SCIENCE CHINA
Information Sciences

April 2020, Vol. 63 140301:1–140301:17
https://doi.org/10.1007/s11432-019-2785-y

• REVIEW •
Special Focus on Deep Learning in Remote Sensing Image Processing

Spatio-temporal fusion for remote sensing data: an overview and new benchmark

Jun LI[1*], Yunfei LI[1], Lin HE[2*], Jin CHEN[3] & Antonio PLAZA[4]

[1]*Guangdong Provincial Key Laboratory of Urbanization and Geo-simulation, School of Geography and Planning, Sun Yat-sen University, Guangzhou* 510275*, China;*
[2]*School of Automation Science and Engineering, South China University of Technology, Guangzhou* 510640*, China;*
[3]*State Key Laboratory of Earth Surface Processes and Resource Ecology, Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing* 100875*, China;*
[4]*Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, Cáceres* E-10071*, Spain*

Received 3 December 2019/Revised 18 January 2020/Accepted 10 February 2020/Published online 9 March 2020

Abstract Spatio-temporal fusion (STF) aims at fusing (temporally dense) coarse resolution images and (temporally sparse) fine resolution images to generate image series with adequate temporal and spatial resolution. In the last decade, STF has drawn a lot of attention and many STF methods have been developed. However, to date the STF domain still lacks benchmark datasets, which is a pressing issue that needs to be addressed in order to foster the development of this field. In this review, we provide (for the first time in the literature) a robust benchmark STF dataset that includes three important characteristics: (1) diversity of regions, (2) long timespan, and (3) challenging scenarios. We also provide a survey of highly representative STF techniques, along with a detailed quantitative and qualitative comparison of their performance with our newly presented benchmark dataset. The proposed dataset is public and available online.

Keywords spatio-temporal fusion (STF), multi-temporal remote sensing data, benchmark data, experimental validation

Citation Li J, Li Y F, He L, et al. Spatio-temporal fusion for remote sensing data: an overview and new benchmark. Sci China Inf Sci, 2020, 63(4): 140301, https://doi.org/10.1007/s11432-019-2785-y

1 Introduction

The analysis of multi-temporal image series is necessary and important in many remote sensing applications, such as vegetation/crop monitoring and estimation [1–7], evapotranspiration estimation [8], atmosphere monitoring [9], land-cover/land-use change detection [10], surface dynamic mapping [11], ecosystem monitoring [12], soil water content analysis [13], and detailed analysis of human-nature interactions [14]. The changes in the surface of the Earth usually occur within short time, and require high spatial resolution in order to be properly modeled. However, the spatial resolution of satellite instruments such as MODIS (250, 500 and 1000 meter-pixels), AVHRR (1100 meter-pixels) and OrbView-2 (1000 meter-pixels) is generally low. While sensors like Landsat-TM/ETM+/OLI and SPOT5-HRG/HRS/VGT are able to provide images with higher spatial resolution, they can just provide temporally sparse image series owing to their long revisit cycle (e.g., 16 days in the case of Landsat, or 26 days in the case of SPOT5). This is a consequence of the tradeoff between the pixel size and the scanning swath of the sensors [15], which means that no available instruments are able to offer temporally dense, high spatial resolution

* Corresponding author (email: lijun48@mail.sysu.edu.cn, helin@scut.edu.cn)

image series. For convenience, in the following we refer to fine and coarse satellite images according to their spatial resolution.

Spatio-temporal fusion (STF) aims at fusing (temporally dense) coarse spatial resolution images and (temporally sparse) fine spatial resolution images to generate image series with adequate temporal and spatial resolution. In the last decade, STF has drawn a lot of attention, with many new STF methods developed. However, to date the datasets used to test different STF approaches are very different, which leads to lack of standardization in the validation of these approaches. In other words, the STF domain still lacks benchmark datasets, which is a pressing issue that needs to be addressed. In order to develop robust benchmark STF datasets, three important characteristics need to be satisfied: (1) diversity of regions, (2) long timespan, and (3) challenging scenarios. We elaborate on these requirements below:

(1) Diversity of regions. There are multiple application domains for STF techniques, such as urban, rural, forest and mountain areas. As a result, any benchmark dataset for STF techniques should include as many different land-cover scenarios as possible. In other words, the more diverse the datasets are, the more comprehensive the assessment of an STF method will be.

(2) Long timespan. The land surface changes for a specific area can be very diverse. As a result, benchmark datasets should cover these possible changes as much as possible. Hence, image sequences covering a long timespan are highly desirable. Such sequences are quite important in order to test the robustness of an STF method. On the other hand, from a methodological viewpoint, the development of deep learning-based methods requires a relatively large number of training data sets, and long timespan data are therefore suitable in this context.

(3) Challenging scenarios. There are several additional challenges for STF method validation, including the spatial resolution gap between fine and coarse images, the characterization of changes in heterogeneous areas, and the prediction of land-cover changes. These aspects should also be included in any relevant benchmark dataset for STF methods.

Based on the three aforementioned aspects, we present a new comprehensive benchmark dataset for evaluation of STF methods. We also review some highly representative STF techniques and provide a quantitative and qualitative comparison of their performance with our newly presented benchmark, which has been made publicly available.

The remainder of the review is organized as follows. Section 2 provides a survey on available STF methods. Section 3 describes our new benchmark dataset, addressing its main characteristics in terms of diversity of regions, long timespan and challenging scenarios. Section 4 provides an experimental comparison of some of the algorithms described in Section 2 with the benchmark dataset in Section 3. Finally, Section 5 concludes the paper with some remarks and hints at plausible future research lines.

2 STF methods

Following the recent review in [15], we divide existing STF methods into five main categories: weight function-based methods, unmixing-based methods, learning-based methods, Bayesian-based methods, and hybrid methods. In the following, we provide details on each of the aforementioned STF categories.

2.1 Weight function-based methods

The spatial and temporal adaptive reflectance fusion model (STARFM) is the first weight function-based STF method developed in the literature. This method first assumes that all the pixels in the coarse images are pure. It uses a weighted strategy to add the reflectance changes between two coarse images to the prior fine image so as to predict the target image. STARFM has been shown to be able to capture phenological changes. However, its performance in highly heterogeneous landscapes and in the task of capturing land-cover changes is limited [16].

STARFM has served as the basis for the development of many weight function-based methods. For instance, the spatial-temporal adaptive algorithm for mapping reflectance change (STAARCH) uses tasseled cap transformations to detect the time of occurrence of land-cover changes from the coarse image series,

successfully improving STARFM in this particular task [17]. The enhanced spatial and temporal adaptive reflectance fusion model (ESTARFM) improves the STARFM prediction accuracy in heterogeneous areas by introducing a conversion coefficient between the reflectance from coarse images and that from fine images [18]. The method in [19] has also been developed to consider sensor observation differences, changing the way in which STARFM calculates pixel weights [19]. The operational data fusion framework of the integrated STARFM integrates STARFM and some pre-processing methods into a unified framework that includes angular corrections on the coarse images, precise and automatic co-registration on both the coarse and fine image pairs, and automatic selection of fine and coarse paired dates [20]. The area-to-point regression kriging-based STARFM uses area-to-point regression kriging to downscale the coarse images, and then applies STARFM [21]. The robust adaptive spatial and temporal fusion model divides the surface changes into shape changes and non-shape changes, then uses a patch-level STARFM to predict the non-shape changes. After that, a principal component analysis (PCA)-based change detection and a nonlocal sparse regression model are used for the prediction of shape changes. Finally, a regression-based high-pass modulation is conducted to further reduce the prediction error and improve the spatial details [22]. The bilateral filter based spatio-temporal method considers the temperature of ground objects in urban areas, and uses bilateral filtering to determine the weights of neighbouring pixels in STARFM to produce high spatio-temporal resolution land surface temperature maps [23]. The weighted combination of kernel-driven and fusion-based methods first uses a kernel-driven method followed by STARFM to generate different high spatio-temporal resolution land surface temperature maps, and then fuses these maps to obtain the final prediction [24]. The spatio-temporal adaptive data fusion algorithm for temperature mapping characterizes the annual cycle of land surface temperature and landscape heterogeneity in urban areas, incorporating temporal changes of radiance into STARFM to improve STARFM for STF of land surface temperature [25]. The method in [26] proposes a downscaling approach that combines the STARFM and the universal triangle method to retrieve daily surface soil moisture. The spatio-temporal enhancement method for medium resolution leaf area index (LAI) first downsamples the LAI map from MODIS images, and then uses STARFM to fuse the obtained information with the LAI map from Landsat images [27]. The spatio-temporal integrated temperature fusion model changes the framework of STARFM to make it able to fuse data collected from arbitrary sensors, i.e., beyond the simple scenario in which only two different sensors are considered for STF [28]. The modified ESTARFM introduces land cover data as an auxiliary source of information for spectrally similar neighboring pixels [29]. Finally, the method in [30] uses phenological information extracted from MODIS vegetation index time series to improve the ESTARFM in the task of predicting the reflectance of paddy rice.

