






Fusing Sentinel-2 and Landsat-8 Surface Reflectance Data via Pixel-Wise Local Normalization

Yunfei Li , Graduate Student Member, IEEE, Qian Shi, Senior Member, IEEE, Lin He , Member, IEEE, Runlin Cai , Liangli Meng, Jun Li , Fellow, IEEE, and Antonio Plaza , Fellow, IEEE

Abstract—Medium spatial resolution surface reflectance image series from the combination of Landsat-8 Operational Land Imager and Sentinel-2 Multispectral Imager observations have great importance to the land surface monitoring tasks, for which great efforts have been paid for blending the two data. However, most of the efforts focus on the image series with spatial resolution of 30 m, which cannot meet the data demand of some applications. Therefore, it is necessary to fuse Landsat-8 and Sentinel-2 images to provide 10-m image series. Currently, there are three means to achieve that, including the area-to-point regression kriging fusion approach (ATPRK), spatiotemporal fusion methods, and deep-learning-based fusion models. However, the ATPRK and spatiotemporal fusion methods suffer from the limited fusion performance, while the deep-learning-based fusion models are hardware dependent, i.e., requiring the graphics processing units, which may not be satisfied sometimes. To address these issues, in this article, we develop a new pixel-wise local normalization-based fusion method (LN-FM) for fusing Sentinel-2 and Landsat-8 images. The newly proposed LN-FM is compared to the ATPRK and three representative spatiotemporal fusion methods in experiments, which use imagery collected from both a rural area and an urban area. The experimental results demonstrate that the newly developed LN-FM exhibits excellent qualitative and quantitative performance, as well as remarkable spatial, spectral, and pixel distribution fidelity. Furthermore, this approach is fast, which may improve its applicability

Index Terms—Landsat-8, local normalization, remote sensing image fusion, Sentinel-2.

Manuscript received 13 June 2022; revised 26 July 2022; accepted 18 August 2022. Date of publication 22 August 2022; date of current version 9 September 2022. This work was supported in part by the National Natural Science Foundation of China under Grant T2225019, Grant 42030111, Grant 61976234, Grant and Grant 62071184; in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2022A1515011615, Grant 2019A1515011057, and Grant 202002030402; and in part by the Guangzhou Science and Technology Program under Grant 202002030395 (Corresponding author: Yunfei Li.)

Yunfei Li, Qian Shi, Runlin Cai, and Liangli Meng are with the Guangdong Provincial Key Laboratory of Urbanization and Geo-simulation, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China (e-mail: liyf18213483@163.com; cairlin5@mail2.sysu.edu.cn; menglli5@mail2.sysu.edu.cn).

Lin He is with the School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China (e-mail: helin@scut.edu.cn).

Jun Li is with the School of Computer Science, China University of Geosciences, Wuhan 430078, China, and also with the Hubei Key Laboratory of Intelligent Geo-Information Processing, China University of Geosciences, Wuhan 430078, China (e-mail: lijuncug@cug.edu.cn).

Antonio Plaza is with the Hyperspectral Computing Laboratory Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, E-10071 Cáceres, Spain (e-mail: aplaza@unex.es).

Digital Object Identifier 10.1109/JSTARS.2022.3200713

I. INTRODUCTION

MEDIUM spatial resolution surface reflectance image series are significant for many applications, such as crop yield estimation [1], cropland mapping [2], evapotranspiration estimation [3], surface dynamic mapping [4], and soil water content analysis [5]. Currently, the two most widely used medium spatial resolution data are from the Landsat-8 Operational Land Imager (OLI) and the Sentinel-2 Multispectral Imager (MSI). The Landsat-8 is a single satellite launched in 2013 by the National Aeronautics and Space Administration (NASA) and United States Geological Survey, who has a 16-day revisit cycle. Differently, the Sentinel-2 includes two satellites, i.e., the Sentinel-2 A and Sentinel-2B, launched in 2015 and 2017, respectively, by the European Space Agency. Thus, the Sentinel-2 data have a remarkable temporal resolution of 5 days. Though the Sentinel-2 data can theoretically provide frequent global observations, cloud contaminate will lead to considerable data loss of it. Therefore, combining the Landsat-8 OLI and Sentinel-2 MSI images has been a popular way to providing the medium spatial resolution image series [6], [7], [8], [9], [10], [11], [12]. In addition, the Landsat-8 OLI and Sentinel-2 MSI images have similar band settings (displayed by the Table I) and the same geographic coordinate system [13], which increases the feasibility of blending them. Furthermore, the NASA has started a Harmonized Landsat and Sentinel-2 (HLS) project to produce a virtual constellation of OLI and MSI surface reflectance data [14], which has been adopted by many researches [15], [16], [17]. Hence, it can be said that the medium spatial resolution image series from the Landsat-8 OLI and Sentinel-2 MSI are with great importance to present efforts about land surface monitoring. Besides, the Landsat-9 with OLI has been launched in September, 2021, which means the merged image series of Landsat OLI and Sentinel-2 MSI will keep their importance within the next 10 years.

Though there have been great efforts paid for blending Landsat-8 and Sentinel-2 images, such as the HLS project of NASA, the combined image series may not satisfy some demands since most of them focus on the spatial resolution of 30 m, which means the Sentinel-2 images are upscaled to match Landsat images with the price of sacrificing their valuable spatial information [13], [18]. As a matter of fact, in some fields, such as precision agriculture [19], [20] and tree species classifications [21], [22], [23], multitemporal remote sensing data with higher spatial resolution is urgently required. Therefore, fusing Sentinel-2 and Landsat-8 images to provide the 10-m image

TABLE I
COMPARISON BETWEEN THE SPECTRAL BANDS OF LANDSAT-8 OLI AND SENTINEL-2 MSI

Sentinel-2 MSI			Lansat-8 OLI		
Band number	Wavelength (nm)	Resolution (m)	Band number	Wavelength (nm)	Resolution (m)
1 (Aerosol)	433–453	60	1 (Aerosol)	430–450	30
2 (blue)	458–523	10	2 (blue)	450–515	30
3 (green)	543–578	10	3 (green)	525–600	30
4 (red)	650–680	10	4 (red)	630–680	30
8 (NIR)	778–892	10	5 (NIR)	845–885	30
8a (narrow NIR)	848–900	20	5 (NIR)	845–885	30
11 (SWIR-1)	1565–1655	20	6 (SWIR-1)	1560–1660	30
12 (SWIR-2)	2100–2280	20	7 (SWIR-2)	2100–2300	30

series is necessary to be discussed. To achieve that, two fusion tasks should be considered, including the self-fusion of Sentinel-2 images (referring to utilize the 10-m bands to sharpen other coarser bands, such as the 20-m SWIR-1 and SWIR-2 bands) and the fusion of Sentinel-2 and Landsat images. In the literature, most Sentinel-2-related fusion works focus on the first task, i.e., self-fusion of Sentinel-2 images [7], [24], [25], [26], while the fusion of Landsat-8 and Sentinel-2 images are relatively rare. Currently, there are three means to achieve that. First, Wang et al. [13] proposed to utilize area-to-point regression kriging (ATPRK) to downscale Landsat-8 images, which achieved remarkable fusion performance, and it has been adopted in soil toxic elements determination [27]. Second, spatiotemporal fusion methods have been adopted to fuse Sentinel-2 and Landsat data. For example, three spatiotemporal fusion approaches are used to fuse the Sentinel-2 and Landsat 8 images to generate the 10-m image series to calculate soil salinity [28]. Besides, deep learning has been introduced into this fusion task. Deep learning has been widely applied in many tasks of remotely sensed image processing, such as classification [29], [30], super-resolution [31], [32], change detection [33], [34], and object detection [35], [36]. Furthermore, some graph-based deep learning models with outstanding performance for remote sensing image classification have been developed recently [37], [38], [39], [40], which indicates the applicability and great potential of deep learning in remotely sensed image processing. For the fusion of Landsat-8 and Sentinel-2 images, some deep learning-based models have been proposed. For example, an extended super-resolution convolutional neural network [18], a degradation-term constrained spatiotemporal fusion network [41], and an enhanced residual dense network [42] have been developed for this fusion, which achieved excellent fusion performance.

Even though existing approaches can provide the 10-m images series, their issues cannot be ignored. The ATPRK has been found suffering from an oversharpening problem, and it cannot maintain the pixels distribution properly [18]. The spatiotemporal fusion methods have been reported to not do well in keeping the shapes of subtle features, and they tend to result in the spectral distortions of the fused images [41] when fusing Landsat-8 and Sentinel-2 images. The deep learning-based fusion models nearly have no flaws in fusion performance. However, these models usually require graphics processing units (GPUs) to

perform its computation due to their highly computational complexity, which may limit their applications. Therefore, a new method for the fusion of Landsat-8 and Sentinel-2 images, which can address aforementioned issues, should be discussed. In this article, we develop a new approach for fusing Sentinel-2 and Landsat-8 images. This new method uses pixel-wise local normalization to transfer the fine spatial details from the Sentinel-2 images to Landsat-8 images to achieve the fusion task. In addition, it is not deep learning-based, for which it is not hardware dependent. To demonstrate its performance, we will apply the proposed method, ATPRK and three representative spatiotemporal fusion methods to carry out the fusion between Sentinel-2 and Landsat-8 images and comprehensively test and evaluate its performance.

The rest of this article is organized as follows. Section II introduces the ATPRK and the spatiotemporal fusion methods as related works. Section III describes our new approach in detail. Section IV evaluates the considered approaches via experiments on two carefully selected study areas. Section V analyzes and discusses the performance of the considered methods in detail. Finally, Section VI concludes this article.

II. RELATED WORKS

A. ATPRK Fusion

The ATPRK approach was initially proposed for downscaling the MODIS images [43], which consists of two parts: regression modeling and area-to-point kriging-based residual downscaling. Then, it was introduced into the pan-sharpening task, and achieved a remarkable performance [44], [45]. Therefore, the ATPRK can be viewed as a representative pan-sharpening approach. Recently, the ATPRK fusion approach has been utilized to conduct the self-fusion of Sentinel-2 images [24] and the fusion between Landsat-8 and Sentinel-2 images [13], in which this approach makes use of 10-m Sentinel-2 images and the 15-m Landsat-8 panchromatic band to downscale the 30-m Landsat-8 bands to 10 m via ATPRK [13]. Roughly speaking, this approach first utilizes the 15-m Landsat-8 panchromatic band to downscale the 30-m bands to 15 m. After that, the 10-m Sentinel-2 images are downscaled to 5 m by bicubic interpolation, which then helps to downscale the 15-m Landsat-8 bands to 5 m. Finally, the 5-m Landsat-8 images are upscaled

to 10 m. More details about the area-to-point regression kriging and the downscaling framework can be found in [13].

B. Spatiotemporal Fusion

Spatiotemporal fusion was developed to blend the temporally sparse fine resolution images and temporally dense coarse resolution images to provide the synthetic data with both high spatial and high temporal resolution simultaneously, which is mainly used for blending the Landsat and MODIS data [1], [2], [4], [5], [46], [47]. Many spatiotemporal fusion methods have been developed, such as the spatial and temporal adaptive reflectance fusion model (STARFM) [46], the flexible spatiotemporal data fusion method (FSDAF) [48], and the sparse representation-based spatiotemporal reflectance fusion model (SPSTFM) [49]. These methods usually require the coarse resolution image of the predicted time and at least one base image pair consisting of a fine and a coarse resolution image from the same day, as an input to predict the missing fine resolution image, for which it is theoretically not appropriate to fuse Sentinel-2 and Landsat images since Sentinel-2 and Landsat-8 images can be barely acquired on the same day. Recently, the authors in [50] proposed a strategy to conduct spatiotemporal fusion with only two images (the coarse resolution image of the predicted time and the base fine resolution image) as input. Specifically, the strategy first aggregates the base fine resolution image to generate a base coarse resolution image. After that, a linear relative radiometric correction is implemented on the coarse resolution image of the predicted time. Finally, the base fine resolution image and the coarse resolution image (as well as the radiometrically corrected coarse resolution image) are input to a spatiotemporal fusion method to generate the fine resolution image at the predicted time. In particular, the authors in [50] tested this strategy for fusing Sentinel-2 and Landsat-8 images, demonstrating its effectiveness. In this work, we adopt the same strategy to adapt classic spatiotemporal fusion methods to the fusion of Sentinel-2 and Landsat-8 images.

