion of remote sensing images," *IEEE T. Geosci. Remote.*, vol. 54, no. 12, pp. 7135-7148, 2016.

[25] A. Li, Y. Bo, Y. Zhu, P. Guo, J. Bi, and Y. He, "Blending multi-resolution satellite sea surface temperature (SST) products using Bayesian maximum entropy method," *Remote Sens. Environ.*, vol. 135, pp. 52-63, 2013.

[26] B. Huang and H. Song, "Spatiotemporal reflectance fusion via sparse representation," *IEEE T. Geosci. Remote.*, vol. 50, no. 10, pp. 3707-3716, 2012.

[27] C. Shang *et al.*, "Spatiotemporal reflectance fusion using a generative adversarial network," *IEEE T. Geosci. Remote.*, vol. 60, pp. 1-15, 2021.

[28] L. Wang, X. Wang, and Q. Wang, "Using 250-m MODIS data for enhancing spatiotemporal fusion by sparse representation," *Photogrammetric Engineering & Remote Sensing*, vol. 86, no. 6, pp. 383-392, 2020.

[29] X. Liu, C. Deng, J. Chanussot, D. Hong, and B. Zhao, "StfNet: A two-stream convolutional neural network for spatiotemporal image fusion," *IEEE T. Geosci. Remote.*, vol. 57, no. 9, pp. 6552-6564, 2019.