There are also several weight function-based methods that are not based on STARFM. The spatial and temporal nonlocal filter-based fusion model uses two regression coefficients to describe the land surface changes and introduces nonlocal filtering to take advantage of the redundancy of fine images to obtain more accurate and robust predictions [31]. The spatiotemporal image-fusion model first calculates the ratio of different coarse images, and then classifies the resulting ratio image into three clusters. Different clusters perform different forms of linear regression to obtain a final prediction [32]. The rigorously-weighted spatiotemporal fusion model uses geo-statistical ordinary kriging and incorporates this approach when calculating the weights of neighboring pixels, using uncertainty analysis to predict the fine image [33]. The semi-physical fusion approach uses the MODIS bidirectional reflectance function (BRDF)/albedo product (and existing Landsat observations) to predict Landsat reflectance [34]. The spatiotemporal reflectance fusion method integrating image inpainting and steering kernel regression fusion model first detects land-cover changes and then fills them with unchanged similar pixels by an exemplar-based inpainting technique. The weights of local neighbouring pixels are adaptively determined by a steering kernel regression to predict fine images [35]. The spatiotemporal model incorporating autoregressive error correction uses a spatiotemporal autoregressive model to minimize autoregressive errors when fitting a relationship between coarse and fine pixels. Then, the relationship is used to predict the target images [36]. The area-to-point regression kriging based approach directly uses area-to-point regression kriging to accomplish STF of Landsat 8 OLI and Sentinel-2 MSI data [37]. The method in [38] uses a temporal high-pass modulation to accomplish STF. The method in [39] first uses a linear injection

model to extract the spatial details from the prior fine image to generate a transitional prediction. Then, a weight strategy similar to that of STARFM is designed to further improve the prediction. The best linear unbiased estimation-based STF method accounts for the phenological characteristics of vegetation, providing annual time series of NDVI data with high spatial resolution as the background field and that with low spatial resolution as the observation field, and then fuses these sources of information using the best linear unbiased estimator to obtain high spatio-temporal resolution NDVI data [40]. The spatio-temporal vegetation index image fusion model proposes a new weighting system to disaggregate the total NDVI change within a moving window to predict the NDVI change for each fine pixel and generate NDVI time series in heterogeneous regions [41]. The method in [42] uses simple linear regression to integrate Landsat and MODIS to generate high spatial resolution evapotranspiration map series. The hybrid color mapping method uses hybrid color mapping to establish the relation between coarse images from different times, and then the relation is utilized on the fine image to obtain the final prediction [43]. The spatial-temporal fraction map fusion model first generates fine resolution fraction change maps by using kernel ridge regression, and then uses a temporal-weighted fusion model to obtain a fine resolution fraction map of the predicted date [44]. Fit-FC uses three models, i.e., regression fitting, spatial filtering and residual compensation to conduct the STF of between Sentinel-2 and Sentinel-3 images [45].

2.2 Unmixing-based methods

The multisensor multiresolution technique (MMT) [46] was the first unmixing-based STF method in the literature. It conducts classification on the prior fine resolution images, assuming that the coarse pixels are linearly mixed by the classes from the classification map. It then unmixes the coarse pixels at the prediction date within a moving window to get the reflectance change of each class, obtaining a final prediction. MMT has been served as a baseline for many other unmixing-based STF methods. For instance, Ref. [47] introduces constraints into the linear unmixing solution to ensure that the values of reflectance changes are positive and within an appropriate range. The method in [48] accounts for the within-class NDVI spatial variability by introducing a locally calibrated multivariate regression model in unmixing. The spatial temporal data fusion approach (STDFA) first classifies multi-NDVI images from multiple fine images to introduce the temporal change information, and then unmixes the coarse pixels to obtain the reflectance change of each class to generate the predicted fine images [49]. The modified spatial and temporal data fusion approach improves the STDFA by using an adaptive window size selection method to select the best window size and moving steps for the disaggregation of coarse pixels [50]. The enhanced spatial and temporal data fusion approach introduces a patch-based ISODATA classification method, sliding window technology, and the temporal-weight concept to improve the STDFA [51]. The unmixing-based spatio-temporal reflectance fusion model first adopts a change trend ratio to physically unmix the MODIS difference images over a given period, and then unmixes the coarse pixels to calculate the change trend ratio to get the final prediction [52]. The regularized spatial unmixing based method (RSpatialU) utilizes prior class spectra to regularize the unmixing process and reduce the unmixing error [53]. The database unmixing method considers that each coarse pixel comprises a mixture of fine pixels, and resolves the coarse pixel into spatially distributed fine pixels based on the statistical relationship of a match-up. It then creates a lookup table for each fine pixel, connecting fine pixels to coarse pixels to predict a fine map from a (previously observed) coarse map [54]. The object based spatial and temporal vegetation index unmixing model first segments the prior fine image to extract the endmembers, and then conducts unmixing to obtain the final prediction [55]. Finally, the NDVI linear mixing growth model uses unmixing to calculate the growth rate to fuse the NDVI from Landsat and MODIS images [56].

2.3 Learning-based methods

The first leaning-based STF method in the literature was called sparse-representation-based spatiotemporal reflectance fusion model (SPSTFM). This method established the relation between reflectance changes from prior coarse images and that from prior fine images by using dictionary learning. Then, it exploits

the resulting information to predict the target fine images [57]. The one-pair learning SPSTFM first improves the spatial resolution of coarse images using dictionary learning, and then adopts a high-pass module to fuse the resulting information with the prior fine image to get the final prediction [58]. The SPSTFM and the one-pair learning SPSTFM have been proved to be effective in the task of capturing land-cover changes [57,58]. Currently, most learning-based STF methods are based on dictionary learning. For instance, the enhanced one-pair learning SPSTFM combines a spatially extended mode and a temporally extended mode to increase the training set. This approach successfully improves the performance of the one-pair learning SPSTFM [59]. The method in [60] exploits high-spectral correlation (across the spectral domain) and high self-similarity (across the spatial domain) to learn a spatio-spectral fusion basis, and then associates temporal changes using a local constraint sparse representation to develop a spatial-spectral-temporal fusion model. The error-bound-regularized semi-coupled dictionary learning method utilizes semi-coupled dictionary learning to address the differences between the high spatial resolution and low spatial resolution images, and then adopts an error-bound-regularized model by imposing error bound regularization [61]. The block sparse Bayesian learning based semi-coupled dictionary learning method explores the inherent characteristics of sparse coefficients and improves the STF results by means of a priori structural sparse constraints [62]. The compressed sensing based spatiotemporal fusion method uses a downsampling process (under the framework of compressed sensing) for reconstruction. With the coupled dictionary to constrain the similarity of sparse coefficients, a new dictionary-based STF method is built [63].

There are also several learning-based methods not based on dictionary learning. The extreme learning machine based STF model is similar to SPSTFM; it performs learning between two fine and coarse images with the extreme learning machine, significantly reducing the processing time [64]. The STF using deep convolutional neural networks (STFDCNN) adopts CNNs to carry out STF. Its core idea is similar to that of the one-pair-SPSTFM, i.e., super-resolution and fusion are conducted via a high-pass module [65]. The two-stream convolutional neural network for spatiotemporal fusion (StfNet) takes into account the temporal dependence and temporal consistency among image sequences in the CNN-based superresolution process [66]. The deep convolutional STF network proposes a three-part (low-frequency information extraction, high-frequency information extraction, fusion) CNN for STF [67]. The method in [68] fuses the fractions of absorbed photosynthetically active radiation from MODIS images and those from Landsat images by utilizing a multiple resolution tree. The method in [69] uses the support vector regression and random forests to model the relationship between the Landsat indicators and MODIS 8-day 1 km evapotranspiration. The resulting relationship is used to predict high spatial resolution evapotranspiration. The method in [70] utilizes the regression-tree to fuse the eMODIS and Landsat 8 information to generate synthetic NDVI data. The hybrid wavelet-artificial intelligence fusion approach method combines wavelet transformation with artificial intelligence approaches to blend MODIS and Landsat 8 data to predict land surface temperature series [71]. We have recently proposed a new CNN-based STF method, named sensor-bias driven spatio-temporal fusion model (BiaSTF) [72]. The main characteristic of this method is that it includes the sensors bias of Landsat and MODIS into the fusion procedure. First, it uses a CNN to improve the reflectance from the coarse images. Then, another CNN is utilized to learn the sensors bias of the two type images. The improved reflectance change and bias contribute to obtaining the final prediction.

2.4 Bayesian-based methods

The spatio-temporal Bayesian data fusion (STBDF) incorporates the temporal correlation information in the image time series and casts the fusion problem as an estimation one, in which the fused image is obtained by the maximum a posteriori estimator. This approach is suitable for heterogeneous landscapes [73]. The method in [74] uses the Bayesian maximum entropy to blend the sea surface temperature from MODIS images and that from AMSR-E. The method in [75] proposes a uniformed Bayesian framework for both spatial-spectral fusion and STF. The method in [63] proposes a spatio-spectral-temporal fusion model which utilizes a maximum posteriori probability to describe an inverse fusion problem.

