HISTIF: A New Spatiotemporal Image Fusion Method for High-Resolution Monitoring of Crops at the Subfield Level

Jiale Jiang, Qiaofeng Zhang, Xia Yao, *Senior Member, IEEE*, Yongchao Tian, Yan Zhu, Weixing Cao, and Tao Cheng, *Senior Member, IEEE*

***Abstract*—Satellite-based time-series crop monitoring at the subfield level is essential to the efficient implementation of precision crop management. Existing spatiotemporal image fusion techniques can be helpful, but they were often proposed to generate medium-resolution images. This study proposed a high-resolution spatiotemporal image fusion method (HISTIF) consisting of filtering for cross-scale spatial matching (FCSM) and multiplicative modulation of temporal change (MMTC). In FCSM, we considered both point spread function effect and geo-registration errors between fine and coarse resolution images. Subsequently, MMTC used pixel-based multiplicative factors to estimate the temporal change between reference and prediction dates without image classification. The performance of HISTIF was evaluated using both simulated and real datasets with one from real Gaofen-1 (GF-1) and simulated Landsat-like/Sentinel-like images, and the other from real GF-1 and real Landsat/Sentinel-2 data on two sites. HISTIF was compared with the existing methods spatial and temporal adaptive reflectance fusion model (STARFM), FSDAF, and Fit-FC. The results demonstrated that HISTIF produced substantial reduction in the fusion error from cross-scale spatial mismatch and accurate reconstruction in spatial details within fields, regardless of simulated or real data. The images predicted by STARFM exhibited pronounced blocky artifacts. While the images predicted by HISTIF and Fit-FC both showed clear within-field variability patterns, HISTIF was able to reduce the spectral distortion more significantly than Fit-FC. Furthermore, HISTIF exhibited the most stable performance across sensors. The findings suggest that HISTIF could be beneficial for the frequent and detailed monitoring of crop growth at the subfield level.**

***Index Terms*—Crops, heterogeneity, image fusion, spatiotemporal fusion, subfield monitoring.**

Manuscript received May 13, 2020; revised July 10, 2020 and August 3, 2020; accepted August 9, 2020. Date of publication August 12, 2020; date of current version August 24, 2020. This work was supported in part by the National Key R&D Program under Grant 2016YFD0300601, in part by Jiangsu Planned Projects for Postdoctoral Research Funds under Grant 2018K229C, in part by the National Natural Science Foundation of China under Grant 41871259 and Grant 31725020, in part by Jiangsu Collaborative Innovation Center for Modern Crop Production, in part by the Academic Program Development of Jiangsu Higher Education Institutions, and in part by the Qinghai Project of Transformation of Scientific and Technological Achievements under Grant 2018-NK-126. *(Corresponding author: Tao Cheng.)*

The authors are with the National Engineering and Technology Center for Information Agriculture, MOE Engineering Research Center of Smart Agriculture, MARA Key Laboratory of Crop System Analysis and Decision Making, Jiangsu Key Laboratory for Information Agriculture, Institute of Smart Agriculture, Nanjing Agricultural University, Nanjing 210095, China (e-mail: jialejiang@njau.edu.cn; 2018101161@njau.edu.cn; yaoxia@njau.edu.cn; yctian@njau.edu.cn; yanzhu@njau.edu.cn; caow@njau.edu.cn; tcheng@njau.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2020.3016135

I. Introduction

MODERN crop production is moving toward precision planting, efficient management, intelligent decision-making, and quantitative implementation at field or even subfield levels [1]–[3]. Subfield-level (within-field) crop monitoring is essential to precision farming, which requires frequent crop growth information at high-spatial resolution for mid-season management [4]. Satellite sensors such as MODIS, VEGETATION, PROBA-V, VIIRS, and Sentinel-3 provide daily observations of the globe, but their kilometric resolution is far from the requirement of subfield-level crop monitoring [5], [6]. The mid-resolution satellite imagery with weekly or bi-weekly revisit frequency, such as Sentinel-2, Landsat, and ASTER, has the potential in monitoring crop growth status over critical growth stages [5], [7], [8], but it is typically difficult to use the imagery for mapping smallholder farms [9]. Those are the main crop management units in many Asian and African countries such as India, China, and Ethiopia, where the size of a typical field ($<$1 ha) is often smaller than the pixels of the mid-resolution images [9]–[11]. To this end, high spatial resolution images ($<$5 m) are desired to overcome the issue of mixed pixels. Although high-resolution satellite imagery is increasingly available in recent years from such satellites as IKONOS, WorldView, Quickbird, SkySat, PlanetScope, and Chinese Gaofen (GF), the use of these data for mapping croplands has still been limited by the low revisit frequency or high cost in image purchasing [12]–[15]. Therefore, the images from current satellite sensors for subfield-level crop monitoring still have to compromise between spatial and temporal resolutions.

Spatiotemporal fusion approaches represent a feasible alternative for generating high spatial and temporal resolution images by blending temporally frequent but spatially coarse imagery (hereafter referred to as "coarse image") and spatially fine but temporally less frequent imagery (hereafter referred to as "fine image"). Table I summarizes a list of representative spatiotemporal fusion methods. Most fusion methods tend to predict both phenological and land cover changes. However, it is preferable to use the methods designed with specific purposes (e.g., mapping or monitoring a specific land cover type) over the generic ones [16]–[18].

For subfield-level crop monitoring, it is essential to consider the phenological change within fields rather than land cover

TABLE I
SUMMARY OF REPRESENTATIVE SPATIOTEMPORAL IMAGE FUSION METHODS

Method [a]	Quantity	Input requirements [b]		Prediction		Reference
		Image pairs	Classification	Phenological change	Land cover change	
STARFM		One or more pairs	Needed	✓	×	[19]
Semi-physical model		One pair	Not needed	✓	×	[20]
ESTARFM		Two pairs	Needed	✓	×	[21]
SPSTFM		Two pairs	Not needed	✓	✓	[3]
One-pair learning		One pair	Not needed	✓	✓	[14]
FSDAF		One pair	Needed	✓	✓	[22]
Fit-FC		One pair	Needed	✓	✓	[23]
RASTFM		One pair	Not needed	✓	✓	[24]
STAARCH	Forest disturbance mapping	A time series of coarse images and two fine images	Needed	✓	×	[17]
Regularized unmixing	Crop monitoring	A time series of fine images	Needed	✓	×	[16]
DHF	FAPAR estimation	A time series of coarse images and several fine images	Needed	✓	×	[25]
Modified ESTARFM	Paddy rice growth monitoring	A time-series of coarse images and two fine images	Needed	✓	×	[18]

[a]Note: This is not an exhaustive list of spatiotemporal image fusion methods but a representation of the most popular ones in the literature. DHF = Decametric Hectometric Fusion, ESTARFM = Enhanced STARFM, FSDAF = Flexible Spatiotemporal DAta Fusion, RASTFM = Robust Adaptive Spatial and Temporal Fusion Model, SPSTFM = Sparse-representation-based SpatioTemporal reflectance Fusion Model, STAARCH = Spatial Temporal Adaptive Algorithm for mapping Reflectance Change, STARFM = Spatial and Temporal Adaptive Reflectance Fusion Model.