[30] H. Zhang, Y. Song, C. Han, and L. Zhang, "Remote sensing image spatiotemporal fusion using a generative adversarial network," *IEEE T. Geosci. Remote.*, vol. 59, no. 5, pp. 4273-4286, 2020.
[31] Z. Tan, M. Gao, X. Li, and L. Jiang, "A flexible reference-insensitive spatiotemporal fusion model for remote sensing images using conditional generative adversarial network," *IEEE T. Geosci. Remote.*, vol. 60, pp. 1-13, 2021.
[32] W. Li, X. Zhang, Y. Peng, and M. Dong, "Spatiotemporal fusion of remote sensing images using a convolutional neural network with attention and multiscale mechanisms," *Int. J. Remote. Sens.*, vol. 42, no. 6, pp. 1973-1993, 2021.
[33] Y. Li, J. Li, L. He, J. Chen, and A. Plaza, "A new sensor bias-driven spatio-temporal fusion model based on convolutional neural networks," *Science China Information Sciences*, vol. 63, no. 4, pp. 1-16, 2020.
[34] J. Chen, L. Wang, R. Feng, P. Liu, W. Han, and X. Chen, "CycleGAN-STF: Spatiotemporal fusion via CycleGAN-based image generation," *IEEE T. Geosci. Remote.*, vol. 59, no. 7, pp. 5851-5865, 2020.
[35] Y. Chen, K. Shi, Y. Ge, and Y. n. Zhou, "Spatiotemporal remote sensing image fusion using multiscale two-stream convolutional neural networks," *IEEE T. Geosci. Remote.*, vol. 60, pp. 1-12, 2021.
[36] D. Guo, W. Shi, M. Hao, and X. Zhu, "FSDAF 2.0: Improving the performance of retrieving land cover changes and preserving spatial details," *Remote Sens. Environ.*, vol. 248, p. 111973, 2020.
[37] X. Li *et al.*, "SFSDAF: An enhanced FSDAF that incorporates sub-pixel class fraction change information for spatio-temporal image fusion," *Remote Sens. Environ.*, vol. 237, p. 111537, 2020.
[38] X. Zhu, E. H. Helmer, F. Gao, D. Liu, J. Chen, and M. A. Lefsky, "A flexible spatiotemporal method for fusing satellite images with different resolutions," *Remote Sens. Environ.*, vol. 172, pp. 165-177, 2016.
[39] Q. Wang, Y. Tang, X. Tong, and P. M. Atkinson, "Virtual image pair-based spatio-temporal fusion," *Remote Sens. Environ.*, vol. 249, p. 112009, 2020.
[40] C. Xu, X. Du, Z. Yan, J. Zhu, S. Xu, and X. Fan, "VSDF: A variation-based spatiotemporal data fusion method," *Remote Sens. Environ.*, vol. 283, p. 113309, 2022.
[41] W. Shi, D. Guo, and H. Zhang, "A reliable and adaptive spatiotemporal data fusion method for blending multi-spatiotemporal-resolution satellite images," *Remote Sens. Environ.*, vol. 268, p. 112770, 2022.
[42] S. Hou *et al.*, "RFSDAF: A New Spatiotemporal Fusion Method Robust to Registration Errors," *IEEE T. Geosci. Remote.*, vol. 60, pp. 1-18, 2021.
[43] Y. Zhao, B. Huang, and H. Song, "A robust adaptive spatial and temporal image fusion model for complex land surface changes," *Remote Sens. Environ.*, vol. 208, pp. 42-62, 2018.
[44] S. Liu, J. Zhou, Y. Qiu, J. Chen, X. Zhu, and H. Chen, "The FIRST model: Spatiotemporal fusion incorrporting spectral autocorrelation," *Remote Sens. Environ.*, vol. 279, p. 113111, 2022.
[45] F. Zhou and D. Zhong, "Kalman filter method for generating time-series synthetic Landsat images and their uncertainty from Landsat and MODIS observations," *Remote Sens. Environ.*, vol. 239, p. 111628, 2020.
[46] J. Zhou *et al.*, "Sensitivity of six typical spatiotemporal fusion methods to different influential factors: A comparative study for a normalized difference vegetation index time series reconstruction," *Remote Sens. Environ.*, vol. 252, p. 112130, 2021.
[47] M. Liu *et al.*, "An Improved Flexible Spatiotemporal DAta Fusion (IFSDAF) method for producing high spatiotemporal resolution normalized difference vegetation index time series," *Remote Sens. Environ.*, vol. 227, pp. 74-89, 2019.
[48] S. Li, L. Xu, Y. Jing, H. Yin, X. Li, and X. Guan, "High-quality vegetation index product generation: A review of NDVI time series reconstruction techniques," *Int. J. Appl. Earth Obs.*, vol. 105, p. 102640, 2021.
[49] M. Wu *et al.*, "An improved high spatial and temporal data fusion approach for combining Landsat and MODIS data to generate daily synthetic Landsat imagery," *Information fusion*, vol. 31, pp. 14-25, 2016.
[50] M. Wu, W. Huang, Z. Niu, C. Wang, W. Li, and B. Yu, "Validation of synthetic daily Landsat NDVI time series data generated by the improved spatial and temporal data fusion approach," *Information Fusion*, vol. 40, pp. 34-44, 2018.
[51] D. Long *et al.*, "Generation of MODIS-like land surface temperatures under all-weather conditions based on a data fusion approach," *Remote Sens. Environ.*, vol. 246, p. 111863, 2020.
[52] J. Quan, W. Zhan, T. Ma, Y. Du, Z. Guo, and B. Qin, "An integrated model for generating hourly Landsat-like land surface temperatures over heterogeneous landscapes," *Remote Sens. Environ.*, vol. 206, pp. 403-423, 2018.
[53] J. Ma, H. Shen, P. Wu, J. Wu, M. Gao, and C. Meng, "Generating gapless land surface temperature with a high spatio-temporal resolution by fusing multi-source satellite-observed and model-simulated data," *Remote Sens. Environ.*, vol. 278, p. 113083, 2022.
[54] S. Wang *et al.*, "A classification-based spatiotemporal adaptive fusion model for the evaluation of remotely sensed evapotranspiration in heterogeneous irrigated agricultural area," *Remote Sens. Environ.*, vol. 273, p. 112962, 2022.
[55] Q. Wang, K. Peng, Y. Tang, X. Tong, and P. M. Atkinson, "Blocks-removed spatial unmixing for downscaling MODIS images," *Remote Sens. Environ.*, vol. 256, p. 112325, 2021.
[56] L. Zhang, Q. Weng, and Z. Shao, "An evaluation of monthly impervious surface dynamics by fusing Landsat and MODIS time series in the Pearl River Delta, China, from 2000 to 2015," *Remote Sens. Environ.*, vol. 201, pp. 99-114, 2017.
[57] M. Bousbaa *et al.*, "High-Resolution Monitoring of the Snow Cover on the Moroccan Atlas through the Spatio-Temporal Fusion of Landsat and Sentinel-2 Images," *Remote Sens-Basel.*, vol. 14, no. 22, p. 5814, 2022.
[58] X. Li *et al.*, "Monitoring high spatiotemporal water dynamics by fusing MODIS, Landsat, water occurrence data and DEM," *Remote Sens. Environ.*, vol. 265, p. 112680, 2021.
[59] B. Chen, L. Chen, B. Huang, R. Michishita, and B. Xu, "Dynamic monitoring of the Poyang Lake wetland by integrating Landsat and MODIS observations," *ISPRS J. Photogramm.*, vol. 139, pp. 75-87, 2018.
[60] C. Huang, Y. Chen, S. Zhang, and J. Wu, "Detecting, extracting, and monitoring surface water from space using optical sensors: A review," *Reviews of Geophysics*, vol. 56, no. 2, pp. 333-360, 2018.
[61] D. Guo, W. Shi, F. Qian, S. Wang, and C. Cai, "Monitoring the spatiotemporal change of Dongting Lake wetland by integrating Landsat and MODIS images, from 2001 to 2020," *Ecological Informatics*, vol. 72, p. 101848, 2022.
[62] B. Zhukov, D. Oertel, F. Lanzl, and G. Reinhackel, "Unmixing-based multisensor multiresolution image fusion," *IEEE T. Geosci. Remote.*, vol. 37, no. 3, pp. 1212-1226, 1999.
[63] R. Zurita-Milla, J. G. Clevers, and M. E. Schaepman, "Unmixing-based Landsat TM and MERIS FR data fusion," *IEEE Geosci. Remote S.*, vol. 5, no. 3, pp. 453-457, 2008.
[64] J. D. Watts, S. L. Powell, R. L. Lawrence, and T. Hilker, "Improved classification of conservation tillage adoption using high temporal and synthetic satellite imagery," *Remote Sens. Environ.*, vol. 115, no. 1, pp. 66-75, 2011.
[65] C. M. Gevaert and F. J. Garcıa-Haro, "A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion," *Remote Sens. Environ.*, vol. 156, pp. 34-44, 2015.
[66] W. Liu, Y. Zeng, S. Li, and W. Huang, "Spectral unmixing based spatiotemporal downscaling fusion approach," *Int. J. Appl. Earth Obs.*, vol. 88, p. 102054, 2020.
[67] D. Jia, C. Cheng, C. Song, S. Shen, L. Ning, and T. Zhang, "A hybrid deep learning-based spatiotemporal fusion method for combining satellite images with different resolutions," *Remote Sens.*, vol. 13, no. 4, p. 645, 2021.
[68] J. Amorós-López *et al.*, "Multitemporal fusion of Landsat/TM and ENVISAT/MERIS for crop monitoring," *Int. J. Appl. Earth Obs.*, vol. 23, pp. 132-141, 2013.
[69] L. Busetto, M. Meroni, and R. Colombo, "Combining medium and coarse spatial resolution satellite data to improve the estimation of sub-pixel NDVI time series," *Remote Sens. Environ.*, vol. 112, no. 1, pp. 118-131, 2008.
[70] Y. Xu, B. Huang, Y. Xu, K. Cao, C. Guo, and D. Meng, "Spatial and temporal image fusion via regularized spatial unmixing," *IEEE Geosci. Remote S.*, vol. 12, no. 6, pp. 1362-1366, 2015.
[71] X. Jiang and B. Huang, "Unmixing-Based Spatiotemporal Image Fusion Accounting for Complex Land Cover Changes," *IEEE T. Geosci. Remote.*, vol. 60, pp. 1-10, 2022.
[72] E. Alpaydın, "Soft vector quantization and the EM algorithm," *Neural networks*, vol. 11, no. 3, pp. 467-477, 1998.
[73] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google Earth Engine: Planetary-scale geospatial analysis for everyone," *Remote Sens. Environ.*, vol. 202, pp. 18-27, 2017.
[74] A. Likas, N. Vlassis, and J. J. Verbeek, "The global k-means clustering algorithm," *Pattern recognition*, vol. 36, no. 2, pp. 451-461, 2003.
[75] Y. Wang, X. Li, P. Zhou, L. Jiang, and Y. Du, "AHSWFM: Automated and Hierarchical Surface Water Fraction Mapping for Small Water Bodies Using Sentinel-2 Images," *Remote Sens-Basel.*, vol. 14, no. 7, p. 1615, 2022.