Then, it constructs an integrated relationship model with all the involved data. Finally, the fused image is obtained by the classical conjugate gradient optimization algorithm. The method in [76] proposes a spatio-spectral-temporalfusion model which utilizes a maximum posteriori probability to describe an inverse fusion problem.

2.5 Hybrid methods

Some STF methods integrate the unmixing strategy, Bayesian theory and weight functions to pursue better performance. For instance, the spatial and temporal reflectance unmixing model (STRUM) conducts unmixing (constrained by Bayesian theory) to the pixels of the change map from coarse images in order to first obtain the reflectance change of each class. Then, it uses weighted functions (in the same way as STARFM) to create a fused image using moving windows [77]. The improved STRUM replaces the classification map by using abundance images obtained from the prior fine images, and then introduces sensor difference adjustment to improve the STRUM method [78]. The improved STARFM with the help of an unmixing-based method first uses an unmixing strategy to get the abundance images from coarse images. Then, the unmixing images replace the coarse images to be fed into SARFM, to obtain the final prediction [79]. The flexible spatiotemporal data fusion (FSDAF) first conducts unmixing to get a transitional prediction and a residual between transitional prediction and the interpolated coarse image in the predicted date. Then, the residual is distributed in fine resolution scales. Finally, a weighted strategy is used to further improve the prediction [80]. The improved FSDAF modifies the FSDAF by introducing a constrained least squares process to combine the increment from unmixing and that from the interpolation of coarse images [81]. The method in [82] combines STARFM and an unmixing strategy in the same way as USTARFM. The spatial-temporal remotely sensed images and land cover maps fusion model uses an unmixing strategy and Bayesian theory to fuse coarse image series and a few land-cover maps from fine images to generate a land-cover map series with high spatial resolution [83]. The enhanced linear STF method characterizes the slope and intercept of the linear model as the residual caused by systematic biases, and then calculates them based on spectral unmixing theory to obtain the final prediction. Then, a weight strategy is used to enhance the prediction [84]. The integrated framework to blend spatiotemporal temperatures adopts a similar idea as the one pursued by FSDAF to blend the land surface temporal changes from Landsat, MODIS and images from a geostationary satellite (FY-2F) [85]. The improved ESTARFM introduces an unmixing strategy to improve the accuracy of spectrally similar pixels in ESTARFM [86]. The NDVI-Bayesian spatiotemporal fusion model first uses a multi-year average MODIS NDVI time series to constrain the unmixing process in a Bayesian framework to obtain the initial downscaled NDVI, and then uses the relation between the initial NDVI and real Landsat NDVI (on other dates) to generate high spatial and temporal resolution NDVI data [87]. Finally, the improved Bayesian data fusion approach improves the performance of STBDF in heterogenous areas by introducing an unmixing stragety [88].

3 Datasets

In this section, we describe our benchmark, which consists of three Landsat-MODIS datasets (called hereinafter AHB dataset, Tianjin dataset and Daxing dataset) that are respectively collected over Ar Horqin Banner of Inner Mongolia province, Tianjin city, and Daxing district of Beijing, China. These datasets are specifically intended to perform STF. All the fine resolution images (i.e., the Landsat images) are acquired by Landsat-8 OLI with six bands, including the blue band (0.45–0.51 μm), the green band (0.53–0.59 μm), the red band (0.64–0.67 μm), the near-infrared band (0.85–0.88 μm), the short-wave infrared-1 band (1.57–1.65 μm), and the short-wave infrared-2 band (2.11–2.29 μm). The coarse resolution images (i.e., the MODIS images with 500 m spatial resolution) are geometrically transformed with respect to the corresponding Landsat images. In the AHB dataset, these images are MOD09GA images, and in the Tianjin and Daxing datasets, these images are MOD02HKM. The bands considered in the MODIS images are selected and reordered to match the Landsat images. Both the Landsat images and the MOD02HKM

Table 1 Summary of the three considered datasets

Dataset	Image size	Pairs	Timespan	Main change
AHB	2480×2800×6	27	2013/05/30 – 2018/12/06	Phenological changes in rural areas
Tianjin	2100×1970×6	27	2013/09/01 – 2019/09/18	Phenological changes in urban areas
Daxing	1640×1640×6	29	2013/09/01 – 2019/11/05	Land-cover changes

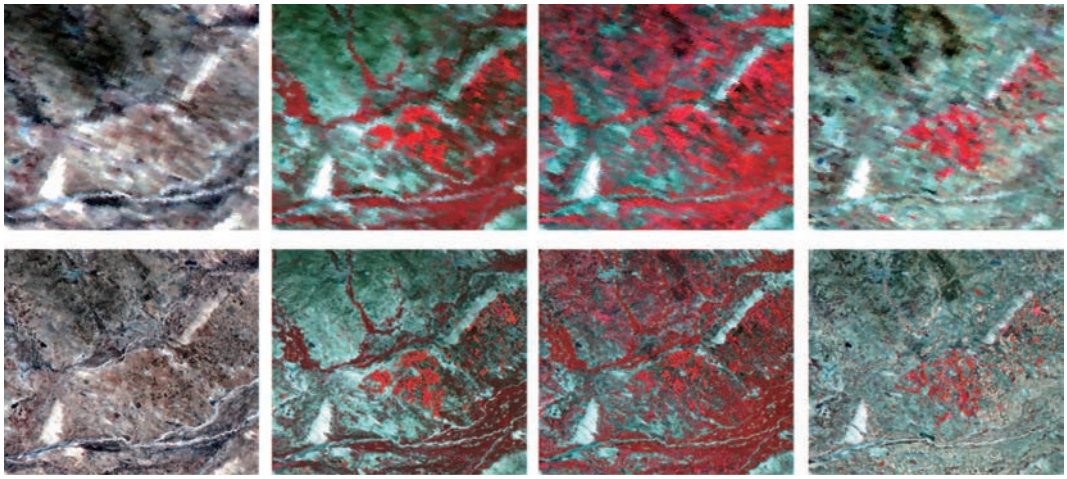

Figure 1 (Color online) Example pairs from the AHB dataset.

images are atmospherically corrected by the Quick Atmospheric Correction (QUAC) algorithm. The summary of them is shown in Table 1. In the following, we provide a detailed description of the three datasets that conform our benchmark.

3.1 AHB dataset

The main purpose of this dataset is to provide a benchmark for testing the accuracy of STF methods in the task of detecting phenological changes in rural areas. The Ar Horqin Banner (43.3619°N, 119.0375°E) is located in the northeast of China. Agriculture and animal husbandry are the major industries of Ar Horqin Banner, for which there are a lot of circular pastures and farmlands. In this site, we collected 27 cloud-free Landsat-MODIS image pairs from 2013/05/30 to 2018/12/06, with a timespan of more than 5 years. This area experienced significant phenological changes owing to the growth of crops and other kinds of vegetation. Four pairs of AHB datasets are displayed in the 1st and 2nd rows of Figure 1, from which we can infer that this area is heterogenous and with significant phenological changes.

3.2 Tianjin dataset

The main purpose of this dataset is to provide a benchmark for testing the accuracy of STF methods in the task of detecting phenological changes in urban areas. Tianjin (39.8625°N, 117.8591°E) is a municipality in the north of China with clear seasonal changes during the year. The Tianjin dataset includes 27 cloud-free Landsat-MODIS image pairs from the 2013/09/01 to 2019/09/18. For illustrative purposes, Figure 2 shows a few sample images from this dataset, from which we can see that there are significant phenological changes in these six pairs.

3.3 Daxing dataset

The main purpose of this dataset is to provide a benchmark for evaluating the performance of STF in the task of detecting land-cover changes. The Daxing dataset includes 29 cloud-free Landsat-MODIS image pairs from 2013/09/01 to 2019/11/05, collected from the Daxing district (39.0009°N, 115.0986°E) located in the south of Beijing city. The Beijing Daxing international airport, which was constructed between December 2014 to September 2019, is exactly inside this site, representing a gradual land cover

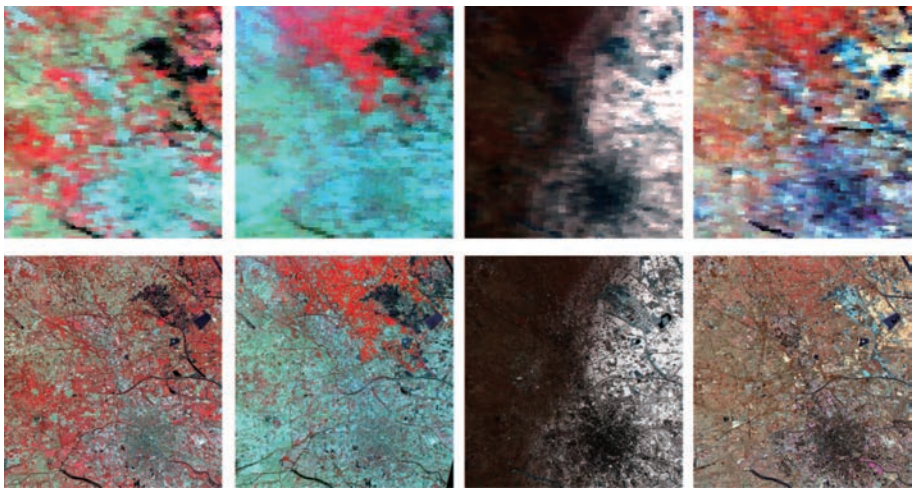

Figure 2 (Color online) Example pairs from the Tianjin dataset.