III. METHOD

In this section, we present the proposed pixel-wise local normalization-based fusion method (LN-FM). Our new approach consists of two main components, which are spatial details transfer and residual compensation. Spatial details transfer refers to extracting the spatial details from the Sentinel-2 images and injecting those into the Landsat-8 images to downscale them. Residual compensation is conducted to further improve the downscaled Landsat-8 images. Both the spatial details transfer and residual compensation are achieved via a new strategy, called pixel-wise local normalization. In the following, we describe the LN-FM in detail by addressing the spatial details transfer and residual compensation steps. Notice that the LN-FM conducts the same operations to each band separately, for which following description of this method is illustrated based on a single band.

A. Spatial Details Transfer

Let \mathbf{S} and $\tilde{\mathbf{S}}$ denote the 10- and 30-m Sentinel-2 images, respectively, while \mathbf{L} and $\tilde{\mathbf{L}}$ are the 10- and 30-m Landsat-8 images, respectively. Notice that the 10-m Landsat-8 images, i.e., the \mathbf{L} , are the goal of this fusion task. The spatial response relation between the \mathbf{S} and $\tilde{\mathbf{S}}$, and that of \mathbf{L} and $\tilde{\mathbf{L}}$ are assumed to be the same, which can be denoted as follow:

$$\tilde{\mathbf{S}} = G(\mathbf{S}) \quad (1)$$

and

$$\tilde{\mathbf{L}} = G(\mathbf{L}) \quad (2)$$

where the $G(\cdot)$ is the spatial response between the 10- and 30-m images. The spatial details transfer processing consists of two steps, which are spatial details extraction and injection. Here, we define the spatial details as the ratio between one pixel and its background, i.e., the sum of its surrounding pixels, for which the spatial details of the Sentinel-2 image are extracted via pixel-wise local normalization, given by

$$\mathbf{D}(x, y) = \frac{\mathbf{S}(x, y)}{\sum_{i=-s}^s \mathbf{S}(x+i, y+i)} \quad (3)$$

where \mathbf{D} denotes the spatial details extracted from \mathbf{S} . x and y are the coordinates of the horizontal and vertical axes, respectively. s controls the range of the local extent, whose size is the $2s + 1$. In this work, the s is set to 1 since $s = 1$ leads to a local range with 3×3 , which matches the $3 \times$ spatial resolution gap between the Sentinel-2 and Landsat-8 images. After that, two assumptions are established. First, we assume the relationship between pixels of 10-m images and their background is constant over time. Second, we assume the 30-m Landsat-8 images perfectly record the background information of 10-m Landsat-8 pixels. Under the two assumptions, the spatial details from (3) can be injected into the 30-m Landsat-8 images, which is the inverse process of the extraction operation

$$\mathbf{L}^t(x, y) = \mathbf{D}(x, y) \sum_{i=-s}^s \tilde{\mathbf{L}}(x+i, y+i). \quad (4)$$

Notice that the $\tilde{\mathbf{L}}$ in this equation has been downsampled by nearest neighbor interpolation, for which it has the same size as that of \mathbf{D} . At this point, we obtain a transitional downsampled Landsat-8 image, i.e., \mathbf{L}^t . However, the fact is that 30-m Landsat-8 images cannot express the background information of 10-m Landsat-8 pixels perfectly, for which the transitional downsampled Landsat-8 image is not reliable enough. To reduce the negative effect of the second assumption, we conduct a calibration processing on it. Here, we assume there is a linear relation between the \mathbf{L}^t and the \mathbf{L} , which is

$$\mathbf{L} = a\mathbf{L}^t + b. \quad (5)$$

To determine the coefficients of the linear model, i.e., the a and b in (5), the \mathbf{S} and $\tilde{\mathbf{S}}$ are utilized. Specifically, we first inject the \mathbf{D} into the $\tilde{\mathbf{S}}$ via the same way expressed by (4), which can be

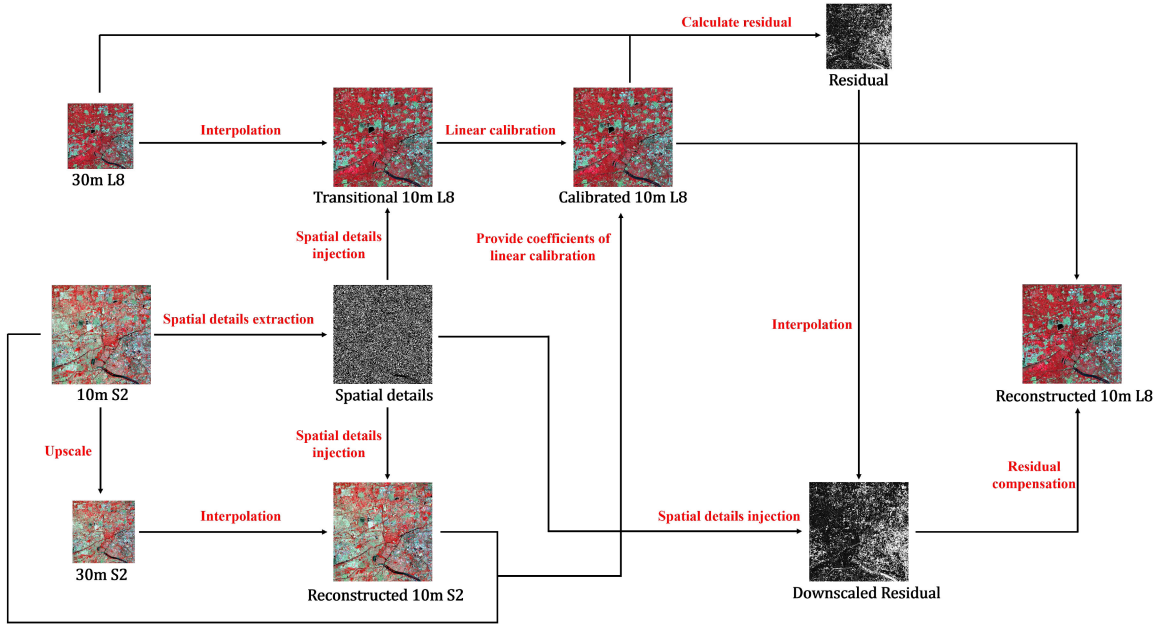


Fig. 1. Flowchart of the proposed LN-FM approach, in which S2 and L8, respectively, refer to Sentinel-2 images and Landsat-8 images.

represented as

$$\mathbf{S}^t(x, y) = \mathbf{D}(x, y) \sum_{i=-s}^s \tilde{\mathbf{S}}(x+i, y+i). \quad (6)$$

Then, a and b can be determined by the \mathbf{S}^t and the \mathbf{S} via regression

$$(\hat{a}, \hat{b}) = \arg \min_{(a, b)} \sum_{i=1}^N \|a\tilde{\mathbf{S}}_i^t + b - \mathbf{S}_i\|^2 \quad (7)$$

where N is the number of samples, i.e., the pixels, while the i refers to the i th sample. After that, the transitional downsampled Landsat-8 image (\mathbf{L}^t) can be calibrated according to (5), and the calibrated image is denoted as \mathbf{L}^c .

B. Residual Compensation

If the \mathbf{L}^c acquired via the spatial details transfer stage is perfect, the $G(\mathbf{L}^c)$ should be equal to the $\tilde{\mathbf{L}}$, which is the known 30-m Landsat-8 image. However, it is not possible due to the uncertainty of the model. To enhance the \mathbf{L}^c , the residual compensation is conducted. First, we define the difference between the $\tilde{\mathbf{L}}$ and the $G(\mathbf{L}^c)$ as the residual, which is

$$\mathbf{R}(x, y) = \tilde{\mathbf{L}}(x, y) - G(\mathbf{L}^c). \quad (8)$$

According to (8), the residual can be viewed as the wrongly downsampled part of the $\tilde{\mathbf{L}}$. Thus, we downsampled it again via introducing the spatial details from the \mathbf{S} as

$$\mathbf{R}_{\text{ds}}(x, y) = \mathbf{D}(x, y) \sum_{i=-s}^s \mathbf{R}(x+i, y+i) \quad (9)$$

where \mathbf{R}_{ds} denotes the downsampled residual. In this equation, \mathbf{R} has been downsampled by nearest neighbor interpolation, which has the same size as the \mathbf{D} . Then, the final downsampled Landsat-8

image, denoted as $\hat{\mathbf{L}}$, is produced by compensating the \mathbf{R}_{ds} to the \mathbf{L}^c as

$$\hat{\mathbf{L}}(x, y) = \mathbf{L}^c(x, y) + \mathbf{R}_{\text{ds}}(x, y). \quad (10)$$

The spatial response relation, i.e., the $G(\cdot)$, is assumed as follow in this work:

$$\tilde{\mathbf{S}}(x, y) = \frac{1}{r^2} \sum_{i=1}^{r^2} \mathbf{S}(x_i, y_i) \quad (11)$$

and

$$\tilde{\mathbf{L}}(x, y) = \frac{1}{r^2} \sum_{j=1}^{r^2} \mathbf{L}(x_j, y_j) \quad (12)$$

where x_i and y_i represent the spatial coordinates of the fine resolution images, while r is the ratio of the spatial resolution between coarse and fine resolution images, which is 3 in this task. The i and j index the fine resolution pixels within the spatial extent of the corresponding coarse resolution pixel. The flowchart of the newly proposed LN-FM is illustrated by Fig. 1

IV. EXPERIMENTS

In this section, we test our newly proposed LN-FM and other four representative fusion methods, including ATPRK [13], STARFM [46], FSDAF [48], and Fit-FC [51]. Note that the ATPRK can be viewed as a modified pan-sharpening approach, while the last three ones are spatiotemporal fusion methods, which are chosen thanks to their availability. Due to the fact that there have been many research works focused on the self-fusion of Sentinel-2 images [7], [24], [25], [26], we do not discuss this aspect in this article. Therefore, we just choose four 10-m Sentinel-2 bands, i.e., the blue, green, red, and near-infrared (NIR) in the experiments. Notice that, in the experiments, all

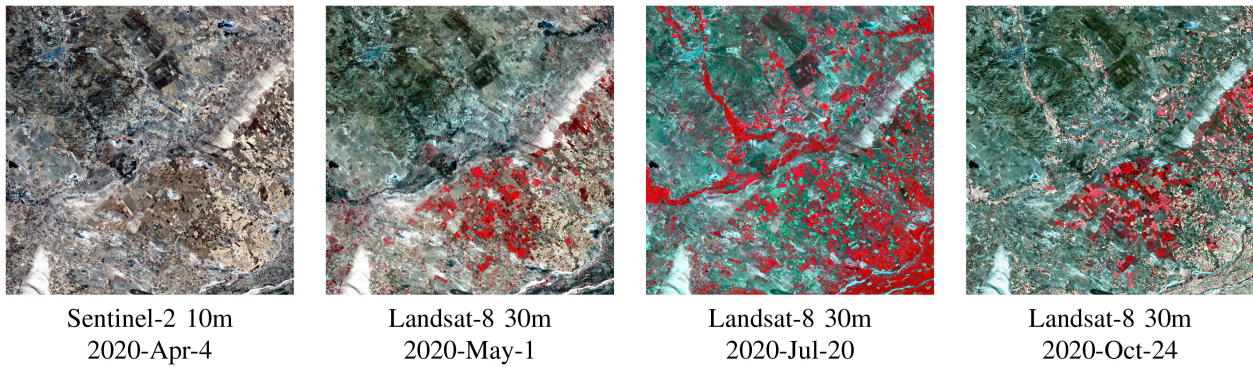


Fig. 2. Sentinel-2 and Landsat-8 images of Ar Horqin Banner, collected on 4 April 2020, 1 May 2020, 20 July 2020, and 24 October 2020, respectively. Three bands (i.e., near-infrared, red, and green) are, respectively, used as R-G-B bands for demonstration.

the considered images are $3\times$ upscaled due to the fact that there are no real 10-m Landsat-8 images that can be used as reference. Therefore, we use the 30-m Sentinel-2 images to downscale the 90-m Landsat-8 images (for ATPRK, the 45-m panchromatic band of Landsat images are also considered), and the downscaled Landsat-8 images will be assessed by comparing them with the reference images, i.e., the real 30-m Landsat-8 image. Five widely used metrics are considered in our experiments to quantitatively assess the fusion results, including root mean square error (RMSE) [52], correlation coefficient (CC) [53], structure similarity (SSIM) [54], erreur relative globale adimensionnelle de synthèse (ERGAS) [55], and spectral angle mapper (SAM) [56]. A smaller RMSE, SAM, and ERGAS (and a bigger CC and SSIM) indicate better performance of the evaluated methods.