[b]Since all methods require at least one coarse image at prediction time for the input data, the table above only lists the additional data required. "One pair" means the pair of coarse and fine images acquired quasi-synchronously.

change during the growth season of a specific crop. The internal structure of farmlands might accidentally change due to natural hazards (e.g., debris flow) or chronically change due to human activities (e.g., urbanization), but in those extreme cases crop growth monitoring would become land cover monitoring. In fact, crops within fields often exhibit different growth status due to variations in crop variety, sowing or planting date, soil condition, and management measure. Therefore, predicting phenological changes is the key to monitoring crop growth dynamics in a growing season, particularly the within-field variability. Although some of the existing fusion methods are considered specifically suited for capturing phenological changes, such as the widely used spatial and temporal adaptive reflectance fusion model (STARFM) [19], the semi-physical model [20], and the Enhanced STARFM (ESTARFM) [21], they do not consider the within-field variability with particular attention. Recently, Liu *et al.* [18] modified ESTARFM by further classifying the paddy rice pixels into four groups based on phenology periods from a time series of coarse images. Similar to its original version, this modified ESTARFM algorithm still depends strongly on the classification accuracy and it is affected adversely by the misclassification errors [18]. Likewise, the STARFM-like methods [22]–[26] and some spatial unmixing-based fusion methods [27], [28] are all classification-based (see Table I). They typically used unsupervised or supervised classification maps to obtain the spatial information [19], [21], [29], which leads to the low spectral variability within each class [16]. To overcome this weakness, Amorós-López *et al.* [16] used a soft clustering to provide the land-cover class proportions for crop monitoring, but their algorithm requires a time series of fine images.

Besides the problem with subfield representation, the existing methods have another intrinsic issue for subfield-level crop monitoring. They lack full consideration of the fusion error from cross-scale spatial mismatch, which is caused by geo-registration errors and point spread function (PSF) calibration errors [15], [27], [30]. Since the majority of spatiotemporal fusion methods operate at pixel level [31], they are highly sensitive to the geo-registration errors between coarse and fine images [32]. In contrast, the feature-level fusion [31] (also termed as object-oriented) methods are less sensitive to the misregistration, such as the sparse-representation-based spatiotemporal reflectance fusion model (SPSTFM) [3] and the one-pair learning model [14]. However, the feature-level methods focus on the shape change rather than the spatial details [30]. In addition to the geo-registration errors, the PSF effect of multiple sensors could also affect the exact match between coarse and fine images [33], [34], but it is often overlooked by the existing fusion methods [16], [27], [35]. It is worth considering the PSF effect from using multiple sensors because the images to be fused are often acquired from sensors with differences in band configuration, instantaneous field of view and orbit height. Since even a minor mismatch would produce a significant bias in the fusion [15], [23], [27], [30] more attention needs to be paid to both geo-registration errors and PSF calibration errors for improved spatiotemporal fusion.

In consideration of the aforementioned limitations, this study proposes a high-resolution spatiotemporal image fusion (HISTIF) model for subfield-level crop monitoring. The goal is to predict frequent high-resolution images over the field sites of major crops. The algorithm consists of two stages, including filtering for cross-scale spatial matching (FCSM) and multiplicative modulation of temporal change (MMTC). FCSM aims to correct the spatial scale mismatch between coarse and fine images by considering both misregistration and PSF effects. MMTC reconstructs subfield variability by accommodating the temporal changes pixel by pixel at fine resolution without

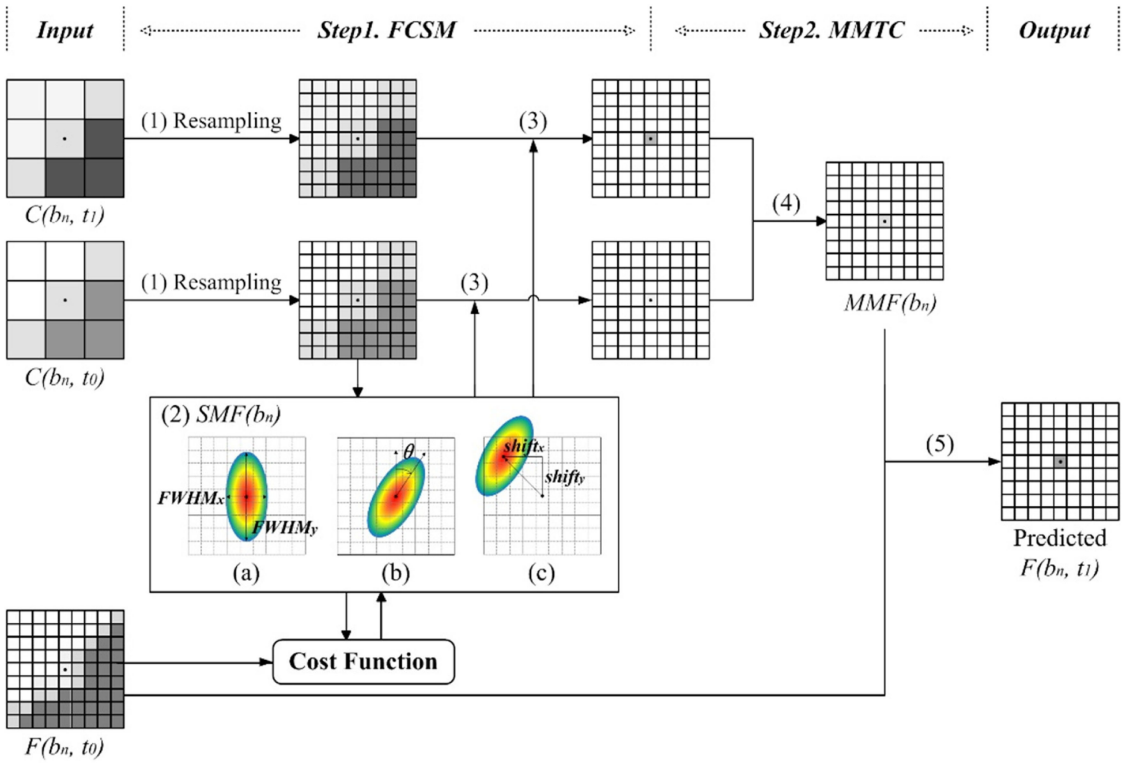

Fig. 1. Technical flowchart of the HISTIF method. (1) Coarse images are resampled to the fine resolution with the nearest neighbor interpolation. (2) Image pair at t_0 is used to determine the parameters of the cross-scale spatial matching filter (SMF) in band b_n, including (a) filter size (FWHM$_x$, FWHM$_y$), (b) rotation (θ) and (c) shift (shift$_x$, shift$_y$). These parameters are optimized through a cost function. (3) Resampled coarse images are convolved with the SMF from (2) to produce spatially reconstructed images. (4) Pixel-wise multiplicative modulation factor (MMF) is calculated from the convolved coarse images at t_0 and t_1. (5) Fine image at t_0 is multiplied by the MMF pixel by pixel to produce the fine image in band b_n at t_1.

classification. Unlike the existing specifically tailored methods, HISTIF requires minimal input data with only one fine image. We evaluated its performance using two mid- and high-resolution images datasets from different sensors and compared it with three traditional or newly published methods in the literature.