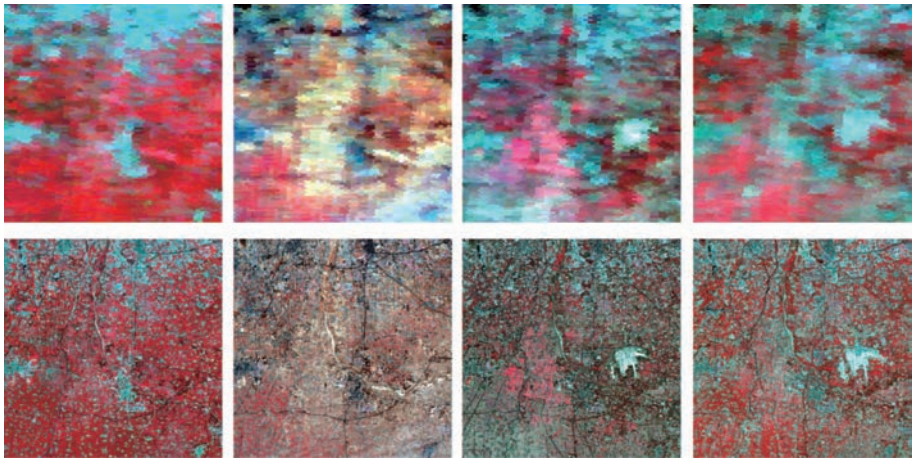

Figure 3 (Color online) Example pairs from the Daxing dataset.

change. In addition, this dataset contains an obvious phenological change as well. Four pairs of the Daxing dataset are displayed in the last two rows of Figure 3, from which we can observe that there are obvious land-cover changes in these pairs.

At this point, we would like to emphasize that the considered benchmark exhibits the following characteristics.

- Diversity of regions: representing rural, urban and mixed regions.
- Long timespan: 5 years or more.
- Challenging scenarios: suffering from phenological changes, land-cover changes, and heterogeneity, with different types of changes.

Finally, we would like to point out that there is significant strip noise in the two short-wave infrared bands in the MODIS images, which brings a great challenge for STF and will not be taken into account in our experiments in Section 4. Nevertheless, we keep these two bands in our datasets as we believe that this issue should be taken into account in future developments. The aforementioned benchmark datasets are available online[1)2)].

1) This is mainly for Chinese scholars. https://pan.baidu.com/s/1ymgud6tnY6XB5CTCXPUfnw.
2) This is more international-friendly. https://drive.google.com/open?id=1yzw-4TaY6GcLPIRNFBpchETrFKno30he.

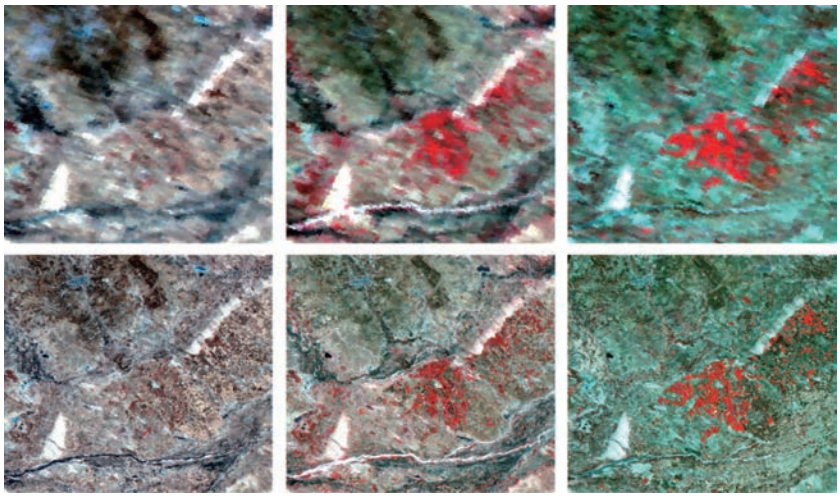

Figure 4 (Color online) Test data from the AHB dataset, where the 1st row displays the MODIS images and the 2nd row displays the Landsat images.

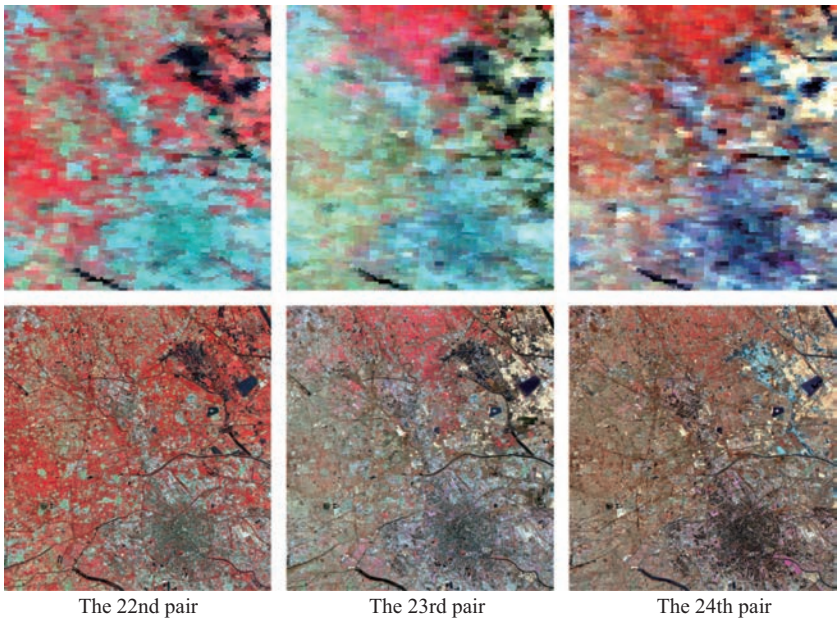

Figure 5 (Color online) Test data from Tianjin dataset.

4 Experiments

4.1 Experimental results of baseline methods

To offer a baseline performance as a reference, in our first experiment we test five standard STF methods, including STARFM [16], ESTARFM [18], FSDAF [80], STFDCNN [65], and BiaSTF [72] on the considered benchmark. The parameters of these methods are set according to the original contributions to ensure their optimal performance. For the AHB dataset, we choose the 23rd, 24th and 25th pairs as the test data. Similarly, the 22nd, 23rd and 24th pairs of Tianjin dataset, the 25th, 26th and 27th pairs of Daxing dataset are selected. Figure 4 displays the three selected image pairs from AHB dataset, while the pairs for Tianjin dataset and Daxing dataset are displayed in Figures 5 and 6, respectively.

Tables 2–4 show the quantitative assessment of the aforementioned standard STF methods. It can be observed that, for the AHB dataset, the ESTARFM and BiaSTF obtain the best performance. Generally

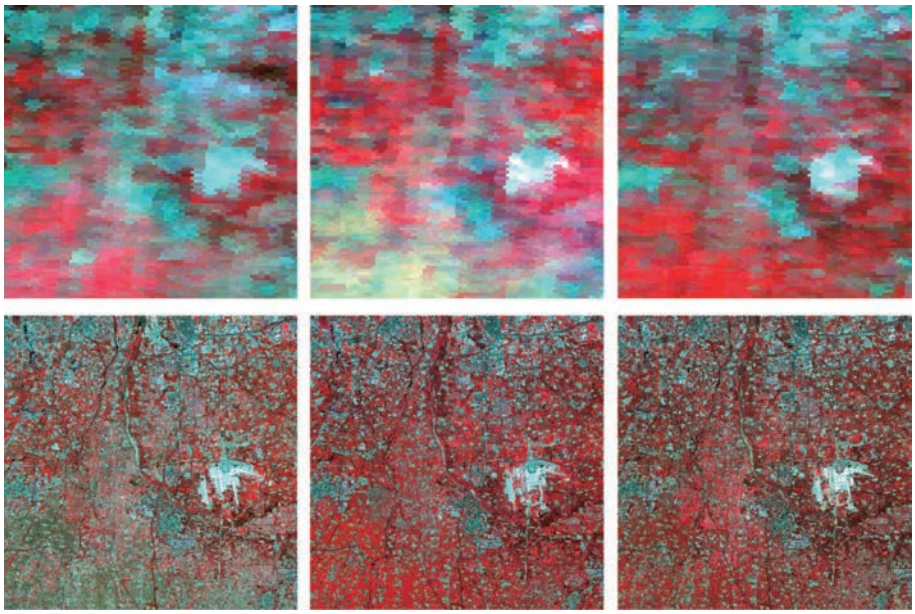

Figure 6 (Color online) Test data from Daxing dataset.