A. Study Areas and Data

Two study areas are selected for experiments. The first one is part of Ar Horqin Banner (43.3619°N , 119.0375°E), Inner Mongolia Province, China, whose spatial extent is $63\text{ km} \times 63\text{ km}$. The Ar Horqin Banner is a small county located in northeast China, whose major industries are agriculture and animal husbandry, comprising farmlands, rangelands, forest, and mountains. A small town is also included. In other words, Ar Horqin Banner is a typical rural area. In this study area, the Sentinel-2 images were acquired on April 4, 2020, while Landsat-8 images were acquired on 1 May, 20 July, and 24 October 2020, respectively, as shown in Fig. 2. Both the Sentinel-2 and Landsat-8 images are Level-2 products, which are surface reflectance images. Differently, the panchromatic bands of the three Landsat-8 images are Level-1 C products collected at the same times (the Level-2 Landsat-8 products do not include those). The size of the Sentinel-2 image is $6300 \times 6300 \times 4$, while that of the Landsat-8 images is $2100 \times 2100 \times 4$. We can observe in Fig. 2 that this area is highly heterogeneous and significant changes occurred from one date to another.

The second study area is part of Tianjin city (43.3619°N , 119.0375°E), China, whose spatial extent is $63\text{ km} \times 63\text{ km}$, too. Different from Ar Horqin Banner, Tianjin city is a municipality located in the north China, and this area is covered by city, towns, factories, farmlands, and rivers. This is a typical

urban area. In the study area, Sentinel-2 images were acquired on 30 January 2020, while Landsat-8 images were acquired on 22 April, 24 May, and 28 August 2020, respectively, as shown in Fig. 3. Similarly, all images are Level-2 products except for the Landsat-8 panchromatic images, which are Level-1 C products. The size of the Sentinel-2 image is $6300 \times 6300 \times 4$, while that of the Landsat-8 images is $2100 \times 2100 \times 4$. It is obvious that there were significant surface changes within the study period in this area (see Fig. 3).

B. Experimental Results

1) *Overall Quantitative Evaluation*: A quantitative evaluation of the six considered fusion methods is given in Table II. Notice that the RMSE, CC, and SSIM in the table are the mean value of all four bands. First, it can be seen the newly proposed LN-FM achieves the best scores in all metrics, and the precedence of LN-FM in ERGAS and SAM is especially remarkable, which demonstrates superior accuracy of LN-FM. Then, it can be found that the FSDAF is competitive in the fusion results of Ar Horqin Banner, which outperforms the ATPRK in all three fusion results. In contrast, the quantitative performance of STRAFM and FITFC is barely satisfactory, and the STARFM achieves the smallest SAM in the result of 1 May 2020. Differently, for the Tianjin dataset, the ATPRK achieves better CC, SSIM, and ERGAS than the three spatiotemporal fusion methods, while the FSDAF is the best in terms of RMSE. Regarding SAM, ATPRK is better in the result of 22 April 2020, while FSDAF performs better in the results of 24 May and 28 August 2020. The STARFM and FITFC still underperform in those cases.

In summary, the quantitative performance of LN-FM is the best among the considered methods. In addition, the FSDAF (one of the three selected spatiotemporal fusion methods) performs remarkably well in the quantitative evaluation, which proves the potential of spatiotemporal fusion methods for fusing Sentinel-2 and Landsat-8 surface reflectance. Finally, the ATPRK exhibits a very good performance in the Tianjin area, which indicates its applicability for the fusion task.

2) *Detailed Comparison*: To further compare the performance of all considered methods, we select the fusion results

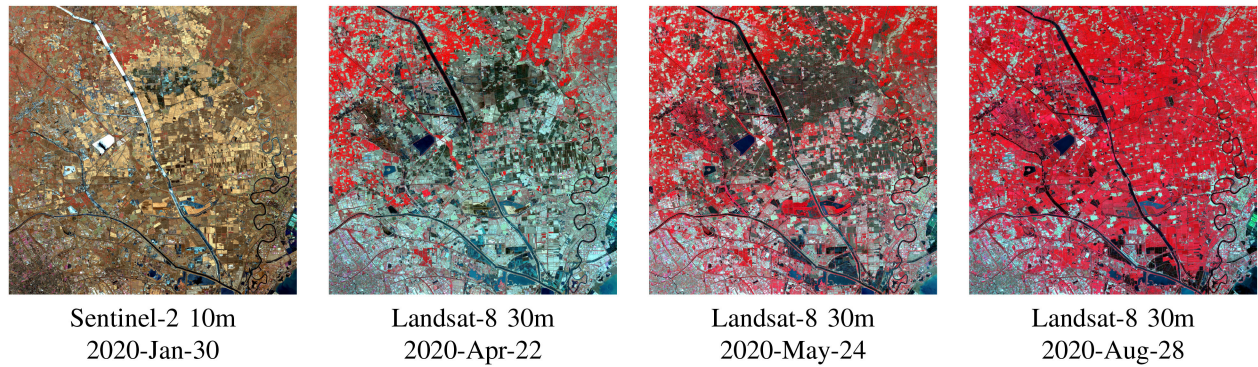


Fig. 3. Sentinel-2 and Landsat-8 images of Tianjin, collected on 30 January 2020, 22 April 2022, 24 May 2020, and 28 August 2020, respectively. Three bands (i.e., near-infrared, red, and green) are, respectively, used as R-G-B bands for demonstration.

TABLE II
QUANTITATIVE ASSESSMENT OF THE CONSIDERED FUSION METHODS

		Ar Horqin Banner					Tianjin					
	date	ATPRK	STARFM	FSDAF	Fit-FC	LN-FM	date	ATPRK	STARFM	FSDAF	Fit-FC	LN-FM
RMSE	May-1	0.0117	0.0116	0.0106	0.0117	0.0103	Apr-22	0.0244	0.0305	0.0230	0.0282	0.0220
	Jul-20	0.0170	0.0184	0.0152	0.0217	0.0136	May-24	0.0266	0.0326	0.0255	0.0309	0.0233
	Oct-24	0.0187	0.0206	0.0167	0.0232	0.0156	Aug-28	0.0264	0.0332	0.0259	0.0310	0.0241
CC	May-1	0.9450	0.9364	0.9450	0.9331	0.9486	Apr-22	0.9005	0.8083	0.8926	0.8360	0.9069
	Jul-20	0.9348	0.9085	0.9378	0.8734	0.9518	May-24	0.8899	0.7971	0.8802	0.8193	0.9051
	Oct-24	0.9120	0.8686	0.9136	0.8357	0.9247	Aug-28	0.8927	0.7808	0.8710	0.8135	0.9000
SSIM	May-1	0.9607	0.9593	0.9449	0.9566	0.9674	Apr-22	0.9183	0.8478	0.8915	0.8664	0.9264
	Jul-20	0.9480	0.9313	0.9367	0.9011	0.9511	May-24	0.9087	0.8349	0.8791	0.8497	0.9233
	Oct-24	0.9068	0.9063	0.9114	0.8776	0.9238	Aug-28	0.8888	0.8309	0.8700	0.8524	0.9221
ERGAS	May-1	0.3306	0.3258	0.3005	0.3241	0.2882	Apr-22	0.9835	1.2729	1.1852	1.1654	0.9030
	Jul-20	0.7142	0.7644	0.6378	0.8829	0.5434	May-24	1.2650	1.6080	1.2657	1.5018	1.1101
	Oct-24	0.6129	0.6679	0.5456	0.7337	0.5119	Aug-28	1.5149	1.9894	1.5642	1.7954	1.3120
SAM	May-1	0.0194	0.0138	0.0140	0.0151	0.0138	Apr-22	0.0932	0.0980	0.0934	0.0905	0.0615
	Jul-20	0.0466	0.0475	0.0400	0.0596	0.0312	May-24	0.1132	0.1281	0.1013	0.1230	0.0804
	Oct-24	0.0317	0.0283	0.0237	0.0326	0.0195	Aug-28	0.1002	0.1082	0.0890	0.1022	0.0650

of 24 October 2020 in Ar Horqin Banner and those of 24 May 2020 in Tianjin to perform a detailed qualitative and quantitative comparison. Fig. 4 shows the fusion results in 24 October 2020, Ar Horqin Banner. Notice that, in Fig. 4, every fusion result consists of two parts, i.e., the complete fused image and the zoomed subarea image, which is marked by a yellow rectangle. We can see that the five involved methods are able to recover the 30-m Landsat-8 image from the 90-m one well, due to the fact that their complete fused images are all visually similar to the ground truth, i.e., Fig. 4(b). However, the subarea reveals some differences. First, the STARFM and Fit-FC perform unsatisfactorily due to the fact that they fail to capture the slight surface change, which is marked by a yellow circle. In contrast, the ATPRK, FSDAF, and LN-FM predict this change well. Although the ATPRK recovers that change, the visual performance of its fused image is undesirable, as it suffers from oversharpening. The fused image of FSDAF appears in better visual form than that of ATPRK, which proves its powerful ability to capture the surface changes. However, although FSDAF captures the slight surface change, it cannot keep the shape of the change well. Compared to FSDAF, the LN-FM can keep the shape of surface features better, which demonstrates its excellent performance. The quantitative assessment of the fusion results is displayed in

Table III, which is basically in agreement with the analysis given by Fig. 4. Specifically, the ATPRK quantitatively outperforms STARFM and Fit-FC due to the fact that the ATPRK achieves better scores in most indices, while the FSDAF exhibits obviously better accuracy than the ATPRK. However, we can see that the LN-FM achieves the best scores in all indices except the SSIM of the blue band, which further reveals the quantitative accuracy of LN-FM. The corresponding scatter plots are given in Fig. 5. Note that the color scheme indicates point density, which increases from blue color to red color. From Fig. 5, we can conclude that the results of LN-FM are better correlated than the results provided by the other methods, which suggests that the fused image of LN-FM is the most similar to the reference Landsat-8 image collected on 24 October 2020.