II. Methods

A. Principle and Implementation of HISTIF

The HISTIF method needs one image pair acquired at reference time t_0 and one coarse image at prediction time t_1 as input (see Fig. 1). The ultimate task of HISTIF is to predict the fine image at t_1 based on the three available images. Before the implementation of HISTIF, all input images should be calibrated to surface reflectance for monitoring temporal changes in crop status. They should be reprojected to the same coordinate system. To preserve the spectral integrity of the original images, we used the nearest neighbor interpolation to resample the coarse images to fine resolution. All of the geo-referenced images are finally cropped to the same area. Let $C(b_n, t_0)$ and $C(b_n, t_1)$ be the preprocessed coarse images in band b_n ($n = 1, 2, 3$, and 4 represents the blue, green, red, and NIR bands, respectively) at t_0 and t_1, respectively. $F(b_n, t_0)$ and $F(b_n, t_1)$ are the fine images in band b_n at t_0 and t_1, respectively. The HISTIF method consists of two steps: FCSM and MMTC.

B. Filtering for Cross-Scale Spatial Matching (FCSM)

In the nearest-neighbor resampled images, all the fine-resolution pixels that fall within a coarse-resolution pixel have the same spectral values. The solution to the blocky artifacts problem is to apply a spatial filter to the resampled images. The most commonly used filter is created by searching for spectrally similar pixels belonging to the same class in a moving window and assigning weights to the selected pixels [19]. Since this traditional filtering is based on the assumption of regular within-class change occurring from t_0 to t_1, it is highly sensitive to the classification accuracy [18] and the geo-registration error between two images [32]. As a result, it could lead to over-smooth predictions. Therefore, we proposed the FCSM not only to remove blocky artifacts, but also to correct the sensor PSF effect and geo-registration errors between coarse and fine images.

The sensor PSF is often composed of several components, including the optical PSF, the detector PSF, the image motion PSF, and the electronic PSF [36]. Given that the last two PSFs can be neglected [37], the equivalent PSF for the GF-1 instrument can be modeled as the first two PSFs (i.e., the detector PSF and the optical PSF) approximated with a 2-D Gaussian function by following the main PSF modeling strategy as described in the following [38]:

$$\mathrm{PSF}(x, y) = \frac{G(x) \cdot G(y)}{\sum G(x) \cdot G(y)} \tag{1}$$

with

$$G(x) = e^{-X^2/2\sigma_x^2},\ G(y) = e^{-Y^2/2\sigma_y^2} \tag{2}$$

where X and Y are the distances to the PSF center in East-West (x) and North-South (y) dimensions, respectively. σ_x and σ_y are the standard deviations of x and y derived from the full width at half maximum (FWHM) of the Gaussian function $G(x, y)$ as follows:

$$\sigma_x = \frac{2\sqrt{2\ln(2)}}{\mathrm{FWHM}_x},\ \sigma_y = \frac{2\sqrt{2\ln(2)}}{\mathrm{FWHM}_y}. \tag{3}$$

To account for the difference in imaging direction between the two sensors, we constructed the cross-scale spatial matching filter (SMF) by multiplying the PSF by the rotation matrix

$$\mathrm{SMF}(\mathrm{FWHM}_x, \mathrm{FWHM}_y, \theta) = \mathrm{PSF} \cdot \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4}$$

where θ is the angle (in degree) relative to the North by rotating the PSF in a counterclockwise direction around its centroid [see Fig. 1(b)].

To reduce the geo-registration errors between coarse and fine images, the SMF is shifted in both East-West and North-South dimensions as measured by shift_x and shift_y [see Fig. 1(c)]. The shifted filter is then convolved with the coarse image as follows:

$$\hat{C}_{\mathrm{SMF}}(b_n, t_0) = C(b_n, t_0) \otimes \mathrm{SMF}(b_n) \tag{5}$$

where $\hat{C}_{\mathrm{SMF}}(b_n, t_0)$ is the filtered version of the resampled image in band n at t_0. As a result of the convolution, the blocky artifacts in $C(b_n, t_0)$ could be eliminated and the cross-scale spatial mismatch between image pairs could be reduced.

The SMF could be determined by three sets of parameters, including the size (FWHM_x, FWHM_y), the rotation (θ) and the shift (shift_x, shift_y). To avoid the subjective parameterization procedures as requested in most of the existing fusion methods, the implementation of FCSM could be automated by only setting up the initial values. These parameters could be estimated by automatically optimizing the cost function defined below as the root mean square error (RMSE) between the filtered image $\hat{C}_{\mathrm{SMF}}(b_n, t_0)$ and the original fine image for each band $F(b_n, t_0)$ at t_0

$$\mathrm{RMSE} = \sqrt{\frac{\sum \left(\hat{C}_{\mathrm{SMF}}(b_n, t_0) - F(b_n, t_0)\right)^2}{N}}. \tag{6}$$

For computational efficiency purposes, the particle swarm optimization (PSO) was used to estimate the five parameters. PSO is a stochastic optimization algorithm and does not require the optimization problem be differentiable. In addition to high computational efficiency, the main advantages of the PSO algorithm also include simple concept, easy implementation, and robustness to control parameters, when compared with other mathematical algorithms and heuristic optimization techniques. For more descriptions of PSO, readers are referred to relevant references [39]–[42]. Given that there are five parameters in HISTIF, the spatial dimension of PSO should be set to 5. In this study, we set the ranges of size, shift, and rotation as 5, 20–180 m, 0–50 m, and 0°–90°, respectively. The maximum number of iterations was set to 100, but the optimization process would stop automatically if there was no significant change after 50 iterations.

Suppose the differences in PSF effect and observation geometry between two sensors are constant from t_0 to t_1, then the coarse image at t_1 ($C(b_n, t_1)$) can also be convolved with the SMF generated from the image pair at t_0 to produce the filtered image at t_1

$$\hat{C}_{\mathrm{SMF}}(b_n, t_1) = C(b_n, t_1) \otimes \mathrm{SMF}(b_n) \tag{7}$$

where $\hat{C}_{\mathrm{SMF}}(b_n, t_1)$ is the filtered coarse image in band n at t_1.