Table 2 Quantitative assessment of experimental results on the AHB dataset

	Band	STARFM	ESTARFM	FSDAF	STFDCNN	BiaSTF
RMSE	Band 1	0.0286	0.0159	0.0300	0.0171	0.0136
	Band 2	0.0355	0.0222	0.0366	0.0244	0.0248
	Band 3	0.0552	0.0405	0.0563	0.0419	0.0426
	Band 4	0.0666	0.0680	0.0675	0.0675	0.0660
CC	Band 1	0.6781	0.5188	0.7006	0.5684	0.7200
	Band 2	0.7033	0.7207	0.7296	0.6382	0.7496
	Band 3	0.7120	0.7376	0.7371	0.6835	0.7634
	Band 4	0.6972	0.7260	0.7232	0.6416	0.7313
SSIM	Band 1	0.7922	0.8432	0.7828	0.8441	0.9026
	Band 2	0.7795	0.8442	0.7818	0.8065	0.8499
	Band 3	0.7263	0.7765	0.7331	0.7489	0.7942
	Band 4	0.7411	0.7574	0.7572	0.7017	0.7649
ERGAS		2.007	1.3899	2.0730	1.849	1.3963
SAM		0.110	0.0800	0.1159	0.0941	0.0720

speaking, BiaSTF achieves better or competitive results when compared to other methods. It is also remarkable that STFDCNN also obtains very promising performance. The other three standard methods generally exhibit worse performance than that obtained by learning based methods, i.e., BiaSTF and STFDCNN.

Finally, Figures 7–9 display the predicted images of all considered STF methods, along with the ground truth pairs. It can be observed that all the methods achieve good performance.

4.2 Experimental results of deep learning based methods

In this subsection, we report the experiments conducted on our benchmark dataset using two deep learning based methods, i.e., STFDCNN [65] and BiaSTF [72], where STFDCNN models the change from coarse images and BiaSTF considers the sensor bias in the modeling. The main reasons why we consider these two CNN-based methods for evaluation purposes are three-fold: (1) the first experiment demonstrates the strong potential of learning based methods, which generally achieve better performance

Table 3 Quantitative assessment of experimental results on the Tianjin dataset

	Band	STARFM	ESTARFM	FSDAF	STFDCNN	BiaSTF
RMSE	Band 1	0.0241	0.0212	0.0226	0.0274	0.0234
	Band 2	0.0312	0.0240	0.0297	0.0389	0.0242
	Band 3	0.0375	0.0342	0.0347	0.0452	0.0310
	Band 4	0.0896	0.1425	0.0872	0.0602	0.0853
CC	Band 1	0.8385	0.8612	0.8462	0.7693	0.8193
	Band 2	0.7740	0.8367	0.7812	0.5980	0.8301
	Band 3	0.6863	0.7957	0.7147	0.6191	0.7985
	Band 4	0.6883	0.3121	0.7155	0.8179	0.7264
SSIM	Band 1	0.8722	0.8886	0.8813	0.8220	0.8582
	Band 2	0.8158	0.8640	0.8222	0.6821	0.8579
	Band 3	0.7430	0.8127	0.7651	0.6752	0.8220
	Band 4	0.6874	0.3258	0.7000	0.8117	0.7204
ERGAS		1.6180	1.7313	1.5304	1.9832	1.4292
SAM		0.1965	0.1656	0.1766	0.1574	0.1443

Table 4 Quantitative assessment of experimental results on the Daxing dataset

	Band	STARFM	ESTARFM	FSDAF	STFDCNN	BiaSTF
RMSE	Band 1	0.0124	0.0152	0.0127	0.0177	0.0121
	Band 2	0.0161	0.0160	0.0159	0.0187	0.0153
	Band 3	0.0221	0.0219	0.0213	0.0251	0.0209
	Band 4	0.0429	0.0509	0.0419	0.0519	0.0456
CC	Band 1	0.9397	0.9338	0.9406	0.9038	0.9478
	Band 2	0.9239	0.9307	0.9284	0.9120	0.9324
	Band 3	0.8962	0.8985	0.9025	0.8768	0.9062
	Band 4	0.7775	0.7079	0.7885	0.7020	0.7486
SSIM	Band 1	0.9556	0.9434	0.9543	0.9226	0.9594
	Band 2	0.9398	0.9429	0.9410	0.9277	0.9452
	Band 3	0.9140	0.9155	0.9176	0.8962	0.9218
	Band 4	0.8011	0.7371	0.8109	0.7320	0.7753
ERGAS		0.9642	1.0436	0.9524	1.3085	0.9328
SAM		0.0673	0.0706	0.0660	0.0794	0.0658

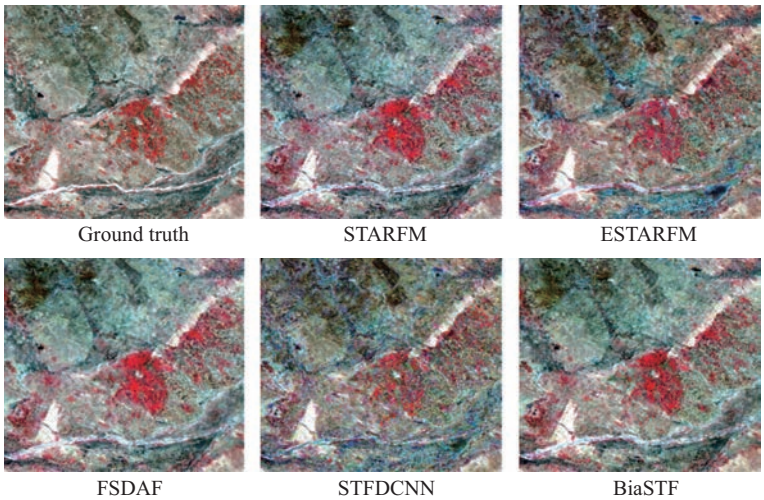

Figure 7 (Color online) Ground truth image and obtained predictions for the AHB dataset.

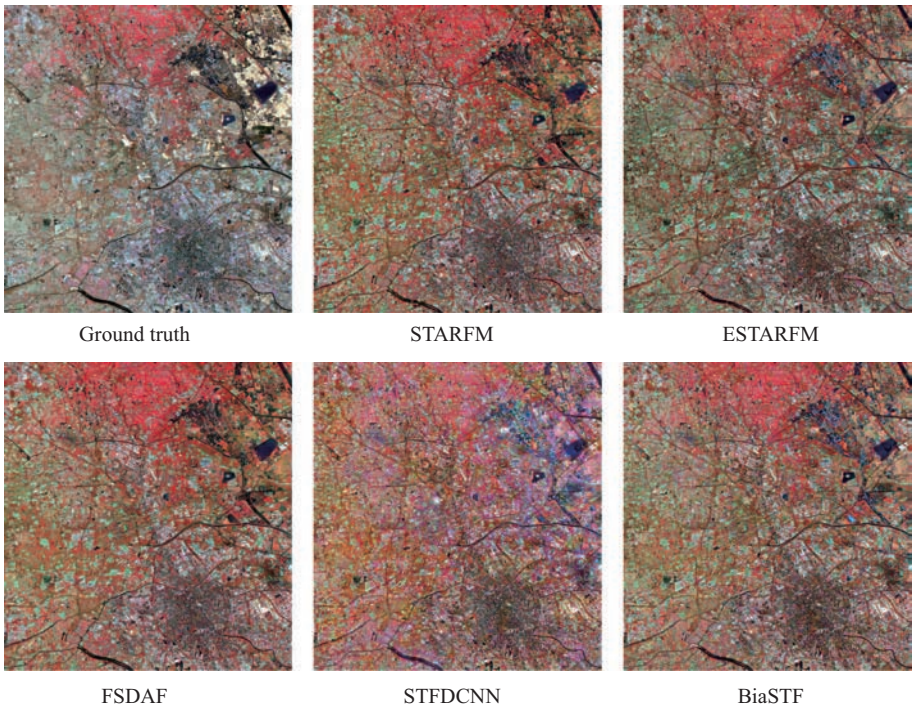

Figure 8 (Color online) Ground truth image and obtained predictions for the Tianjin dataset.