The fusion results for Tianjin of 24 May 2020 are displayed in Fig. 6, from which we can see that the five considered methods can reconstruct the 30-m Landsat-8 images successfully. Similar to the fusion results obtained for the Ar Horqin Banner, the subarea tells the difference. There is a bright parcel in the subarea of the reference Landsat-8 image, which can be used to test the qualitative performance of the involved methods. First, it is obvious the STARFM and Fit-FC cannot capture that parcel, while the ATPRK, FSDAF, and LN-FM can recover it. Although

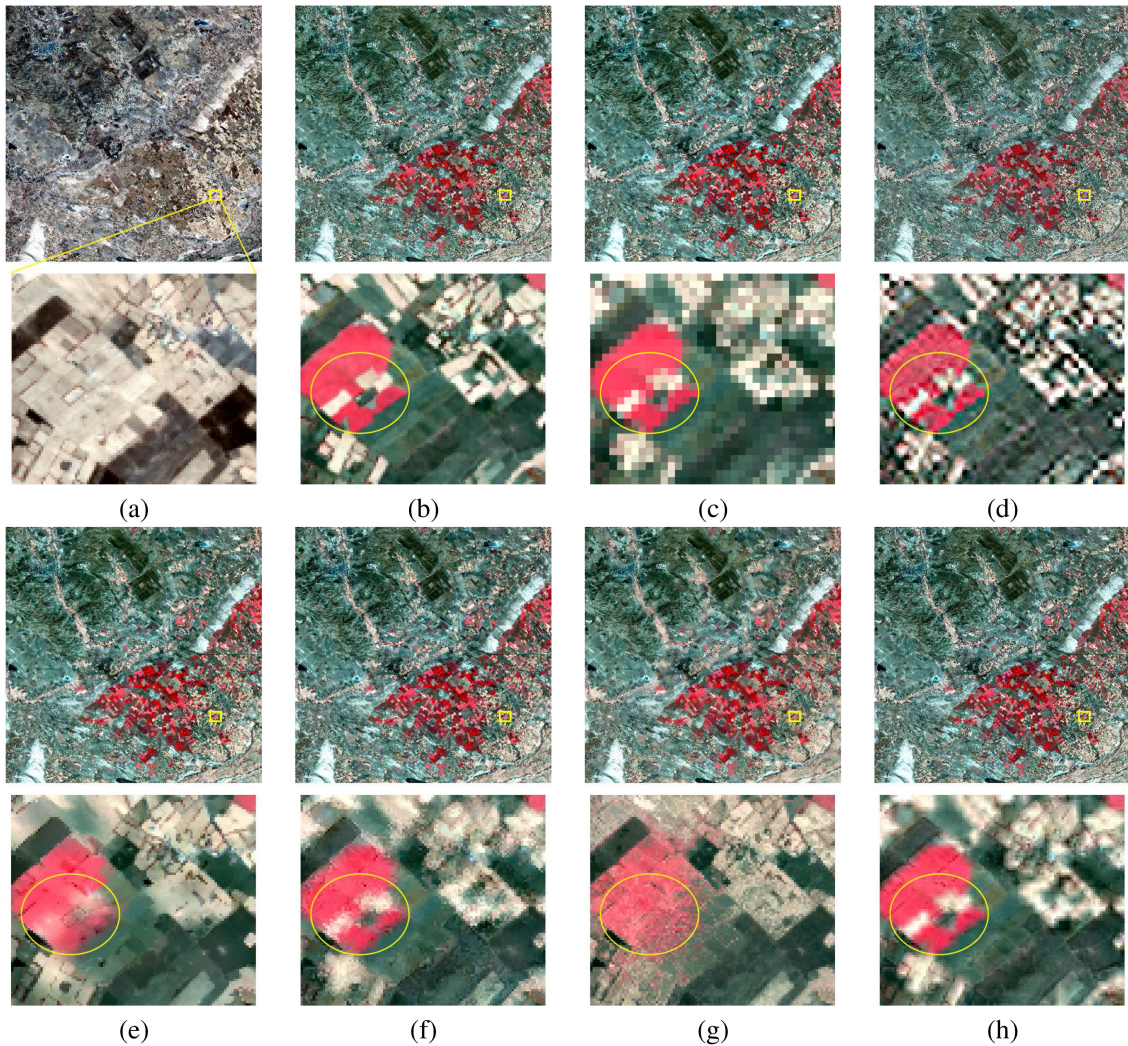


Fig. 4. Fusion results in 24 October 2020, Ar Horqin Banner. (a) 30-m Sentinel-2 image of April 4, 2020. (b) and (c) 30- and 90-m Landsat-8 image of 24 October 2020, respectively. (d)–(h) Fused 30-m Landsat-8 images by ATPRK, STARFM, FSDAF, Fit-FC, and LN-FM, respectively.

TABLE III
QUANTITATIVE ASSESSMENT OF THE AR HORQIN BANNER FUSION RESULTS OF OCTOBER 24, 2020

	RMSE				CC				SSIM				ERGAS	SAM
	Blue	Green	Red	NIR	Blue	Green	Red	NIR	Blue	Green	Red	NIR		
ATPRK	0.0107	0.0153	0.0210	0.0280	0.9027	0.9069	0.9133	0.9252	0.8973	0.9013	0.9082	0.9206	0.6129	0.0317
STARFM	0.0120	0.0163	0.0225	0.0315	0.8482	0.8679	0.8768	0.8817	0.9254	0.9106	0.8990	0.8902	0.6679	0.0283
FSDAF	0.0098	0.0133	0.0183	0.0256	0.8995	0.9130	0.9185	0.9235	0.8979	0.9110	0.9163	0.9206	0.5456	0.0237
Fit-FC	0.0124	0.0177	0.0252	0.0373	0.8320	0.8405	0.8409	0.8293	0.9168	0.8901	0.8663	0.8373	0.7337	0.0326
LN-FM	0.0092	0.0128	0.0170	0.0235	0.9125	0.9202	0.9306	0.9357	0.9117	0.9194	0.9295	0.9346	0.5119	0.0195

ATPRK and FSDAF can capture the bright parcel, unfortunately the result suffers from oversharpening and overblurring (some lines in the subarea are blurred in the fused image of FSDAF), respectively. The LN-FM can not only capture the bright parcel, but also keep the lines complete, which proves its qualitative performance again. Table IV shows a quantitative evaluation of these fusion results. Still, the ATPRK outperforms STARFM and Fit-FC. Differently, the ATPRK achieves better scores than FSDAF in three visible bands (and the ERGAS metric), while the FSDAF is better in the NIR band (and the SAM metric). The

LN-FM shows its outstanding quantitative performance again by outperforming all other methods according to all tested indices. Especially, the ERGAS and SAM of LN-FM are much smaller than those of other methods. Fig. 7 displays the scatter plots of the fusion results. Still, the scores of the proposed method are closer to the diagonal, which indicates that the fused image is the best among all considered fusion methods.

In summary, the LN-FM achieves the best qualitative and quantitative performance in all experiments. Besides, the ATPRK and FSDAF are remarkably good in terms of qualitative

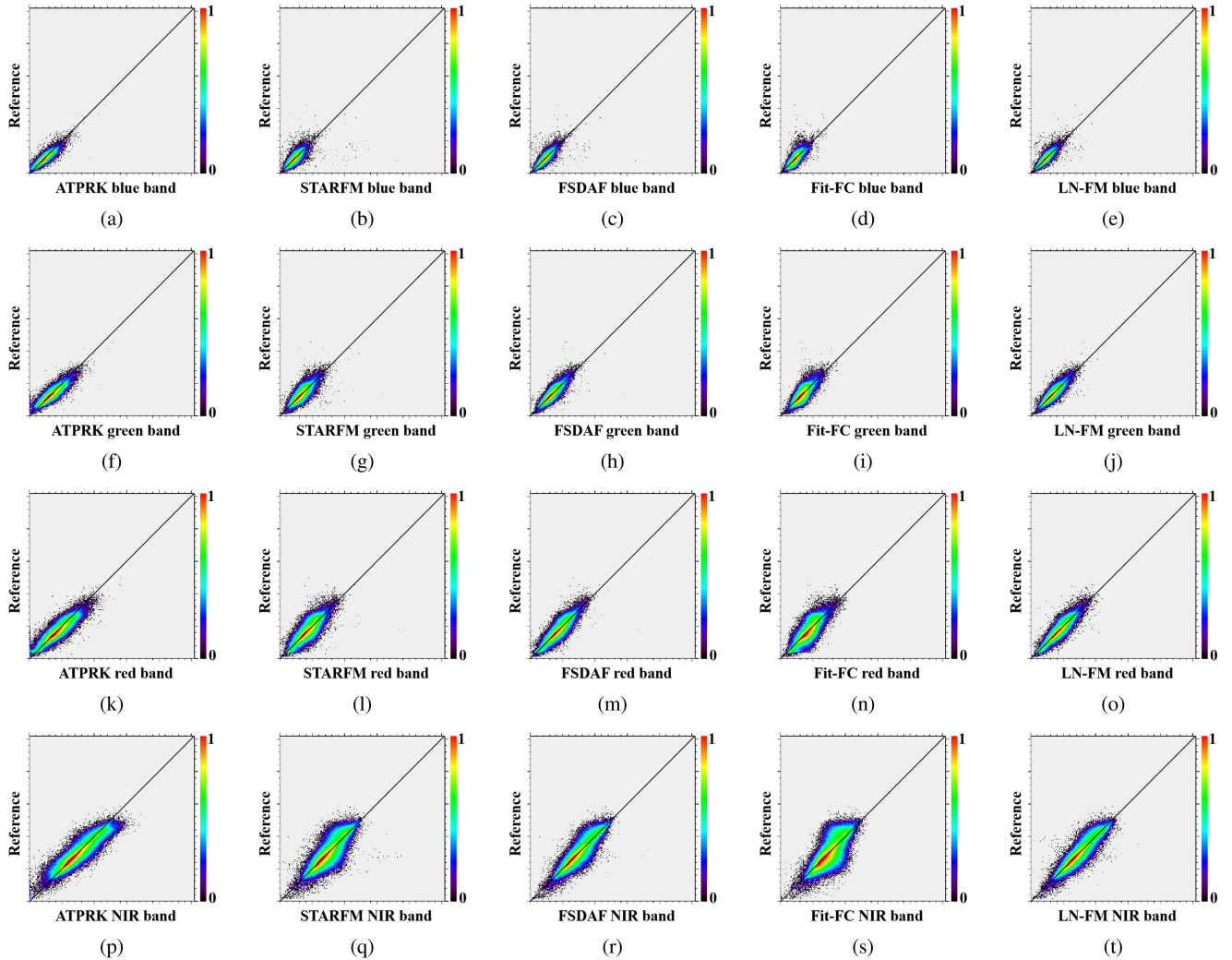


Fig. 5. Scatter plots of the fusion results of 24 October 2020, Ar Horqin Banner, compared with the reference Landsat-8 image. (a)–(e) are the scatter plots of blue bands from ATPRK, STARFM, FSDAF, Fit-FC, and LN-FM, respectively, while (f)–(j), (k)–(o), and (p)–(t) show those of the green band, the red band, and the NIR band, respectively. The color scheme indicates point density, which increases from blue color to red color.

TABLE IV
QUANTITATIVE ASSESSMENT FOR TIANJIN FUSION RESULTS OF MAY 24, 2020

	RMSE				CC				SSIM				ERGAS	SAM
	Blue	Green	Red	NIR	Blue	Green	Red	NIR	Blue	Green	Red	NIR		
ATPRK	0.0170	0.0183	0.0246	0.0465	0.8902	0.8915	0.8906	0.8875	0.9198	0.9173	0.9062	0.8915	1.2650	0.1132
STARFM	0.0222	0.0242	0.0316	0.0524	0.7865	0.7826	0.7877	0.8316	0.8496	0.8391	0.8195	0.8313	1.6080	0.1281
FSDAF	0.0174	0.0192	0.0247	0.0409	0.8725	0.8700	0.8768	0.9016	0.8717	0.8694	0.8757	0.8997	1.2657	0.1013
Fit-FC	0.0202	0.0221	0.0301	0.0512	0.8172	0.8149	0.8056	0.8396	0.8681	0.8601	0.8327	0.8380	1.5018	0.1230
LN-FM	0.0147	0.0171	0.0212	0.0402	0.9085	0.8962	0.9094	0.9064	0.9352	0.9231	0.9239	0.9110	1.1101	0.0804

and quantitative performance as well. However, the ATPRK exhibits serious oversharpening problems. Finally, STARFM and Fit-FC are barely satisfactory in terms of both qualitative and quantitative performance.