C. Multiplicative Modulation of Temporal Change (MMTC)

The second stage aims to accommodate the temporal changes in reflectance caused by crop phenology between the images at two resolutions. In order to characterize both between-class heterogeneity and within-field variability, the proposed method quantifies the temporal changes per pixel rather than per class as done in the fusion methods that need land-cover maps as the input [21]. The pixel-wise changes for band b_n in coarse images from t_0 to t_1 can be represented by the multiplicative modulation factor (MMF)

$$\mathrm{MMF}(b_n) = \frac{\hat{C}_{\mathrm{SMF}}(b_n, t_1)}{\hat{C}_{\mathrm{SMF}}(b_n, t_0)}. \tag{8}$$

A multiplicative factor was chosen instead of an additive factor, because the addition operator as used in many fusion methods would inevitably produce negative values when applying linear functions to model the spectral changes from t_0 to t_1. This way the MMF carries all the true information about temporal changes from the coarse images at two dates. Suppose the pixel-level temporal changes in reflectance are regular from t_0 to t_1 and consistent on different scales [20], then the reflectance of the fine image in band b_n at t_1 could be determined through the MMF.

According to (7) and (8), the fine image in band b_n at t_1 can be predicted from a coarse image at t_1 and one image pair at t_0 as follows:

$$F(b_n, t_1) = \frac{C(b_n, t_1) \otimes \mathrm{SMF}(b_n)}{C(b_n, t_0) \otimes \mathrm{SMF}(b_n)} \times F(b_n, t_0). \tag{9}$$

III. Experiments

A. Study Sites and Image Data

Simulated imagery represents an ideal situation without sensor differences and serves as a basic test for validating the proposed method. Real images from currently operational satellite instruments could be employed to address the complicated factors in real-world applications and to evaluate the practicability of HISTIF. Therefore, both simulated and real datasets were used in this study. The real images were acquired over two study sites, with one large agricultural area (Site 1 in Fig. 2) and one small

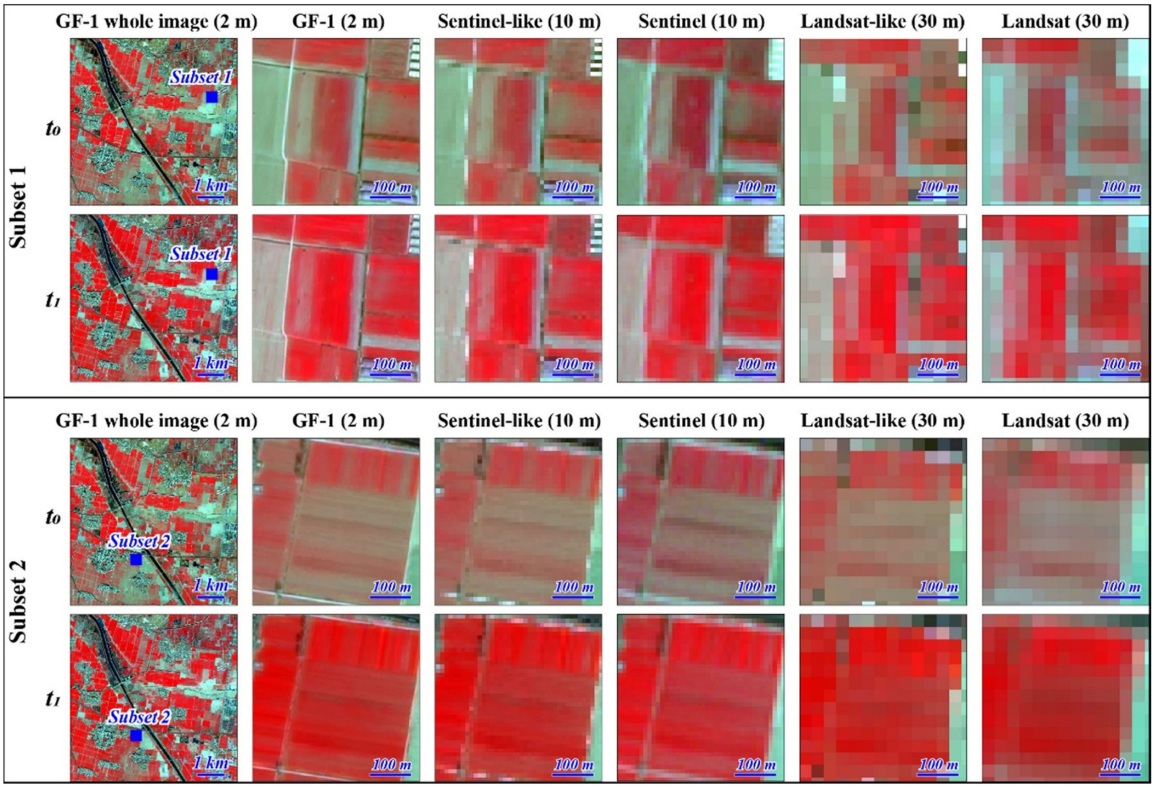

Fig. 2. Test subset images of Site 1 for t_0 and t_1 at fine and coarse resolutions. The Sentinel-like and Landsat-like reflectance images were simulated from the corresponding GF-1 data, so t_0 and t_1 of them are the same as those of GF-1 on Site 1 in Table II. All false-color images are displayed with NIR-red-green as RGB. The statistical assessments of the real GF∼Sentinel and GF∼Landsat datasets are listed in Table VIII in the Appendix.

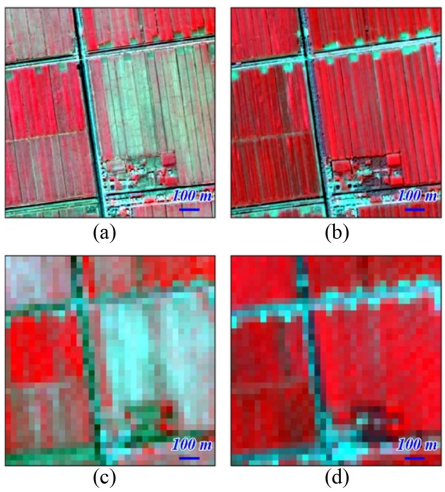

Fig. 3. Test images of Site 2 at fine and coarse resolutions for the wheat site. (a) and (b) are the GF-1 images acquired on (a) 12 March 2015 and (b) 22 April 2015, respectively. (c) and (d) are the corresponding Landsat-8 images acquired on 10 March 2015 and 27 April 2015, respectively. All false-color images are displayed with NIR-red-green as RGB. The statistical assessments of the GF∼Landsat dataset are listed in Table VIII in the Appendix.

cropland (Site 2 in Fig. 3). The selection of study sites was based on the following facts:

1) the site was mainly covered by crop fields;
2) the croplands experienced obvious changes in reflectance between the two dates due to crop growth; and
3) the spectral variations within fields (i.e., at subfield level) were significant.