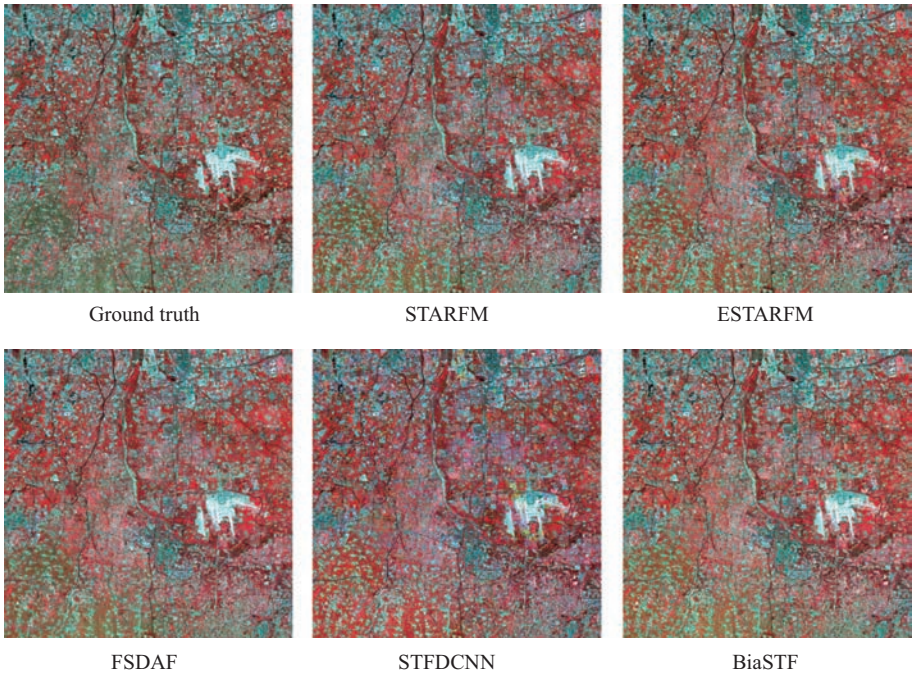

Figure 9 (Color online) Ground truth image and obtained predictions for the Daxing dataset.

than the other methods; (2) these methods are based on deep learning concepts, which are the current state-of-the-art for STF methods; (3) one of the main characteristics of our new benchmark dataset is that it offers the possibility to evaluate deep learning methods. We decide to focus on the STFDCNN and BiaSTF methods for validation purposes owing to their performance in the previous experiment. For more details, we refer to the original contribution in [65, 72]. Notice again that the parameters of the two methods are set according to [65, 72] to ensure their optimal performance. Furthermore, in all experiments, the last ten pairs are used for testing and the rest are used for training. Concerning the

Table 5 Quantitative assessment of the obtained results in the three considered benchmark datasets

	AHB dataset			Tianjin dataset			Daxing dataset		
	Pair	STFDCNN	BiaSTF	Pair	STFDCNN	BiaSTF	Pair	STFDCNN	BiaSTF
RMSE	18th	0.0270	0.0180	18th	0.0413	0.0509	20th	0.0379	0.0379
	19th	0.0316	0.0304	19th	0.0417	0.0515	21st	0.0386	0.0355
	20th	0.0325	0.0247	20th	0.0633	0.0846	22nd	0.0425	0.0456
	21st	0.0477	0.0394	21st	0.0463	0.0459	23rd	0.0333	0.0394
	22nd	0.0265	0.0237	22nd	0.0471	0.0399	24th	0.0410	0.0394
	23rd	0.0222	0.1910	23rd	0.0429	0.0409	25th	0.0258	0.0231
	24th	0.0377	0.0367	24th	0.0549	0.0910	26th	0.0283	0.0234
	25th	0.0295	0.0436	25th	0.0502	0.0497	27th	0.0315	0.0292
	26th	0.0277	0.0296	26th	0.0321	0.0264	28th	0.0289	0.0295
	27th	0.0308	0.0397	27th	0.0333	0.0355	29th	0.0463	0.0309
CC	18th	0.7455	0.8898	18th	0.7479	0.8198	20th	0.8117	0.8471
	19th	0.6883	0.8204	19th	0.6501	0.8198	21st	0.7583	0.8095
	20th	0.6489	0.8011	20th	0.4795	0.6049	22nd	0.7100	0.7285
	21st	0.5427	0.7877	21st	0.6683	0.7573	23rd	0.8179	0.8837
	22nd	0.7183	0.8159	22nd	0.7558	0.7842	24th	0.7329	0.8834
	23rd	0.6382	0.6593	23rd	0.7010	0.7935	25th	0.8456	0.8869
	24th	0.6329	0.7411	24th	0.6797	0.7616	26th	0.8486	0.8837
	25th	0.6136	0.7698	25th	0.7282	0.7889	27th	0.8406	0.8686
	26th	0.6216	0.7030	26th	0.8648	0.9096	28th	0.8653	0.8705
	27th	0.4892	0.6218	27th	0.8809	0.8552	29th	0.7276	0.8050
SSIM	18th	0.8222	0.9243	18th	0.7874	0.8106	20th	0.8433	0.8607
	19th	0.7740	0.8552	19th	0.7320	0.6545	21st	0.7868	0.8405
	20th	0.7301	0.8495	20th	0.5516	0.6373	22nd	0.7509	0.7231
	21st	0.6268	0.7552	21st	0.7126	0.7730	23rd	0.8278	0.8581
	22rd	0.8016	0.8637	22rd	0.7585	0.8086	24th	0.7603	0.8599
	23nd	0.8148	0.8372	23nd	0.7477	0.8146	25th	0.8719	0.9040
	24th	0.7753	0.8279	24th	0.7196	0.7560	26th	0.8696	0.9004
	25th	0.7584	0.7043	25th	0.7556	0.7964	27th	0.8565	0.8783
	26th	0.7781	0.7911	26th	0.8766	0.9202	28th	0.8845	0.8843
	27th	0.6853	0.6996	27th	0.8868	0.8677	29th	0.7184	0.8352
ERGAS	18th	0.8165	0.5345	18th	2.1370	2.4416	20th	1.3017	1.3017
	19th	1.1292	1.2540	19th	1.7090	1.9960	21st	1.4559	1.1097
	20th	1.2808	0.7552	20th	2.5530	1.9970	22nd	1.9421	1.8862
	21st	2.7802	2.5355	21st	2.3168	2.2868	23rd	1.5581	1.4893
	22rd	3.9073	3.7836	22rd	2.0626	1.7757	24th	1.5646	1.9691
	23nd	1.1872	1.3057	23nd	1.9832	1.5954	25th	1.1212	0.9930
	24th	1.8490	1.3963	24th	1.8920	1.7864	26th	1.3085	0.9328
	25th	1.9249	7.5260	25th	2.4567	2.1757	27th	1.8015	1.5930
	26th	1.8163	1.9451	26th	1.7705	1.4810	28th	1.5409	1.5033
	27th	2.7139	2.2231	27th	1.8223	1.6771	29th	2.3941	1.0790
SAM	18th	0.0908	0.0394	18th	0.1393	0.1345	20th	0.0825	0.0663
	19th	0.1204	0.0790	19th	0.1200	0.1210	21st	0.1343	0.1052
	20th	0.1422	0.0809	20th	0.1429	0.1645	22nd	0.1640	0.2008
	21st	0.2570	0.2461	21st	0.1843	0.1791	23rd	0.1300	0.0919
	22nd	0.2756	0.2247	22nd	0.1925	0.1422	24th	0.1266	0.0896
	23rd	0.1299	0.1265	23rd	0.1574	0.1443	25th	0.0746	0.0689
	24th	0.0941	0.0720	24th	0.1699	0.1462	26th	0.0794	0.0658
	25th	0.1864	0.3368	25th	0.1567	0.1577	27th	0.0782	0.0706
	26th	0.2056	0.2623	26th	0.1050	0.0836	28th	0.0768	0.0729
	27th	0.3030	0.2663	27th	0.1007	0.0865	29th	0.1604	0.0935

quantitative metrics, we select five widely used ones to assess the results, including the root mean square error (RMSE), the structure similarity (SSIM) [89], the correlation coefficient (CC), the erreur relative global adimensionnelle de synthese (ERGAS) [90], and the spectral angle mapper (SAM).

A quantitative evaluation of our experimental results is summarized in Table 5. Notice that the RMSE, CC and SSIM in this table refer to the average of all four bands. It can be observed that, in general, both

the STFDCNN and BiaSTF methods provide relatively good results, while in most cases the BiaSTF outperforms the STFDCNN. This is because the BiaSTF considers the sensors bias in the modeling.

5 Conclusion and future lines

In this review, we introduce a new robust benchmark dataset for the evaluation of spatio-temporal fusion (STF) algorithms. The proposed benchmark possesses three important characteristics: (1) diversity of regions, (2) long timespan, and (3) challenging scenarios, and comprises Landsat and MODIS images collected over Inner Mongolia province, Tianjin city, and Daxing district of Beijing, China, respectively. This article also provides a survey of highly representative STF techniques, along with a detailed quantitative and qualitative comparison of the performance of some of the most representative STF techniques (including traditional ones and deep learning-based ones) with our newly presented benchmark dataset (which is especially suitable for evaluation deep learning-based STF methods). Our experimental results suggest that deep learning methods that model sensors bias lead to better results in terms of STF. Our future work will be mainly focused on evaluating additional STF methods and expanding the benchmark with additional pairs covering challenging scenarios over a diversity of regions and with long timespan in the considered cases.