V. ANALYSIS AND DISCUSSION

In this section, we analyze and discuss issues that have not been widely explored in the previous section regarding the performance of the compared methods, including their spatial

fidelity, spectral fidelity and computational efficiency. We also discuss the parameter settings of the proposed LN-FM.

A. Spatial Fidelity Analysis

To analyze the spatial fidelity of the five considered methods, we conduct boundary extraction on the fusion results of 24 October 2020, Ar Horqin Banner (see Fig. 4), and those of 24 May 2020, Tianjin (see Fig. 6), via Laplace operator. Figs. 8 and 9 show the boundary maps obtained from the green bands

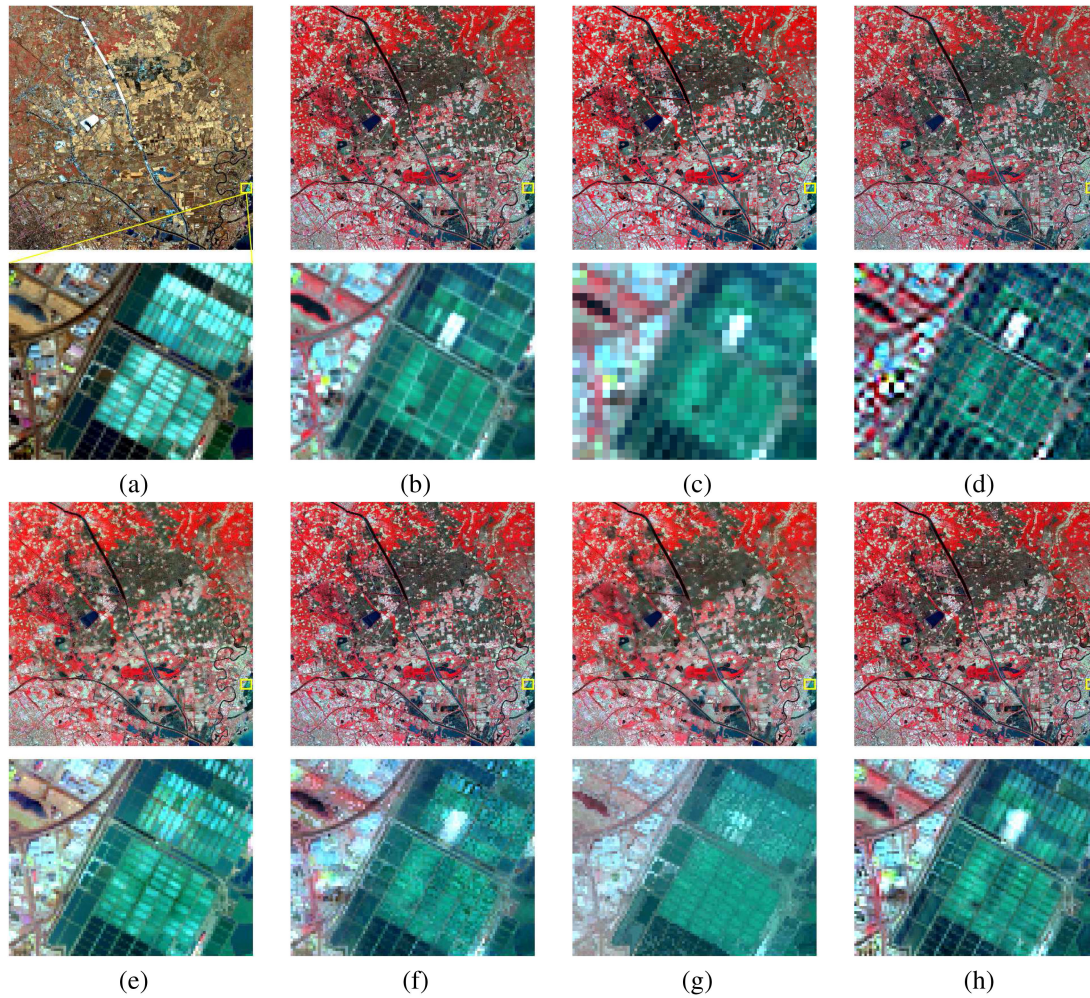


Fig. 6. Fusion results of 24 May 2020, Tianjin. (a) 30-m Sentinel-2 image of 30 January 2020. (b) and (c) 30- and 90-m Landsat-8 image of 24 May 2020, respectively. (d)–(h) Fused 30-m Landsat-8 images by ATPRK, STARFM, FSADF, Fit-FC, and LN-FM, respectively.

TABLE V
AAD OF THE BOUNDARY MAPS BETWEEN THE CONSIDERED METHODS AND THE REFERENCES

Band	Ar Horqin Banner					Tianjin				
	ATPRK	STARFM	FSADF	Fit-FC	LN-FM	ATPRK	STARFM	FSADF	Fit-FC	LN-FM
blue	0.2519	0.1539	0.1441	0.1376	0.1421	0.2815	0.2352	0.2358	0.2089	0.2027
green	0.3148	0.2443	0.2325	0.2389	0.2324	0.3063	0.2755	0.2743	0.2518	0.2516
red	0.3453	0.2977	0.2845	0.3047	0.2755	0.3380	0.3268	0.3165	0.3087	0.2853
NIR	0.3607	0.3328	0.3194	0.3514	0.3121	0.3935	0.3864	0.3733	0.3853	0.3762
Average	0.3182	0.2572	0.2451	0.2582	0.2405	0.3298	0.3060	0.3000	0.2887	0.2790

of the reference Landsat-8 images and the fused images of all involved methods, in which all results comprise the complete boundary maps and the zoomed subarea boundary maps. We can see that the boundaries of the ATPRK results are more dense than those of the reference images in Figs. 8 and 9, which reveals the oversharping problem of ATPRK. In Fig. 8, the boundary maps of STARFM, FSADF, Fit-FC, and LN-FM are visually similar. As a result, we focus on the subarea boundary maps to tell which one is better. Apparently, the boundaries of LN-FM are closer to those of the reference images. Among

the three spatiotemporal fusion methods, the boundaries of FSADF are most similar to those of the reference images. In Fig. 9, we can see that the boundaries of Fit-FC are more sparse than those of the reference image. Among STARFM, FSADF, and LN-FM, the subarea maps demonstrate that the boundaries of LN-FM are still closer to those obtained from the reference images. To quantitatively assess the boundary maps from the considered methods, we calculate the absolute average difference (AAD) between them and those from the reference, as shown in Table V. From this table, we can conclude that

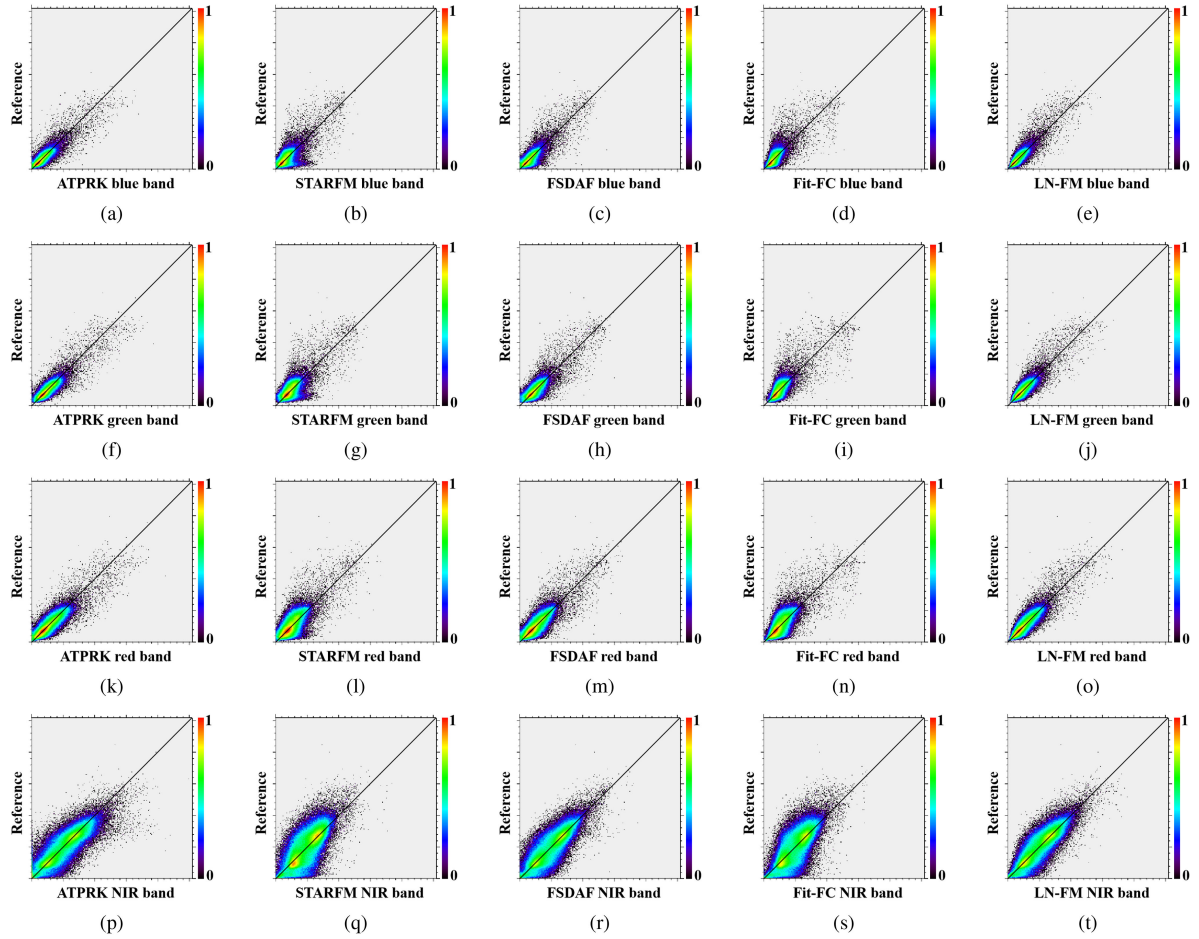


Fig. 7. Scatter plots of fusion results of 24 May 2020, Tianjin, compared with the reference Landsat-8 image. (a)–(e) are the scatter plots of blue bands from ATPRK, STARFM, FSDAF, Fit-FC, and LN-FM, respectively, while (f)–(j), (k)–(o), and (p)–(t) show those of green band, red band, and NIR band, respectively. The color scheme indicates point density, which increases from blue color to red color.

TABLE VI
QUANTITATIVE EVALUATION OF THE CLASSIFICATION AGREEMENT BETWEEN THE FUSED IMAGES OBTAINED BY THE FIVE CONSIDERED METHODS AND THE REFERENCE LANDSAT-8 IMAGES

		Ar Horqin Banner					Tianjin				
		ATPRK	STARFM	FSDAF	Fit-FC	LN-FM	ATPRK	STARFM	FSDAF	Fit-FC	LN-FM
MLR	OA	0.9056	0.8859	0.9078	0.8614	0.9225	0.7889	0.7172	0.7929	0.7465	0.8325
	Kappa	0.7656	0.6929	0.7547	0.6149	0.7940	0.6946	0.5866	0.6980	0.6277	0.7545
SVM	OA	0.9052	0.8859	0.9074	0.8625	0.9225	0.7849	0.7124	0.7842	0.7397	0.8273
	Kappa	0.7627	0.6878	0.7506	0.6137	0.7914	0.6893	0.5803	0.6860	0.6185	0.7474
RF	OA	0.8962	0.8786	0.9006	0.8554	0.9136	0.7737	0.7171	0.7741	0.7367	0.8200
	Kappa	0.7451	0.6789	0.7397	0.6063	0.7736	0.6729	0.5849	0.6700	0.6136	0.7360
KNN	OA	0.9073	0.8826	0.9043	0.8565	0.9179	0.7920	0.7297	0.7888	0.7428	0.8248
	Kappa	0.7679	0.6890	0.7490	0.6121	0.7844	0.7009	0.6072	0.6944	0.6245	0.7451

ATPRK achieves the biggest AAD in all four bands, while the LN-FM is generally the best method. Among STARFM, FSDAF, and Fit-FC, the FSDAF exhibits the smallest AAD in all bands.