The simulated dataset includes fine images from real GF-1 data at t_0 and t_1 and corresponding simulated of Sentinel-like and Landsat-like images on Site 1. Simulated coarse images were obtained by averaging 2-m GF-1 surface reflectance images to 10 m (Sentinel-like) and 30 m (Landsat-like) resolutions. As a result, the image pairs were in the same size and each coarse pixel contained corresponding GF-1 pixels completely (i.e., 5×5 and 15×15 GF-1 pixels were exactly contained by one Sentinel-like and one Landsat-like pixel, respectively).

GF-1 satellite is the first in the series of Chinese High Resolution Earth Observation System launched in April 2013. It was equipped with two panchromatic/multispectral (PMS) and four wide-field-of-view (WFV) cameras. The WFV/GF-1 was designed as high temporal resolution of 4 days but low spatial resolution of 16 m. The PMS/GF-1 instrument acquired 2 m resolution panchromatic images and 8 m resolution multispectral images in blue, green, red, and near-infrared bands, but its revisit time is 41 days and the frequency of acquiring cloud-free images could be even lower. Because of the high spatial resolution and open data policy, the images from the PMS/GF-1 (hereafter refer to GF-1) were used as the fine images in this study. The panchromatic and multispectral images were calibrated from digital number to radiance then to reflectance using the Fast Line-of-Sight Atmospheric Analysis of Spectral Hyper-cubes model. The reflectance images were geometrically corrected using the rational polynomial coefficient ortho-rectification and

TABLE II
DATA USED FOR EVALUATING HISTIF AND EXISTING METHODS

	PMS/GF-1	Sentinel-2	OLI/Landsat-8
Band (wavelength range in μm and band number in parentheses)			
Blue	0.45 - 0.52 (1)	0.46 - 0.52 (2)	0.45 - 0.51 (2)
Green	0.52 - 0.59 (2)	0.54 - 0.58 (3)	0.53 - 0.59 (3)
Red	0.63 - 0.69 (3)	0.65 - 0.68 (4)	0.64 - 0.67 (4)
NIR	0.77 - 0.89 (4)	0.78 - 0.89 (8)	0.85 - 0.88 (5)
Spatial scale			
Spatial resolution (m)	2	10	30
Scale factor		5	15
Acquisition time			
Site 1 t_0	6 February 2018	4 February 2018	5 February 2018
Site 2 t_1	11 March 2018	11 March 2018	9 March 2018
Site 2 t_0	12 March 2015		10 March 2015
Site 2 t_1	22 April 2015		27 April 2015

the associated digital elevation model file. After atmospheric and geometric corrections, the panchromatic and multispectral images were pansharpened using the Nearest Neighbor Diffusion algorithm to create 2 m resolution multispectral images. All of these operations were performed in ENVI 5.3.

The coarse images of real data were acquired from the Operational Land Imager (OLI) on Landsat 8 and the multispectral instrument on Sentinel-2A/B. The surface reflectance products of Landsat-8 OLI Level-2 Data are freely available via the United States Geological Survey and were downloaded for subsequent analysis. For the Sentinel-2 images, the product of top-of-atmosphere reflectance (Level-1C)[1] was transformed to the surface reflectance with the Sen2Cor processor released by the European Space Agency. To ensure image pairs were well co-registered, we first resampled the GF-1 images to 10 and 30 m, corresponding to the spatial resolution of Sentinel-2 and Landsat-8, respectively. Then, the coarse image was georectified to the resampled image with ground control points. For the spatiotemporal fusion of GF-1 PMS and Landsat-8 OLI images (hereafter referred to as GF∼Landsat fusion) and the fusion of GF-1 PMS and Sentinel-2 images (hereafter referred to as GF∼Sentinel fusion), this study only used the blue, green, red, and NIR bands (Bands 1, 2, 3, and 4 in GF-1 PMS; Bands 2, 3, 4, and 5 in OLI; Bands 2, 3, 4, 8 in Sentinel-2) (see Table II).

The two selected sites are located in Jiangsu province, which is part of the Yangtze-Huai Plain and one of the major crop production regions in China. The major crops in this region are winter wheat and rice, which are planted in a rotation cropping system. The Site 1 (117.53°E, 34.12°N) occupies 6×6 km^2, corresponding to 3000×3000 pixels for GF-1 images, 600 ×600 pixels for Sentinel-2 images, and 200×200 pixels for Landsat-8 images. The GF-1, Sentinel-2, and Landsat-8 images were acquired quasi-synchronously at reference time t_0 (4 to 6 February) and prediction time t_1 (9 to 11 March), roughly from tillering stage to jointing stage, in the winter wheat season of 2018 (see Table II). For visual assessment of the fusion performance, we selected two typical subset images on Site 1 (see Fig. 2). Although Sentinel-2 images can record the rough distribution of the farmland, the outline is very blurry and the real Sentinel-2 images are more blurred than the simulated Sentinel-like ones (see Fig. 2). Landsat-8 images with 30 m resolution cannot capture the spatial detail between and within fields. The Site 2 (119.094°E, 33.396°N) covers a spatial area of 1050 by 1050 m, corresponding to 525 by 525 pixels for GF-1 and 35 by 35 pixels for Landsat. The GF-1 images were acquired on March 12, 2015 (t_0) and April 22, 2015 (t_1), and the two Landsat-8 images were acquired quasi-synchronously on March 10, 2015 (t_0) and April 27, 2015 (t_1) (see Table II). According to Fig. 3, it is clear that Site 2 experienced strong phenological changes in reflectance between the early jointing and heading growth stages of winter wheat. The spectral variation within wheat fields was significant due to the difference in genetic type and sowing date, especially in March [see Fig. 3(a)]. For both sites, temporal changes were regular as observed on GF∼Sentinel and GF∼Landsat image pairs with different scale differences (see Figs. 2, 3, and 12).

[1][Online]. Available: https://scihub.copernicus.eu/

B. Comparison and Evaluation

As the major step of HISTIF, the SMF process should be evaluated for validating the stability over time. We used the image pair of real data at t_0 as the reference data to optimize the parameters of SMF. In order to verify the assumption of stable SMF performance over the time interval, we also adjusted the parameters of SMF at t_1. Furthermore, the SMFs at both t_0 and t_1 were applied to predict the fine image at t_1, respectively.