Acknowledgements This work was supported in part by National Natural Science Foundation of China (Grant Nos. 61771496, 61571195), National Key Research and Development Program of China (Grant No. 2017YFB0502900), and Guangdong Provincial Natural Science Foundation (Grant No. 2017A030313382). The authors would like to thank the developers of STARFM, ESTARFM, FSDAF and STFDCNN algorithms for sharing their codes.

References

1 Shen M, Tang Y, Chen J, et al. Influences of temperature and precipitation before the growing season on spring phenology in grasslands of the central and eastern Qinghai-Tibetan Plateau. Agric For Meteor, 2011, 151: 1711–1722
2 Amorós-López J, Gómez-Chova L, Alonso L, et al. Multitemporal fusion of Landsat/TM and ENVISAT/MERIS for crop monitoring. Int J Appl Earth Observation GeoInf, 2013, 23: 132–141
3 Johnson M D, Hsieh W W, Cannon A J, et al. Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods. Agric For Meteor, 2016, 218–219: 74–84
4 Liao C, Wang J, Dong T, et al. Using spatio-temporal fusion of Landsat-8 and MODIS data to derive phenology, biomass and yield estimates for corn and soybean. Sci Total Environ, 2019, 650: 1707–1721
5 Nduati E, Sofue Y, Matniyaz A, et al. Cropland mapping using fusion of multi-sensor data in a complex urban/peri-urban area. Remote Sens, 2019, 11: 207
6 Zhang M, Lin H, Wang G X, et al. Estimation of vegetation productivity using a Landsat 8 time series in a heavily urbanized area, central China. Remote Sens, 2019, 11: 133
7 Hwang T, Song C, Bolstad P V, et al. Downscaling real-time vegetation dynamics by fusing multi-temporal MODIS and Landsat NDVI in topographically complex terrain. Remote Sens Environ, 2011, 115: 2499–2512
8 Knipper K R, Kustas W P, Anderson M C, et al. Evapotranspiration estimates derived using thermal-based satellite remote sensing and data fusion for irrigation management in California vineyards. Irrig Sci, 2019, 37: 431–449
9 Pan Y Q, Shen F, Wei X D. Fusion of Landsat-8/OLI and GOCI data for hourly mapping of suspended particulate matter at high spatial resolution: a case study in the Yangtze (Changjiang) estuary. Remote Sens, 2018, 10: 158
10 Yang X, Lo C P. Using a time series of satellite imagery to detect land use and land cover changes in the Atlanta, Georgia metropolitan area. Int J Remote Sens, 2002, 23: 1775–1798
11 Heimhuber V, Tulbure M G, Broich M. Addressing spatio-temporal resolution constraints in Landsat and MODIS-based mapping of large-scale floodplain inundation dynamics. Remote Sens Environ, 2018, 211: 307–320
12 Pastick N J, Wylie B K, Wu Z T. Spatiotemporal analysis of Landsat-8 and Sentinel-2 data to support monitoring of dryland ecosystems. Remote Sens, 2018, 10: 791–806
13 Chiesi M, Battista P, Fibbi L, et al. Spatio-temporal fusion of NDVI data for simulating soil water content in heterogeneous Mediterranean areas. Eur J Remote Sens, 2019, 52: 88–95
14 Li X C, Zhou Y Y, Asrar G R, et al. Response of vegetation phenology to urbanization in the conterminous United States. Glob Change Biol, 2017, 23: 2818–2830
15 Zhu X L, Cai F Y, Tian J Q, et al. Spatiotemporal fusion of multisource remote sensing data: literature survey, taxonomy, principles, applications, and future directions. Remote Sens, 2018, 10: 527
16 Gao F, Masek J G, Schwaller M R, et al. On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance. IEEE Trans Geosci Remote Sens, 2006, 44: 2207–2218
17 Hilker T, Wulder M A, Coops N C, 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, 2009, 113: 1613–1627