In summary, we can conclude that the spatial fidelity of ATPRK is not high due to its oversharpening problem. In contrast, the LN-FM exhibits remarkably good spatial fidelity, which can preserve the boundaries well. Among STARFM, FSDAF, and Fit-FC, the FSDAF exhibits the best spatial fidelity.

B. Spectral Fidelity Analysis

To explore the spectral fidelity of the considered methods, we conduct classification on the fusion results of 24 October 2020, Ar Horqin Banner (see Fig. 4), and those of 24 May 2020, Tianjin (see Fig. 6), using four classic classifiers, including the multinomial logistic regression (MLR) [57], support vector machine (SVM) [58], random forest (RF) [59], and K-nearest neighbors (KNN) [60]. The classifiers are trained by the samples randomly

TABLE VII
ABSOLUTE DIFFERENCE OF AVERAGE AND STANDARD DEVIATION BETWEEN THE FUSED IMAGES OF THE CONSIDERED METHODS AND THE REFERENCE LANDSAT-8 IMAGES ($\times 10^{-3}$)

		Ar Horqin Banner					Tianjin				
		ATPRK	STARFM	FSDAF	Fit-FC	LN-FM	ATPRK	STARFM	FSDAF	Fit-FC	LN-FM
Blue	Average	0.0052	0.0148	0.0398	0.0279	0.0022	0.0595	1.9024	1.5638	0.1390	0.0101
	Stdev	2.6056	1.4730	1.2845	3.5295	0.8845	1.9291	2.8656	1.1968	5.5562	1.5358
Green	Average	0.0067	0.0172	0.0373	0.0613	0.0006	0.0200	2.0792	1.7955	0.0563	0.0187
	Stdev	3.8218	2.4279	2.1066	4.7185	1.3382	2.0343	3.8761	1.1679	5.7164	1.1149
Red	Average	0.0171	0.0161	0.0449	0.0825	0.0055	0.1066	2.3160	1.7335	0.0906	0.0092
	Stdev	5.2137	3.7414	3.0897	6.8094	2.2088	3.1049	6.7079	2.2112	7.6759	2.4403
NIR	Average	0.0048	0.3080	0.1525	0.3815	0.0067	0.5432	1.9804	1.6838	0.1279	0.0609
	Stdev	7.0178	6.8851	5.0674	11.1076	3.2352	5.9279	14.371	5.8407	14.936	3.2255

TABLE VIII
COMPUTING TIME COMPARISON FOR THE CONSIDERED FUSION METHODS

Ar Horqin Banner					
	ATPRK	STARFM	FSDAF	Fit-FC	LN-FM
time (seconds)	642	98	3589	2529	222
Tianjin					
	ATPRK	STARFM	FSDAF	Fit-FC	LN-FM
time (seconds)	649	92	3603	2536	219

TABLE IX
QUANTITATIVE ASSESSMENT FOR LN-FM WITH DIFFERENT VALUES OF PARAMETER s

Tianjin										
s	1	2	3	4	5	6	7	8	9	10
ERGAS	1.1101	1.2211	1.3635	1.4755	1.5616	1.6342	1.6966	1.7504	1.7987	1.8427
SAM	0.0804	0.0866	0.0973	0.1063	0.1130	0.1184	0.1230	0.1268	0.1301	0.1331
Ar Horqin Banner										
s	1	2	3	4	5	6	7	8	9	10
ERGAS	0.5119	0.5919	0.6780	0.7458	0.7996	0.8452	0.8845	0.9191	0.9504	0.9791
SAM	0.0195	0.0216	0.0246	0.0272	0.0294	0.0313	0.0330	0.0345	0.0359	0.0371

chosen from the clustered pixels of the reference Landsat-8 images, which are divided into four clusters by K-means. After the training, the classifiers are utilized to classify the fused images of the five methods, and the reference Landsat-8 images, whose agreement is measured by using the overall accuracy (OA) and Kappa coefficient. Table VI displays the OA and Kappa of the classification. It is clear that the LN-FM outperforms all other methods with all considered indices, which indicates that the spectra of the fused image via LN-FM are closer to those of the reference Landsat-8 images in both Ar Horqin Banner and Tianjin. Besides, the FSDAF and ATPRK achieve similar quantitative performance, which is in fact better than that of STARFM and Fit-FC. We can conclude that the spectral fidelity of LN-FM is the best among all the compared methods, while FSDAF and ATPRK also perform relatively well.

C. Pixel Distribution Fidelity Analysis

Figs. 10 and 11 show a pixel distribution comparison between the fused images of the considered methods and the reference

Landsat-8 images of 24 October 2020, Ar Horqin Banner and 24 May 2020, Tianjin, respectively. Table VII illustrates the absolute difference of their average and standard deviation. From Fig 10, we can conclude that the curves of STARFM, FSDAF, and LN-FM are similar to the reference curves, while those of ATPRK and Fit-FC obviously differ from the references, which means the STARFM, FSDAF, and LN-FM keep the pixel distribution better than ATPRK and Fit-FC in Ar Horqin Banner. To figure out which one is the best among STARFM, FSDAF, and LN-FM, we can pay attention to the results in Table VII, which demonstrates that the dominance of LN-FM in terms of average and standard deviation is quite remarkable. Fig. 11 shows the pixel distribution curves of Tianjin, from which we can conclude that the curves of ATPRK and FSDAF are different from reference ones, which indicates that the ATPRK, FSDAF, and Fit-FC cannot preserve the pixel distribution well in the Tianjin data. The STARFM and LN-FM have visually similar curves. However, Table VII reveals that the LN-FM outperforms STARFM. In summary, we can conclude that the LN-FM exhibits better pixel distribution fidelity than the other

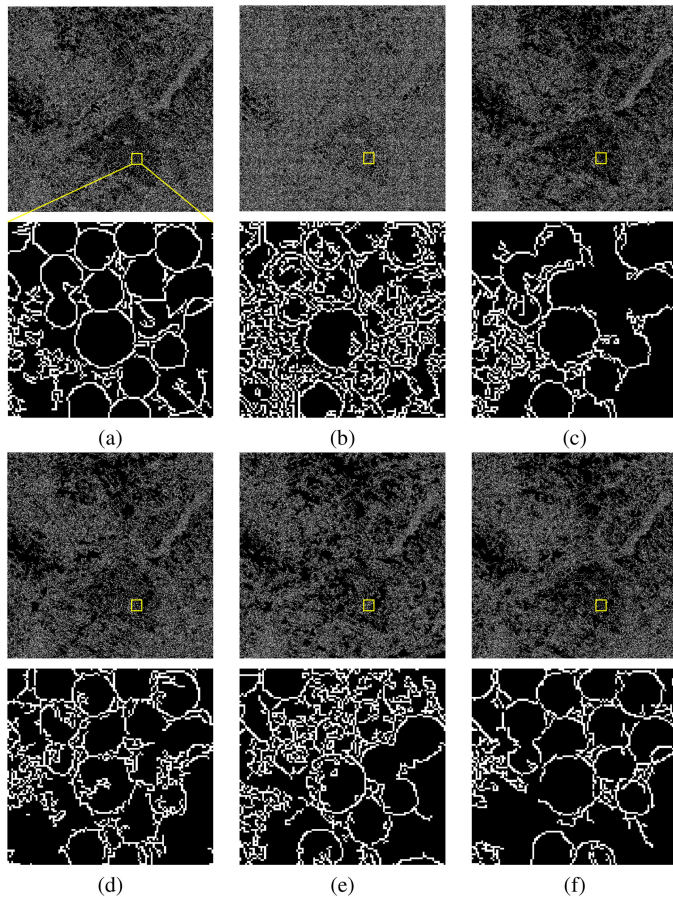


Fig. 8. Comparison of boundary maps for the fusion results of 24 October 2020, Ar Horqin Banner, where (a) is the boundary map extracted from green band of reference Landsat-8 image, while (b)–(f) are the boundary maps extracted from the fused images of ATPRK, STARFM, FSDAF, Fit-FC, and LN-FM, respectively.

four methods, as shown in Figs. 10 and 11 and Table VII. Besides, the STARFM is the best among three spatiotemporal fusion methods, and also outperforms ATPRK. Finally, the ATPRK cannot perform satisfactorily in terms of pixel distribution fidelity, which is coincident to the conclusion of [18].

D. Evaluation of Computational Complexity

Table VIII shows the average computing time of all considered methods in Ar Horqin Banner and Tianjin experiments. We can see that the faster one is STARFM, whose mean computing time is 98 s in Ar Horqin Banner and 92 s in Tianjin. Remember that the codes of STARFM and FSDAF are ENVI IDL version (<http://www.chen-lab.club/>), while those of ATPRK and Fit-FC are MATLAB version (<https://github.com/qunmingwang/>), which are all open and available. Recently, a GPU version of FSDAF has been developed [61], which is available, too. However, we do not select that under the consideration of fair comparison since other methods are all CPU based. The newly proposed LN-FM is coded by Python. All experiments are conducted on a computer with 3.60 GHz Inter(R) Core(TM) i7-9700 K CPU. In comparison, our LN-FM (which achieves

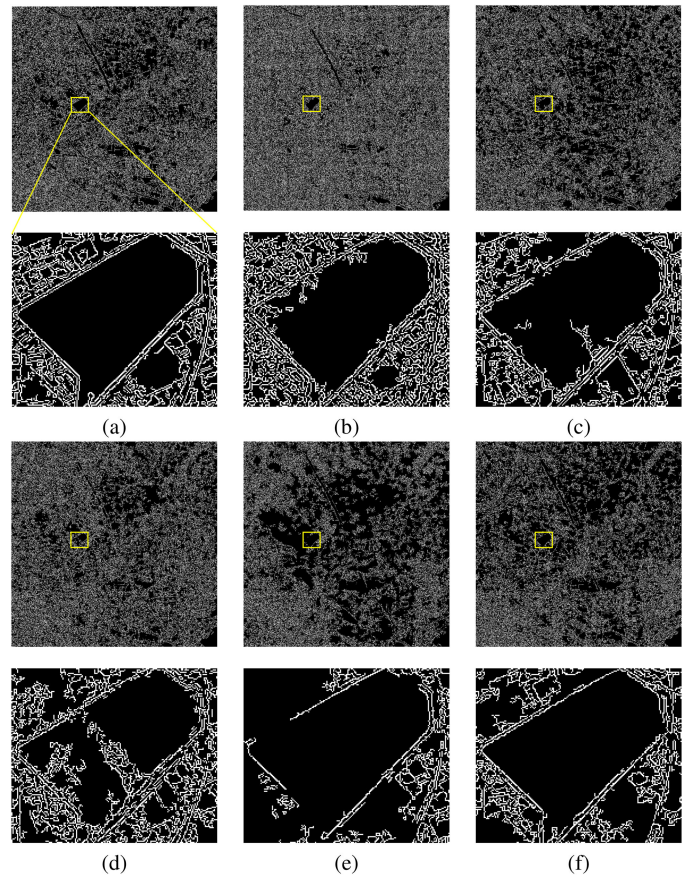


Fig. 9. Comparison of boundary maps for the fusion results of 24 May 2020, Tianjin, where (a) is the boundary map extracted from green band of reference Landsat-8 image, while (b)–(f) are the boundary maps extracted from the fused images of ATPRK, STARFM, FSDAF, Fit-FC, and LN-FM, respectively.

the best performance in experiments) takes 222 and 219 s in Ar Horqin Banner and Tianjin experiments, respectively. The ATPRK spends 642 s and 649 s on the experiments in Ar Horqin Banner and Tianjin, respectively, which is faster than FSDAF and Fit-FC. The FSDAF and Fit-FC are quite time consuming, with both spending thousands of seconds. To sum up, the ATPRK can achieve a moderate fusion performance, and its computing time is moderate too. Among the three spatiotemporal fusion methods, the FSDAF has the best fusion performance but it is time consuming. Differently, the STARFM is fast while its fusion performance is quite limited. In contrast to the other four methods, the proposed LN-FM not only achieves the best performance in the fusion of Sentinel-2 and Landsat-8 images, but also exhibits relatively low computational complexity.