The performance of HISTIF was compared with three spatiotemporal image fusion methods, STARFM [19], flexible spatiotemporal data fusion (FSDAF) [43], and Fit-FC [29]. The STARFM is a typical fusion method for phenological change prediction and is usually used as the benchmark for evaluation of a new image fusion method [29], [30], [44]. The FSDAF is suitable for heterogeneous landscapes and mainly designed for applications related to land cover changes. The Fit-FC is a recently developed method for characterizing both phenological and land-cover changes. In addition, these three algorithms can be implemented using only three available input images (i.e., one image pair at t_0 and one coarse image at t_1) and are available in open source codes. Therefore, the three fusion algorithms were considered as reference methods and compared with the HISTIF method. All the parameters of STARFM, FSDAF, and Fit-FC were determined by following the corresponding instructions and the trial-and-error tests.

For performance evaluation, the fine images predicted by all four methods were compared visually with the true observation from GF-1 at t_1. Each band of predicted images was compared quantitatively using the correlation coefficient (CC), the RMSE, and the mean absolute difference (MAD)

$$\mathrm{CC} = \frac{\sum (x-\overline{x})\cdot(y-\overline{y})}{\sqrt{\sum (x-\overline{x})^2 \cdot \sum (y-\overline{y})^2}} \tag{10}$$

$$\mathrm{RMSE} = \sqrt{\frac{(x-y)^2}{n}} \tag{11}$$

$$\mathrm{MAD} = \frac{|x-y|}{n} \tag{12}$$

TABLE III
OPTIMAL PARAMETERS OF SMF FOR EACH BAND AT REFERENCE AND PREDICTION TIMES

Band	Time	FWHM		Shift (m) [a]		Rotation angle (°)	RMSE [b]	RMSE' [c]	ΔRMSE [d] (RMSE'-RMSE)
		$FWHM_x$	$FWHM_y$	$shift_x$	$shift_y$				
Site 1 GF~Sentinel fusion									
Blue	t_0	24	10	4 *E*	0	37	0.0236		
	t_1	26	14	4 *E*	0	37	0.0229	0.0230	0.0001
Green	t_0	20	18	4 *E*	0	37	0.0283		
	t_1	18	18	6 *E*	0	39	0.0231	0.0235	0.0004
Red	t_0	24	10	6 *E*	0	37	0.0338		
	t_1	22	10	6 *E*	0	35	0.0299	0.0301	0.0002
NIR	t_0	24	10	4 *E*	0	37	0.0409		
	t_1	22	12	6 *E*	0	35	0.0331	0.0333	0.0002
Site 1 GF~Landsat fusion									
Blue	t_0	40	34	30 *E*	10 *N*	56	0.0340		
	t_1	36	30	24 *E*	8 *N*	59	0.0336	0.0337	0.0001
Green	t_0	44	56	24 *E*	8 *N*	48	0.0293		
	t_1	42	50	26 *E*	8 *N*	46	0.0352	0.0353	0.0001
Red	t_0	60	44	24 *E*	8 *N*	50	0.0369		
	t_1	54	40	20 *E*	8 *N*	50	0.0445	0.0446	0.0001
NIR	t_0	74	30	28 *E*	8 *N*	48	0.0513		
	t_1	70	24	30 *E*	8 *N*	48	0.0561	0.0562	0.0001
Site 2 GF~Landsat fusion									
Blue	t_0	20	24	30 *E*	28 *S*	55	0.0241		
	t_1	28	30	34 *E*	32 *S*	60	0.0062	0.0063	0.0001
Green	t_0	20	24	28 *E*	30 *S*	55	0.0122		
	t_1	28	30	32 *E*	32 *S*	62	0.0087	0.0088	0.0001
Red	t_0	30	30	28 *E*	28 *S*	55	0.0279		
	t_1	28	30	30 *E*	30 *S*	51	0.0194	0.0194	0
NIR	t_0	26	30	30 *E*	28 *S*	54	0.0514		
	t_1	20	28	32 *E*	26 *S*	49	0.1022	0.1023	0.0001

[a] Note: *E*, *S*, and *N* represent shifts to the east, the south, and the north, respectively.
[b] RMSE is the root mean square error in SMF between the predicted image and the corresponding fine observation image.
[c] RMSE' represents the RMSE between the predicted image based on SMF at t_0 and the fine-resolution observation image at t_1.
[d] ΔRMSE is the difference between RMSE and RMSE'.

where x and y are observed and predicted values, respectively, and n is the number of pixels. The ideal value of CC is 1. The linear relationship between predicted and actual reflectance is stronger when the CC value is greater. The value of RMSE and MAD closer to 0 indicates more accurate prediction and better performance of spatiotemporal fusion algorithm. Furthermore, pixel-wise comparisons between predicted and actual reflectance for each band were shown in density plots.

IV. RESULTS

A. Stability of SMF Across Time

Table III shows the calibration details of the SMF at both t_0 and t_1 for the real datasets. In all cases, the shapes of SMF characterized by FWHM were mostly unsymmetrical. The optimal FWHM was no less than 10 or 30 m, i.e., beyond one Sentinel-2 or no less than 10 or 30 m, i.e., beyond one Sentinel-2 or Landsat-8 pixel. The shifts of SMF for all bands on Site 1 were in the northeast and east direction in the case of GF~Landsat fusion and GF~Sentinel fusion, respectively. Conversely, the shifts of SMF on Site 2 were in the southeast direction for GF~Landsat fusion. The optimal rotation angle varied from 46° to 62° for GF~Landsat fusion on the two sites, while the rotation angles for GF~Sentinel fusion on Site 1 were relatively constant between different bands (from 35° to 39°). When replacing the SMF at t_1 with the optimal SMF at t_0, the RMSE between the SMF predicted image and the GF-1 image at t_1 (RMSE') were still low and the differences in RMSE (ΔRMSE) were minimal in all cases. Given that the maximum value of ΔRMSE was as small as 0.0004, the SMF could be regarded stable across time. Therefore, it is feasible to characterize the spatial scale relationship between the available Landsat-8/Sentinel-2 image and the unknown GF-1 image at t_1 using the SMF obtained from the available image pair at t_0.