[76] J. Rover, B. K. Wylie, and L. Ji, "A self-trained classification technique for producing 30 m percent-water maps from Landsat data," *Int. J. Remote. Sens.*, vol. 31, no. 8, pp. 2197-2203, 2010.
[77] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5-32, 2001.
[78] M. K. Gumma *et al.*, "Agricultural cropland extent and areas of South Asia derived using Landsat satellite 30-m time-series big-data using random forest machine learning algorithms on the Google Earth Engine cloud," *GISci. Remote Sens.*, vol. 57, no. 3, pp. 302-322, 2020.
[79] Z. Mitraka, F. Del Frate, and F. Carbone, "Nonlinear spectral unmixing of landsat imagery for urban surface cover mapping," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 9, no. 7, pp. 3340-3350, 2016.
[80] F. Priem, A. Okujeni, S. van der Linden, and F. Canters, "Comparing map-based and library-based training approaches for urban land-cover fraction mapping from Sentinel-2 imagery," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 78, pp. 295-305, 2019.
[81] Q. Yuan *et al.*, "Deep learning in environmental remote sensing: Achievements and challenges," *Remote Sens. Environ.*, vol. 241, p. 111716, 2020.
[82] H. Yuan *et al.*, "Retrieving soybean leaf area index from unmanned aerial vehicle hyperspectral remote sensing: Analysis of RF, ANN, and SVM regression models," *Remote Sens-Basel.*, vol. 9, no. 4, p. 309, 2017.
[83] Y. Wang, G. Foody, X. Li, Y. Zhang, P. Zhou, and Y. Du, "Regression-based surface water fraction mapping using a synthetic spectral library for monitoring small water bodies," *GIScience & Remote Sensing*, vol. 60, no. 1, p. 2217573, 2023.