E. Parameter Settings of LN-FM

There is just one parameter involved in the proposed LN-FM, which is the s . This parameter controls the extent of the local normalization, whose size is $2s + 1$. In our experiments, the s is set to 1 due to the fact that this leads to the local range of 3×3 , which matches the spatial resolution gap between Sentinel-2 and Landsat-8 images. To figure out the impact of s on the performance LN-FM, we select the Ar Horqin Banner Sentinel-2 and

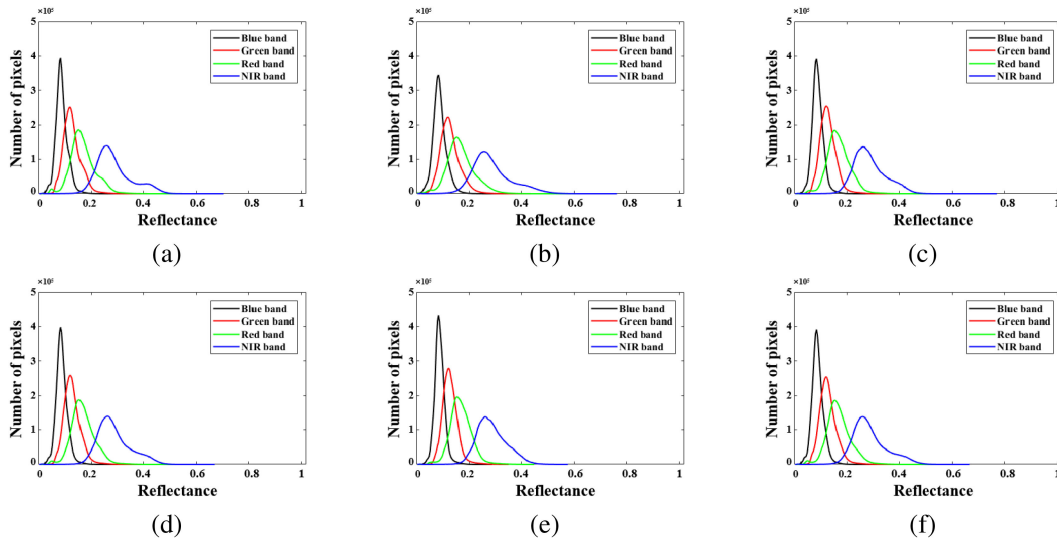


Fig. 10. Pixel distribution comparison among the fusion results of 24 October 2020, Ar Horqin Banner, where (a) is the result from the reference Landsat-8 image and (b)–(f) are the results from the fused images of ATPRK, STARFM, FSDAF, Fit-FC and LN-FM, respectively.

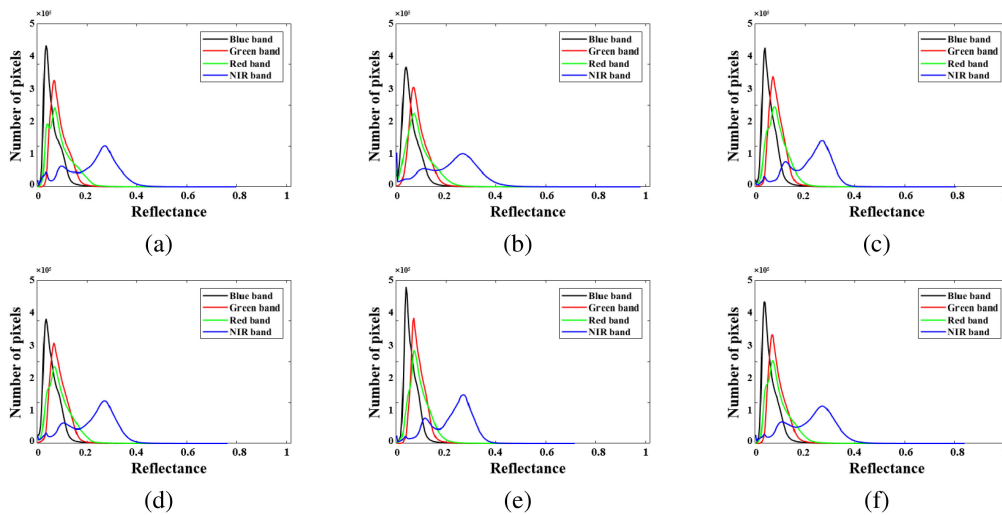


Fig. 11. Pixel distribution comparison among the fusion results of 24 May 2020, Tianjin, where (a) is the result from the reference Landsat-8 image and (b)–(f) are the results from the fused images of ATPRK, STARFM, FSDAF, Fit-FC and LN-FM, respectively.

Landsat-8 images of 4 April and 24 October 2020, respectively, and Tianjin Sentinel-2 and Landsat-8 images of 30 January and 24 May 2020, respectively, to conduct the fusion by LN-FM using different values of s . The ERGAS and SAM are chosen for quantitative evaluation. The results are displayed in Table IX, from which we can learn that the ERGAS and SAM scores both increase with the increase of s (notice that smaller ERGAS and SAM values indicate better fusion results). Therefore, the bigger the s , the worse the performance of LN-FM. These results confirm our introspection that the best value of s for LN-FM in our experiments is 1.

VI. CONCLUSION

It is feasible to fuse Landsat-8 and Sentinel-2 images to acquire the dense image series with spatial resolution of 10 m,

which is important for some applications. However, existing fusion approaches suffer from limited fusion performance (ATPRK and spatiotemporal fusion methods) or highly computational complexity (deep learning-based fusion models). To address these problems, we develop a new pixel-wise local normalization-based fusion method (LN-FM) for fusing Sentinel-2 and Landsat-8 surface reflectance in this study. For the purpose of testing the performance of the proposed LN-FM, the experiments are conducted on two typical study areas (a rural area and an urban area). Four representative fusion approaches, including ATPRK, STARFM, FSDAF, and Fit-FC, are selected for comparison, and five widely used metrics, including RMSE, CC, SSIM, ERGAS, and SAM, are chosen for quantitative assessment. In addition, the spatial, spectral, and pixel distribution fidelity, as well as the computation efficiency of the LN-FM are also investigated. According to the experimental results and

analysis, it can be concluded that the proposed LN-FM can not only exhibit the best qualitative and quantitative performance among all five involved methods, but also keep the spatial, spectral, and pixel distribution fidelity of fused images well. Furthermore, the LN-FM is a fast approach, which may improve its applicability. In the future, on the one hand, we plan to further improve the LN-FM method in order to adapt it to other different satellite sensors, on the other hand, we will try to apply it on the cloud computing platform of remote sensing data, such as the Google Earth Engine.

REFERENCES

- [1] C. Liao, J. Wang, T. Dong, J. Shang, J. Liu, and Y. Song, "Using spatio-temporal fusion of Landsat-8 and MODIS data to derive phenology, biomass and yield estimates for corn and soybean," *Sci. Total Environ.*, vol. 650, pp. 1707–1721, 2019.
- [2] E. Nduati, Y. Sofue, A. Matniyaz, J. Park, W. Yang, and A. Kondoh, "Cropland mapping using fusion of multi-sensor data in a complex urban/peri-urban area," *Remote Sens.*, vol. 11, no. 2, 2019, Art. no. 207.
- [3] K. R. Knipper et al., "Evapotranspiration estimates derived using thermal-based satellite remote sensing and data fusion for irrigation management in California vineyards," *Irrigation Sci.*, vol. 37, no. 3, pp. 431–449, 2019.
- [4] V. Heimhuber, M. G. Tulbure, and M. Broich, "Addressing spatio-temporal resolution constraints in Landsat and MODIS-based mapping of large-scale floodplain inundation dynamics," *Remote Sens. Environ.*, vol. 211, pp. 307–320, 2018.
- [5] C. Martaet al., "Spatio-temporal fusion of NDVI data for simulating soil water content in heterogeneous mediterranean areas," *Eur. J. Remote Sens.*, vol. 52, no. 1, pp. 88–95, 2019.
- [6] B. P. Page, L. G. Olmanson, and D. R. Mishra, "A harmonized image processing workflow using Sentinel-2/MSI and Landsat-8/OLI for mapping water clarity in optically variable lake systems," *Remote Sens. Environ.*, vol. 231, 2019, Art. no. 111284.
- [7] F. Palsson, J. Sveinsson, and M. Ulfarsson, "Sentinel-2 image fusion using a deep residual network," *Remote Sens.*, vol. 10, no. 8, 2018, Art. no. 1290.
- [8] L. Liu et al., "Mapping cropping intensity in China using time series Landsat and Sentinel-2 images and Google earth engine," *Remote Sens. Environ.*, vol. 239, 2020, Art. no. 111624.
- [9] J. Wang et al., "Estimating leaf area index and aboveground biomass of grazing pastures using Sentinel-1, Sentinel-2 and Landsat images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 154, pp. 189–201, 2019.
- [10] S. Gascoin, M. Grizonnet, M. Bouchet, G. Salgues, and O. Hagolle, "Theia snow collection: High-resolution operational snow cover maps from Sentinel-2 and Landsat-8 data," *Earth Syst. Sci. Data*, vol. 11, no. 2, pp. 493–514, 2019.
- [11] C. Schwatke, D. Scherer, and D. Dettmering, "Automated extraction of consistent time-variable water surfaces of lakes and reservoirs based on Landsat and Sentinel-2," *Remote Sens.*, vol. 11, no. 9, 2019, Art. no. 1010.
- [12] M. D. Nguyen, O. B. Villanueva, D. D. Bui, P. T. Nguyen, and L. Ribbe, "Harmonization of Landsat and Sentinel 2 for crop monitoring in drought prone areas: Case studies of Ninh Thuan (Vietnam) and Bekaa (Lebanon)," vol. 12, no. 2, 2019, Art. no. 281.
- [13] Q. Wang, G. A. Blackburn, A. O. Onojeghwo, J. Dash, and P. M. Atkinson, "Fusion of Landsat 8 OLI and Sentinel-2 MSI data," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 99, pp. 3885–3899, Jul. 2017.
- [14] M. Claverie et al., "The harmonized Landsat and Sentinel-2 surface reflectance data set," *Remote Sens. Environ.*, vol. 219, pp. 145–161, 2018.
- [15] A. Dkb, B. Jmg, C. Ekm, A. Mm, D. Le, and A. Maf, "Continental-scale land surface phenology from harmonized Landsat 8 and Sentinel-2 imagery," *Remote Sens. Environ.*, vol. 240, 2020, Art. no. 111685.
- [16] J. Per, C. Zhanzhang, M. Eli, F. Mark, and E. Lars, "A method for robust estimation of vegetation seasonality from Landsat and Sentinel-2 time series data," *Remote Sens.*, vol. 10, no. 4, 2018, Art. no. 635.
- [17] R. Shang and Z. Zhu, "Harmonizing Landsat 8 and Sentinel-2: A time-series-based reflectance adjustment approach," *Remote Sens. Environ.*, vol. 235, 2019, Art. no. 111439.
- [18] Z. Shao, J. Cai, P. Fu, L. Hu, and T. Liu, "Deep learning-based fusion of Landsat-8 and Sentinel-2 images for a harmonized surface reflectance product," *Remote Sens. Environ.*, vol. 235, 2019, Art. no. 111425.
- [19] J. Segarra, M. L. Buchailot, J. L. Araus, and S. C. Kefauver, "Remote sensing for precision agriculture: Sentinel-2 improved features and applications," *Agronomy*, vol. 10, no. 5, 2020, Art. no. 641.
- [20] A. Kayad, M. Sozzi, S. Gatto, F. Marinello, and F. Pirotti, "Monitoring within-field variability of corn yield using Sentinel-2 and machine learning techniques," *Remote Sens.*, vol. 11, no. 23, 2019, Art. no. 2873.
- [21] I. Markus, V. Francesco, and A. Clement, "First experience with Sentinel-2 data for crop and tree species classifications in central europe," *Remote Sens.*, vol. 8, no. 3, 2016, Art. no. 166.
- [22] A. Hocio and A. Lewandowska, "Mapping forest type and tree species on a regional scale using multi-temporal Sentinel-2 data," *Remote Sens.*, vol. 11, no. 8, 2019, Art. no. 929.
- [23] M. Persson, E. Lindberg, and H. Reese, "Tree species classification with multi-temporal Sentinel-2 data," *Remote Sens.*, vol. 10, no. 11, 2018, Art. no. 1794.
- [24] Q. Wang, W. Shi, Z. Li, and P. M. Atkinson, "Fusion of Sentinel-2 images," *Remote Sens. Environ.*, vol. 187, pp. 241–252, 2016.
- [25] M. Gargiulo, A. Mazza, R. Gaetano, G. Ruello, and G. Scarpa, "Fast super-resolution of 20 m Sentinel-2 bands using convolutional neural networks," *Remote Sens.*, vol. 11, no. 22, 2019, Art. no. 2635.
- [26] J. Wu, Z. He, and J. Hu, "Sentinel-2 sharpening via parallel residual network," *Remote Sens.*, vol. 12, no. 2, 2020, Art. no. 279.
- [27] V. Khosravi, A. Gholizadeh, and M. Saberioon, "Soil toxic elements determination using integration of Sentinel-2 and Landsat-8 images: Effects of fusion techniques on model performance," Mar. 2022.
- [28] L. Han et al., "Using spatiotemporal fusion algorithms to fill in potentially absent satellite images for calculating soil salinity: A feasibility study," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 111, 2022, Art. no. 102839.
- [29] W. Zhao and S. Du, "Spectral spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, Aug. 2016.
- [30] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.
- [31] J. M. Haut, M. E. Paoletti, R. Fernandez-Beltran, J. Plaza, A. Plaza, and J. Li, "Remote sensing single-image superresolution based on a deep compendium model," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 9, pp. 1432–1436, Sep. 2019.
- [32] S. Lei, Z. Shi, and Z. Zou, "Super-resolution for remote sensing images via local-global combined network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 8, pp. 1243–1247, Aug. 2017.
- [33] X. Cao, Y. Ji, L. Wang, B. Ji, L. Jiao, and J. Han, "SAR image change detection based on deep denoising and CNN," *IET Image Process.*, vol. 13, no. 9, pp. 1509–1515, 2019.
- [34] J. Liu et al., "Convolutional neural network-based transfer learning for optical aerial images change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 1, pp. 127–131, Jan. 2020.
- [35] W. Ouyang et al., "DeepID-Net: Object detection with deformable part based convolutional neural networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 7, pp. 1320–1334, Jul. 2017.
- [36] S. Zhang, G. He, H. Chen, N. Jing, and Q. Wang, "Scale adaptive proposal network for object detection in remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 6, pp. 864–868, Jun. 2019.
- [37] X. Wu, D. Hong, and J. Chanussot, "Convolutional neural networks for multimodal remote sensing data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5517010-1–5517010-10.
- [38] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021.
- [39] D. Hong et al., "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, May 2021.
- [40] B. Zhang et al., "Progress and challenges in intelligent remote sensing satellite systems," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1814–1822, 2022.
- [41] J. Wu, L. Lin, T. Li, Q. Cheng, C. Zhang, and H. Shen, "Fusing Landsat 8 and Sentinel-2 data for 10-m dense time-series imagery using a degradation-term constrained deep network," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 108, 2022, Art. no. 102738.
- [42] S. Xiong, S. Du, X. Zhang, S. Ouyang, and W. Cui, "Fusing Landsat-7, Landsat-8 and Sentinel-2 surface reflectance to generate dense time series images with 10 m spatial resolution," *Int. J. Remote Sens.*, vol. 43, no. 5, pp. 1630–1654, 2022.