B. Test on Simulated Data Over a Large Area (Site 1)

Figs. 4 and 5 show the predicted fine images at t_1 from simulated datasets (both GF~Sentinel and GF~Landsat) of subset 1 and subset 2 on Site 1, respectively. Generally, the predictions for the simulated GF~Sentinel fusion for all the four methods were more accurate than those for the simulated GF~Landsat fusion. The predicted images by the four methods were closer to the observation image, and the spatial details in the subset images were clearer. For the simulated GF~Landsat fusion, the predicted images by STARFM and FSDAF exhibited some blurry patches, and the spatial heterogeneity within fields were unclear [e.g., the areas marked with white dashed ellipses in Figs. 4(e) and (f) and 5(e) and (f)]. On the contrary, the images predicted by Fit-FC [see Fig. 4(g)] and HISTIF [see Figs. 4(h) and 5(h)] were visibly close to the original image GF(t_1). However, some

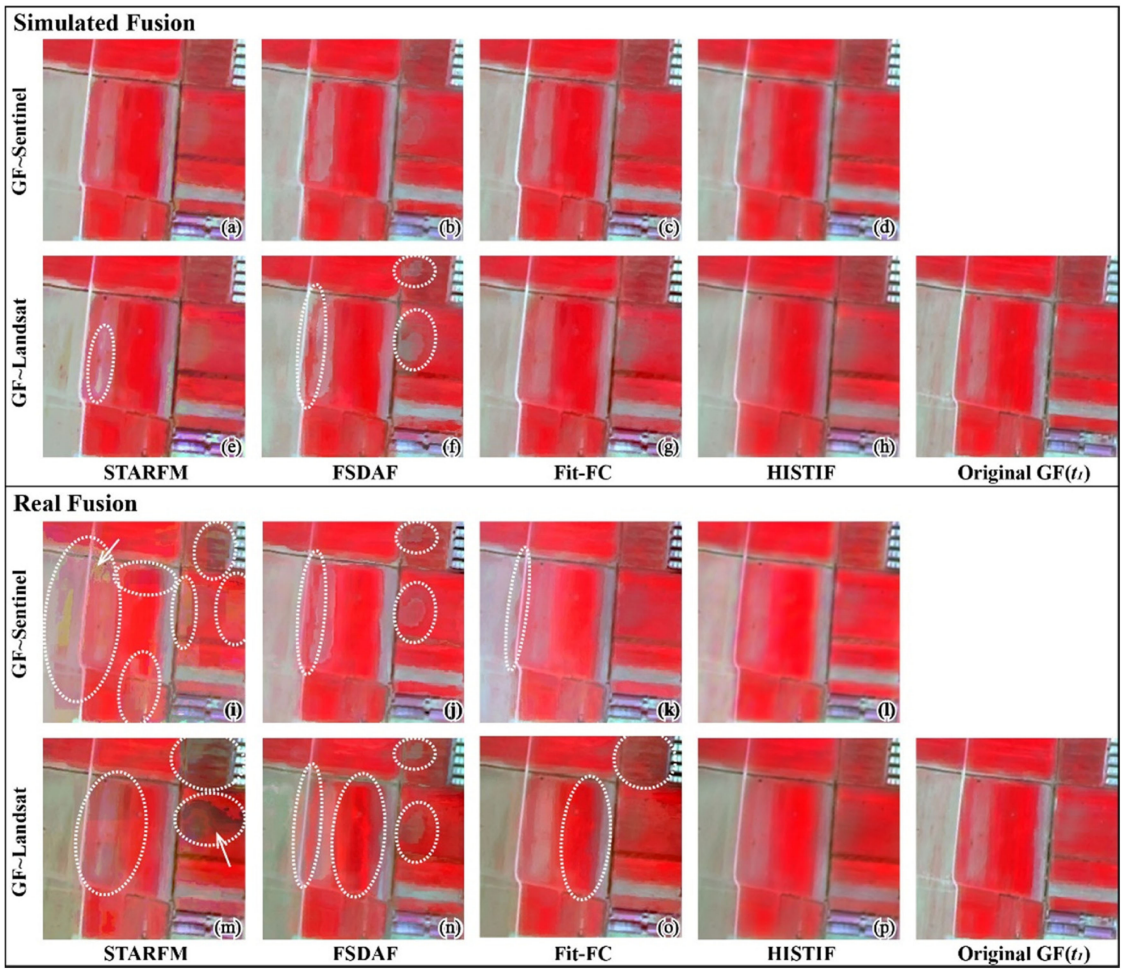

Fig. 4. Comparison of the predicted fine images with the four fusion methods (STARFM, FSDAF, Fit-FC, and HISTIF) for both simulated and real fusion and the observation image GF(t_1) of subset 1 on Site 1. All false-color images are displayed with NIR-red-green as RGB. The white dashed ellipses denote the outlying areas in the predicted images. The density scatters of the whole images predicted by the four methods are shown in Figs. 13–15 in the Appendix.

negative values were unexpectedly predicted by Fit-FC in the simulated GF∼Landsat fusion [see Figs. 13(c), (g), (k), and (o)].

In Table IV, the statistical assessments for the simulated GF∼Landsat and GF∼Sentinel fusions show differences in prediction accuracy of the four methods. Generally, HISTIF yielded higher accuracy than the other three. Although STARFM and Fit-FC achieved the same MAD (0.0085) for the red band as HISTIF did in the simulated GF∼Landsat fusion, the MAD values of HISTIF were the lowest for blue and NIR bands. The lowest mean RMSE was also seen for HISTIF in the simulated GF∼Sentinel fusion, which was 19%, 10%, and 10% lower than the RMSE values for STARFM, FSDAF, and Fit-FC, respectively.

C. Test on Real Data Over a Large Area (Site 1)

Compared to the predicted images from the simulated datasets, the predictions for the real datasets were much poorer for STARFM, FSDAF, and Fit-FC (see Figs. 4 and 5). On the contrary, the predicted images by HISTIF appeared close to the actual GF-1 observations regardless of simulated or real datasets (see Figs. 4 and 5). In the predicted images of HISTIF on subset 1 [see Fig. 4(l) and (p)], clear objects could be seen around the wheat fields, including bare soil, greenhouses, buildings, irrigation channels, and roads. Conversely, the image predicted by STARFM [see Fig. 4(i) and (m)] appeared blurry and exhibited unclear boundaries at the junction of fields and other surface features. Moreover, spectral distortion appeared noticeably in the STARFM predictions. For example, the area denoted by the white arrow in Fig. 4(i) should be light red but is predicted inaccurately as green; the area with arrow in Fig. 4(m) should be mixed gray and red but is predicted inaccurately as black. This problem also appeared in the predicted images by STARFM on subset 2, especially near the border of the farmland [see Fig. 5(i) and (m)]. Compared to STARFM, predictions by the FSDAF [see Fig. 4(j) and (n), and Fig. 5(j) and (n)] and Fit-FC [see Fig. 4(k) and (o) and Fig. 5(k) and (o)] were improved but blurry field boundaries still existed with some noticeable noises (e.g., the areas marked with white dashed ellipses). With respect to the fusion performance for the area with within-field variability, the STARFM prediction was the blurriest. The within-field variability in the FSDAF and Fit-FC predicted images were clearer. However, the predicted images of FSDAF have some mottled features, such as the area marked by