Xiaodong Li received the B.S. degree in geographic information system from China University of Geosciences, Wuhan, China, in 2006, and the M.S. and Ph.D. degrees in physical geography from the Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan, in 2009 and 2012, respectively. He is currently a Professor with the Innovation Academy for Precision Measurement Science and Technology, Chinese Academy of Sciences. He has been supported by the Youth Innovation Promotion Association of the Chinese Academy of Sciences and Hubei Province Natural Science Fund for Distinguished Young Scholars. His research interests include superresolution mapping and fusion of remotely sensed imagery.

Yalan Wang received the B.S. degree in geographic information science from the Chengdu University of Technology, Chengdu, China, in 2020. She is pursuing the Ph.D. degree in physical geography with the Innovation Academy for Precision Measurement Science and Technology, Chinese Academy of Sciences, Wuhan, and also with the University of Chinese Academy of Sciences, Beijing, China. Her research interests include water mapping of remotely sensed imagery and remote sensing applications in water resources.

Yihang Zhang received the B.S. degree in land resource management from China University of Geosciences, Wuhan, China, in 2012, and the Ph.D. degree in physical geography from the Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan, in 2017.
From 2015 to 2016, he was a Visiting Ph.D. Student supervised by P. M. Atkinson with the Lancaster Environment Centre, Faculty of Science and Technology, Lancaster University, Lancaster, U.K. Since July 2021, he has been a Post-Doctoral Fellow with the Lancaster Environment Centre joint supported by Lancaster University and CSC. He is currently an Associate Professor with Innovation Academy for Precision Measurement Science and Technology, Chinese Academy of Sciences. His research interests include global forest cover mapping and downscaling of remotely sensed imagery.

Shuwei Hou received the B.S. degree in electrical engineering and automation, the M.S. degree in pattern recognition and intelligent system, and the Ph.D. degree in Space science and technology from Xidian University, Xi'an, China, in 2002, 2005, and 2021 respectively. She joined the China Academy of Space Technology, Xi'an, where she is currently a senior engineer. Her research interests include processing of remote sensing image data, computer vision, and pattern recognition.

Pu Zhou received the B.S. degree in physical geography from China University of Geosciences, Wuhan, China, in 2019. He is pursuing the Ph.D. degree in physical geography with the Innovation Academy for Precision Measurement Science and Technology, Chinese Academy of Sciences, Wuhan, and also with the University of Chinese Academy of Sciences, Beijing, China. His research interests include remotely sensed image fusion and deep learning.

Xia Wang received the B.S. and M.S. degrees in land resource management from the China University of Geosciences, Wuhan, in 2012 and 2014, respectively, followed by her Ph.D. degree from Wuhan University in 2018. From 2021 to 2022, she serves as a visiting scholar at the Lancaster Environment Centre, Faculty of Science and Technology, Lancaster University, Lancaster, U.K. Currently, she holds the position of assistant researcher at Wuhan Botanical Garden, Chinese Academy of Sciences. Her research pursuits encompass the field of ecological environment assessment and environmental remote sensing monitoring.

Yun Du received the B.S. degree in geomorphology and quaternary geology from Nanjing University, Nanjing, China, in 1989, the M.S. degree in physical geography from the Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan, China, in 1992, and the Ph.D. degree in historical geography from Wuhan University, Wuhan, in 1999. He is currently a Professor with Innovation Academy for Precision Measurement Science and Technology, Chinese Academy of Sciences. His research interests include environment monitoring and evaluation.

Giles M. Foody earned the B.Sc. and Ph.D. degrees from the University of Sheffield, Sheffield, U.K., in 1983 and 1986, respectively.
He is Professor of Geographical Information Science in the School of Geography, University of Nottingham, Nottingham, U.K. He has produced nine books and more than 260 refereed journal articles and his work has been cited more than 38,000 times (*h* index = 95). His main research interests lie at the interface between remote sensing, ecology, and informatics with a core focus on image classification for land surface cover mapping and monitoring applications at scales ranging from the subpixel to global.
He served as the founding Editor-in-Chief of *Remote Sensing Letters*. He holds additional editorial roles on over ten other journals including *Landscape Ecology*, the *International Journal of Remote Sensing*, *Remote Sensing of Environment*, *Remote Sensing* and the *International Journal of Applied Earth Observation and Geoinformation*. He was elected a Fellow of the IEEE (FIEEE) in 2013 and a member of Academia Europaea (MAE) in 2021.