- [43] Q. Wang, W. Shi, P. M. Atkinson, and Y. Zhao, "Downscaling MODIS images with area-to-point regression kriging," *Remote Sens. Environ.*, vol. 166, pp. 191–204, 2015.
- [44] Q. Wang, W. Shi, and P. M. Atkinson, "Area-to-point regression kriging for pan-sharpening," *ISPRS J. Photogrammetry Remote Sens.*, vol. 114, pp. 151–165, 2016.
- [45] Q. Wang, W. Shi, P. M. Atkinson, and E. Pardo-Iguzquiza, "A new geostatistical solution to remote sensing image downscaling," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 386–396, Jan. 2016.
- [46] F. Gao, J. G. Masek, M. R. Schwaller, and F. Hall, "On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2207–2218, Aug. 2006.
- [47] M. Zhang, H. Lin, G. Wang, H. Sun, and Y. Cai, "Estimation of vegetation productivity using a Landsat 8 time series in a heavily urbanized area, central China," *Remote Sens.*, vol. 11, no. 2, 2019, Art. no. 133.
- [48] 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.
- [49] B. Huang, "Spatiotemporal reflectance fusion via sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3707–3716, Oct. 2012.
- [50] J. Wu, Q. Cheng, H. Li, S. Li, and H. Shen, "Spatiotemporal fusion with only two remote sensing images as input," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 6206–6219, 2020.
- [51] Q. Wang and P. M. Atkinson, "Spatio-temporal fusion for daily Sentinel-2 images," *Remote Sens. Environ.*, vol. 204, pp. 31–42, 2018.
- [52] 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," *Sci. China Inf. Sci.*, vol. 63, no. 4, 2020, Art. no. 140302.
- [53] J. Li, Y. Li, L. He, J. Chen, and A. Plaza, "Spatio-temporal fusion for remote sensing data: An overview and new benchmark," *Sci. China Inf. Sci.*, vol. 63, no. 4, 2020, Art. no. 140301.
- [54] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [55] D. Renza, E. Martinez, and A. Arquero, "A new approach to change detection in multispectral images by means of ERGAS index," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 1, pp. 76–80, Jan. 2013.
- [56] B. Park, W. Windham, K. Lawrence, and D. Smith, "Contaminant classification of poultry hyperspectral imagery using a spectral angle mapper algorithm," *Biosyst. Eng.*, vol. 96, no. 3, pp. 323–333, 2007.
- [57] J. Li, A. Plaza, J. M. Bioucas-Dias, and A. Plaza, "Hyperspectral image segmentation using a new Bayesian approach with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3947–3960, Oct. 2011.
- [58] G. M. Foody and A. Mathur, "Toward intelligent training of supervised image classifications: Directing training data acquisition for SVM classification," *Remote Sens. Environ.*, vol. 93, no. 1–2, pp. 107–117, 2004.
- [59] J. Ham, Y. Chen, M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, Mar. 2005.
- [60] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967.
- [61] H. Gao, X. Zhu, Q. Guan, X. Yang, and X. Peng, "cuFSDAF: An enhanced flexible spatiotemporal data fusion algorithm parallelized using graphics processing units," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, no. 99, pp. 1–16, May 2021.



Yunfei Li (Graduate Student Member, IEEE) received the B.S. degree in 2018 from Jilin University, Changchun, China, and the M.S. degree in 2021 from Sun Yat-sen University, Guangzhou, China, where he is currently working toward the Ph.D. degree.

His research interests include the fusion of multi-source remote sensing data.



Qian Shi (Senior Member, IEEE) received the B.S. degree in sciences and techniques of remote sensing from Wuhan University, Wuhan, China, in 2010, and the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, in 2015.

She is currently an Associate Professor with the School of Geography and Planning, Sun Yat-sen University, Guangzhou, China. Her research interests include remote sensing image classification, including

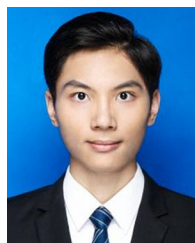
deep learning, active learning, and transfer learning.



Lin He (Member, IEEE) received the B.S. degree in instrumentation engineering from the Xi'an Institute of Technology, Xi'an, China, in 1995, the M.S. degree in instrumentation engineering from Chongqing University, Chongqing, China, in 2003, and the Ph.D. degree in pattern recognition and intelligent systems from Northwestern Poly-technical University, Xi'an, in 2007.

Since 2007, he has been with the School of Automation Science and Engineering, South China University of Technology, Guangzhou, China, where he

is currently an Associate Professor. His current research interests include statistical pattern recognition, hyperspectral image processing, and high-dimensional signal processing.



Runlin Cai received the B.S. degree in geographical science from the South China Normal University, Guangzhou, China, in 2018. He is currently working toward the M.S. degree with the School of Geography and Planning, Sun Yat-sen University, Guangzhou.

His research interests include hyperspectral image processing and geospatial data analysis.



Meng Liangli received the B.S. and M.S. degrees from Nanning Normal University, Nanning, China, in 2017 and 2020, respectively. She is currently working toward the Ph.D. degree with Sun Yat-sen University, Guangzhou, China.

Her research interests include the fusion of multi-source remote sensing data.



Jun Li (Fellow, IEEE) received the B.S. degree in geographic information systems from Hunan Normal University, Changsha, China, in 2004, the M.E. degree in remote sensing from Peking University, Beijing, China, in 2007, and the Ph.D. degree in electrical engineering from the Instituto de Telecomunicações, Instituto Superior Técnico (IST), Universidade Técnica de Lisboa, Lisbon, Portugal, in 2011.

She is currently a Full Professor with the China University of Geosciences, Wuhan China. Her main research interests include remotely sensed hyperspectral

image analysis, signal processing, supervised/semisupervised learning, and active learning.

Dr. Li is the Editor-in-Chief for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. She has been a Guest Editor for several journals, including the PROCEEDINGS OF THE IEEE and the *ISPRS Journal of Photogrammetry and Remote Sensing*.



Antonio Plaza (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in computer engineering from the University of Extremadura, Cáceres, Spain, in 1999 and 2002, respectively.

He is currently the Head of the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, University of Extremadura. He has authored more than 600 publications, including 200 JCR journal papers (145 in IEEE journals), 23 book chapters, and 285 peer-reviewed conference proceeding papers. His main research inter-

ests include hyperspectral data processing and parallel computing of remote sensing data.

Dr. Plaza has guest edited ten special issues on hyperspectral remote sensing for different journals. He is a Fellow of the IEEE for contributions to “hyperspectral data processing and parallel computing of Earth observation data.” He was the recipient of the recognition of Best Reviewers for the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS (in 2009) and IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (in 2010), for which he served as an Associate Editor in 2007–2012. He is also an Associate Editor for IEEE ACCESS, and was a member of the Editorial Board of the IEEE GEOSCIENCE AND REMOTE SENSING NEWSLETTER (2011–2012) and the IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE (2013). He was also a member of the steering committee of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. He was also the recipient of the Best Column Award of the IEEE SIGNAL PROCESSING MAGAZINE in 2015, the 2013 Best Paper Award of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the most highly cited paper (2005–2010) in the *Journal of Parallel and Distributed Computing*, and best paper awards at the IEEE International Conference on Space Technology and the IEEE Symposium on Signal Processing and Information Technology. He served as the Director of Education Activities for the IEEE Geoscience and Remote Sensing Society (GRSS) in 2011–2012, and as the President of the Spanish Chapter of IEEE GRSS in 2012–2016. He has reviewed more than 500 manuscripts for over 50 different journals. He is currently serving as the Editor-in-Chief for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING journal. Additional information: <http://www.umbc.edu/rssipl/people/aplaza>.