18 Zhu X L, Chen J, Gao F, et al. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. Remote Sens Environ, 2010, 114: 2610–2623
19 Shen H, Wu P, Liu Y, et al. A spatial and temporal reflectance fusion model considering sensor observation differences. Int J Remote Sens, 2013, 34: 4367–4383
20 Wang P J, Gao F, Masek J G. Operational data fusion framework for building frequent Landsat-like imagery. IEEE Trans Geosci Remote Sens, 2014, 52: 7353–7365
21 Wang Q, Zhang Y, Onojeghuo A O, et al. Enhancing spatio-temporal fusion of MODIS and Landsat data by incorporating 250 m MODIS data. IEEE J Sel Top Appl Earth Observ Remote Sens, 2017, 10: 4116–4123
22 Zhao Y, Huang B, Song H. A robust adaptive spatial and temporal image fusion model for complex land surface changes. Remote Sens Environ, 2018, 208: 42–62
23 Huang B, Wang J, Song H H, et al. Generating high spatiotemporal resolution land surface temperature for urban heat island monitoring. IEEE Geosci Remote Sens Lett, 2013, 10: 1011–1015
24 Xia H, Chen Y, Li Y, et al. Combining kernel-driven and fusion-based methods to generate daily high-spatial-resolution land surface temperatures. Remote Sens Environ, 2019, 224: 259–274
25 Weng Q, Fu P, Gao F. Generating daily land surface temperature at Landsat resolution by fusing Landsat and MODIS data. Remote Sens Environ, 2014, 145: 55–67
26 Xu C, Qu J J, Hao X, et al. Downscaling of surface soil moisture retrieval by combining MODIS/Landsat and in situ measurements. Remote Sens, 2018, 10: 210
27 Houborg R, McCabe M F, Gao F. A spatio-temporal enhancement method for medium resolution LAI (STEM-LAI). Int J Appl Earth Observ GeoInf, 2016, 47: 15–29
28 Wu P, Shen H, Zhang L, et al. Integrated fusion of multi-scale polar-orbiting and geostationary satellite observations for the mapping of high spatial and temporal resolution land surface temperature. Remote Sens Environ, 2015, 156: 169–181
29 Fu D, Chen B, Wang J, et al. An improved image fusion approach based on enhanced spatial and temporal the adaptive reflectance fusion model. Remote Sens, 2013, 5: 6346–6360
30 Liu M, Liu X, Wu L, et al. A modified spatiotemporal fusion algorithm using phenological information for predicting reflectance of paddy rice in southern China. Remote Sens, 2018, 10: 772
31 Cheng Q, Liu H, Shen H, et al. A spatial and temporal nonlocal filter-based data fusion method. IEEE Trans Geosci Remote Sens, 2017, 55: 4476–4488
32 Hazaymeh K, Hassan Q K. Spatiotemporal image-fusion model for enhancing the temporal resolution of Landsat-8 surface reflectance images using MODIS images. J Appl Remote Sens, 2015, 9: 096095
33 Wang J, Huang B. A rigorously-weighted spatiotemporal fusion model with uncertainty analysis. Remote Sens, 2017, 9: 990
34 Roy D P, Ju J, Lewis P, et al. Multi-temporal MODIS-Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data. Remote Sens Environ, 2008, 112: 3112–3130
35 Wu B, Huang B, Cao K, et al. Improving spatiotemporal reflectance fusion using image inpainting and steering kernel regression techniques. Int J Remote Sens, 2017, 38: 706–727
36 Wang J, Huang B. A spatiotemporal satellite image fusion model with autoregressive error correction (AREC). Int J Remote Sens, 2018, 39: 6731–6756
37 Wang Q, Blackburn G A, Onojeghuo A O, et al. Fusion of Landsat 8 OLI and Sentinel-2 MSI data. IEEE Trans Geosci Remote Sens, 2017, 55: 3885–3899
38 Malleswara Rao J, Rao C V, Senthil Kumar A, et al. Spatiotemporal data fusion using temporal high-pass modulation and edge primitives. IEEE Trans Geosci Remote Sens, 2015, 53: 5853–5860
39 Sun Y, Zhang H, Shi W. A spatio-temporal fusion method for remote sensing data using a linear injection model and local neighbourhood information. Int J Remote Sens, 2019, 40: 2965–2985
40 Yin G, Li A, Jin H, et al. Spatiotemporal fusion through the best linear unbiased estimator to generate fine spatial resolution NDVI time series. Int J Remote Sens, 2018, 39: 3287–3305
41 Liao C, Wang J, Pritchard I, et al. A spatio-temporal data fusion model for generating NDVI time series in heterogeneous regions. Remote Sens, 2017, 9: 1125
42 Bhattarai N, Quackenbush L J, Dougherty M, et al. A simple Landsat-MODIS fusion approach for monitoring seasonal evapotranspiration at 30 m spatial resolution. Int J Remote Sens, 2015, 36: 115–143
43 Kwan C, Budavari B, Gao F, et al. A hybrid color mapping approach to fusing MODIS and Landsat images for forward prediction. Remote Sens, 2018, 10: 520
44 Zhang Y, Foody G M, Ling F, et al. Spatial-temporal fraction map fusion with multi-scale remotely sensed images. Remote Sens Environ, 2018, 213: 162–181
45 Wang Q, Atkinson P M. Spatio-temporal fusion for daily Sentinel-2 images. Remote Sens Environment, 2017, 204: S0034425717305096
46 Zhukov B, Oertel D, Lanzl F, et al. Unmixing-based multisensor multiresolution image fusion. IEEE Trans Geosci Remote Sens, 1999, 37: 1212–1226
47 Zurita-Milla R, Clevers J, Schaepman M E. Unmixing-based Landsat TM and MERIS FR data fusion. IEEE Geosci Remote Sens Lett, 2008, 5: 453–457
48 Maselli F, Rembold F. Integration of LAC and GAC NDVI data to improve vegetation monitoring in semi-arid environments. Int J Remote Sens, 2002, 23: 2475–2488
49 Niu Z. 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, 2012, 6: 063507
50 Wu M Q, Huang W, Niu Z, et al. Generating daily synthetic Landsat imagery by combining Landsat and MODIS Data. Sensors, 2015, 15: 24002–24025
51 Zhang W, Li A, Jin H, et al. An enhanced spatial and temporal data fusion model for fusing Landsat and MODIS surface reflectance to generate high temporal Landsat-like data. Remote Sens, 2013, 5: 5346–5368
52 Huang B, Zhang H. Spatio-temporal reflectance fusion via unmixing: accounting for both phenological and land-cover changes. Int J Remote Sens, 2014, 35: 6213–6233
53 Xu Y, Huang B, Xu Y Y, et al. Spatial and temporal image fusion via regularized spatial unmixing. IEEE Geosci Remote Sens Lett, 2015, 12: 1362–1366
54 Mizuochi H, Hiyama T, Ohta T, et al. Development and evaluation of a lookup-table-based approach to data fusion for seasonal wetlands monitoring: an integrated use of AMSR series, MODIS, and Landsat. Remote Sens Environ, 2017, 199: 370–388
55 Lu M, Chen J, Tang H, et al. Land cover change detection by integrating object-based data blending model of Landsat and MODIS. Remote Sens Environ, 2016, 184: 374–386
56 Rao Y, Zhu X, Chen J, et al. An improved method for producing high spatial-resolution NDVI time series datasets with multi-temporal MODIS NDVI data and Landsat TM/ETM+ images. Remote Sens, 2015, 7: 7865–7891
57 Huang B, Song H H. Spatiotemporal reflectance fusion via sparse representation. IEEE Trans Geosci Remote Sens, 2012, 50: 3707–3716
58 Song H H, Huang B. Spatiotemporal satellite image fusion through one-pair image learning. IEEE Trans Geosci Remote Sens, 2013, 51: 1883–1896
59 Li D, Li Y, Yang W, et al. An enhanced single-pair learning-based reflectance fusion algorithm with spatiotemporally extended training samples. Remote Sens, 2018, 10: 1207
60 Zhao C, Gao X, Emery W J, et al. An integrated spatio-spectral-temporal sparse representation method for fusing remote-sensing images with different resolutions. IEEE Trans Geosci Remote Sens, 2018, 56: 3358–3370
61 Wu B, Huang B, Zhang L. An error-bound-regularized sparse coding for spatiotemporal reflectance fusion. IEEE Trans Geosci Remote Sens, 2015, 53: 6791–6803
62 Wei J, Wang L, Liu P, et al. Spatiotemporal fusion of remote sensing images with structural sparsity and semi-coupled dictionary learning. Remote Sens, 2017, 9: 21
63 Wei J, Wang L, Liu P, et al. Spatiotemporal fusion of MODIS and Landsat-7 reflectance images via compressed sensing. IEEE Trans Geosci Remote Sens, 2017, 55: 7126–7139
64 Liu X, Deng C, Wang S, et al. Fast and accurate spatiotemporal fusion based upon extreme learning machine. IEEE Geosci Remote Sens Lett, 2016, 13: 2039–2043
65 Song H H, Liu Q, Wang G, et al. Spatiotemporal satellite image fusion using deep convolutional neural networks. IEEE J Sel Top Appl Earth Observations Remote Sens, 2018, 11: 821–829
66 Liu X, Deng C, Chanussot J, et al. StfNet: a two-stream convolutional neural network for spatiotemporal image fusion. IEEE Trans Geosci Remote Sens, 2019, 57: 6552–6564
67 Tan Z, Yue P, Di L, et al. Deriving high spatiotemporal remote sensing images using deep convolutional network. Remote Sens, 2018, 10: 1066
68 Tao X, Liang S, Wang D, et al. Improving satellite estimates of the fraction of absorbed photosynthetically active radiation through data integration: methodology and validation. IEEE Trans Geosci Remote Sens, 2018, 56: 2107–2118
69 Ke Y, Im J, Park S, et al. Downscaling of MODIS one kilometer evapotranspiration using Landsat-8 data and machine learning approaches. Remote Sens, 2016, 8: 215
70 Boyte S P, Wylie B K, Rigge M B, et al. Fusing MODIS with Landsat 8 data to downscale weekly normalized difference vegetation index estimates for central Great Basin rangelands, USA. GISci Remote Sens, 2018, 55: 376–399
71 Moosavi V, Talebi A, Mokhtari M H, et al. A wavelet-artificial intelligence fusion approach (WAIFA) for blending Landsat and MODIS surface temperature. Remote Sens Environ, 2015, 169: 243–254
72 Li Y F, Li J, He L, et al. A sensor-bias driven spatio-temporal fusion model based on convolutional neural networks. Sci China Inf Sci, 2020, 63: 140302
73 Xue J, Leung Y, Fung T. A Bayesian data fusion approach to spatio-temporal fusion of remotely sensed images. Remote Sens, 2017, 9: 1310
74 Li A, Bo Y, Zhu Y, et al. Blending multi-resolution satellite sea surface temperature (SST) products using Bayesian maximum entropy method. Remote Sens Environ, 2013, 135: 52–63
75 Huang B, Zhang H, Song H, et al. Unified fusion of remote-sensing imagery: generating simultaneously high-resolution synthetic spatial-temporal-spectral earth observations. Remote Sens Lett, 2013, 4: 561–569
76 Shen H, Meng X, Zhang L. An integrated framework for the spatio-temporal-spectral fusion of remote sensing images. IEEE Trans Geosci Remote Sens, 2016, 54: 7135–7148
77 Gevaert C M, García-Haro F J. A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion. Remote Sens Environ, 2015, 156: 34–44
78 Ma J, Zhang W, Marinoni A, et al. An improved spatial and temporal reflectance unmixing model to synthesize time series of Landsat-like images. Remote Sens, 2018, 10: 1388
79 Xie D, Zhang J, Zhu X, et al. An improved STARFM with help of an unmixing-based method to generate high spatial and temporal resolution remote sensing data in complex heterogeneous regions. Sensors, 2016, 16: 207
80 Zhu X L, Helmer E H, Gao F, et al. A flexible spatiotemporal method for fusing satellite images with different resolutions. Remote Sens Environ, 2016, 172: 165–177

81 Liu M, Yang W, Zhu X, et al. An improved flexible spatiotemporal data fusion (IFSDAF) method for producing high spatiotemporal resolution normalized difference vegetation index time series. Remote Sens Environ, 2019, 227: 74–89

82 Cui J, Zhang X, Luo M. Combining linear pixel unmixing and STARFM for spatiotemporal fusion of Gaofen-1 wide field of view imagery and MODIS imagery. Remote Sens, 2018, 10: 1047

83 Li X, Ling F, Foody G M, et al. 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, 2017, 196: 293–311

84 Ping B, Meng Y S, Su F Z. An enhanced linear spatio-temporal fusion method for blending Landsat and MODIS data to synthesize Landsat-Like imagery. Remote Sens, 2018, 10: 881

85 Quan J, Zhan W, Ma T, et al. An integrated model for generating hourly Landsat-like land surface temperatures over heterogeneous landscapes. Remote Sens Environ, 2018, 206: 403–423

86 Liu W, Zeng Y, Li S, et al. An improved spatiotemporal fusion approach based on multiple endmember spectral mixture analysis. Sensors, 2019, 19: 2443

87 Liao L, Song J, Wang J, et al. Bayesian method for building frequent Landsat-Like NDVI datasets by integrating MODIS and Landsat NDVI. Remote Sens, 2016, 8: 452

88 Xue J, Leung Y, Fung T. An unmixing-based Bayesian model for spatio-temporal satellite image fusion in heterogeneous landscapes. Remote Sens, 2019, 11: 324

89 Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process, 2004, 13: 600–612

90 Renza D, Martinez E, Arquero A. A new approach to change detection in multispectral images by means of ERGAS index. IEEE Geosci Remote Sens Lett, 2013, 10: 76–80