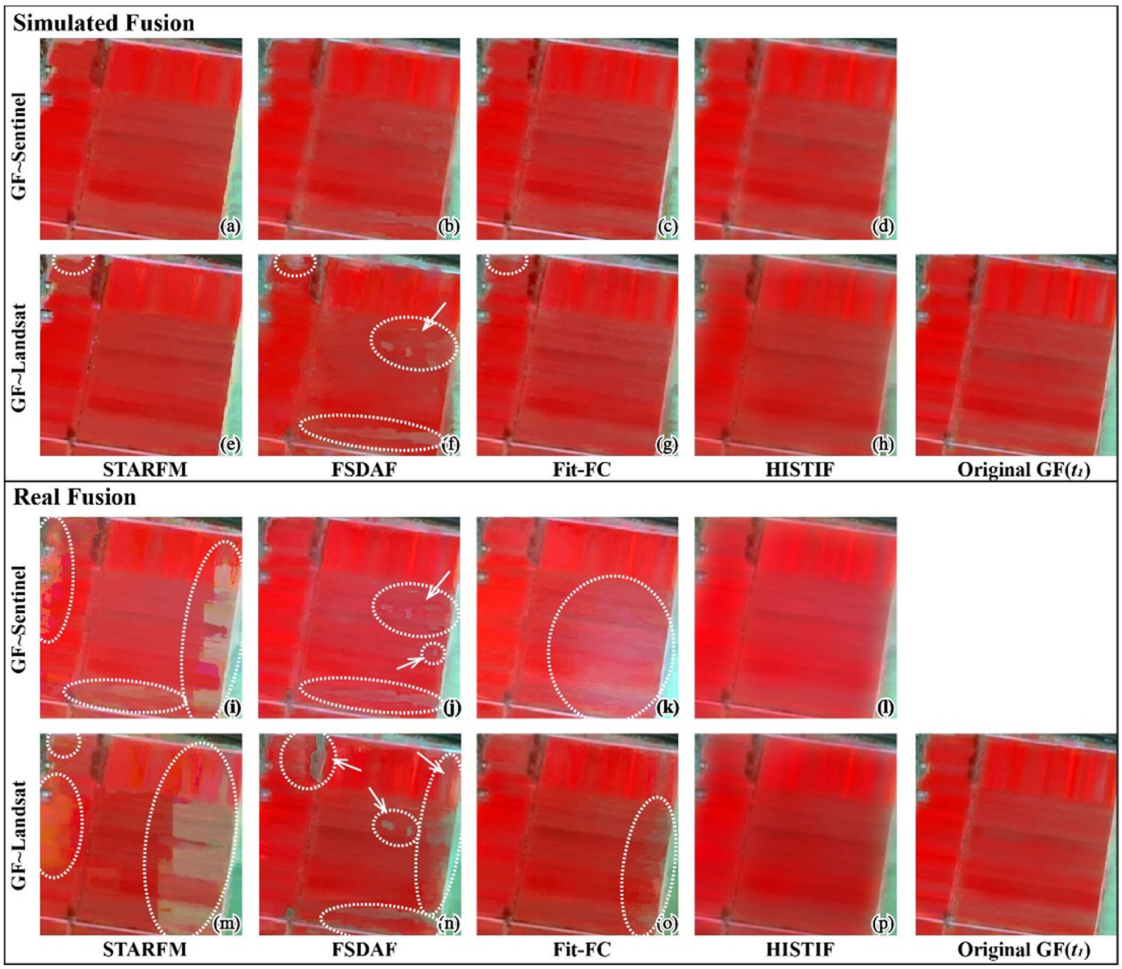

Fig. 5. Comparison of the predicted fine images with the four fusion methods (STARFM, FSDAF, Fit-FC, and HISTIF) for both simulated and real fusion and the observation image GF(t_1) of subset 2 on Site 1. All false-color images are displayed with NIR-red-green as RGB. The white dashed ellipses denote the outlying areas in the predicted images. The density scatters of the whole images predicted by the four methods are shown in Figs. 13–16 in the Appendix.

TABLE IV
STATISTICAL ASSESSMENT OF BANDWISE PREDICTIONS WITH THE FOUR SPATIOTEMPORAL IMAGE FUSION METHODS FOR THE SIMULATED DATA ON SITE 1

		GF~Landsat fusion				GF~Sentinel fusion			
		STARFM	FSDAF	Fit-FC	HISTIF	STARFM	FSDAF	Fit-FC	HISTIF
CC	Blue	0.9486	0.9350	0.9442	**0.9496**	0.9610	0.9666	0.9636	**0.9735**
	Green	0.9404	0.9270	0.9390	**0.9429**	0.9549	0.9622	0.9623	**0.9701**
	Red	0.9560	0.9461	0.9579	**0.9592**	0.9683	0.9739	0.9751	**0.9801**
	NIR	0.9678	0.9641	0.9699	**0.9726**	0.9767	0.9816	0.9822	**0.9847**
	Mean	0.9532	0.9431	0.9528	**0.9561**	0.9652	0.9711	0.9708	**0.9771**
RMSE	Blue	0.0083	0.0092	0.0087	**0.0082**	0.0072	0.0067	0.0070	**0.0060**
	Green	0.0108	0.0119	0.0111	**0.0106**	0.0094	0.0086	0.0086	**0.0077**
	Red	0.0129	0.0143	0.0128	**0.0125**	0.0110	0.0100	0.0098	**0.0087**
	NIR	0.0196	0.0207	0.0189	**0.0181**	0.0167	0.0148	0.0146	**0.0135**
	Mean	0.0129	0.0140	0.0129	**0.0124**	0.0111	0.0100	0.0100	**0.0090**
MAD	Blue	0.0054	0.0059	0.0055	**0.0052**	0.0044	0.0039	0.0041	**0.0034**
	Green	**0.0070**	0.0079	0.0072	0.0071	0.0059	0.0052	0.0053	**0.0045**
	Red	**0.0085**	0.0096	**0.0085**	**0.0085**	0.0070	0.0061	0.0062	**0.0053**
	NIR	0.0134	0.0145	0.0133	**0.0127**	0.0111	0.0098	0.0098	**0.0088**
	Mean	0.0086	0.0095	0.0086	**0.0084**	0.0071	0.0063	0.0064	**0.0055**

Note: Bold values indicate the best accuracy by row.

TABLE V
STATISTICAL ASSESSMENT OF BANDWISE PREDICTIONS WITH THE FOUR SPATIOTEMPORAL IMAGE FUSION METHODS FOR THE REAL DATA ON SITE 1

		GF~Landsat fusion				GF~Sentinel fusion			
		STARFM	FSDAF	Fit-FC	HISTIF	STARFM	FSDAF	Fit-FC	HISTIF
CC	Blue	0.8862	0.8871	0.8722	**0.9153**	0.9024	0.9133	0.9002	**0.9241**
	Green	0.8833	0.8904	0.8932	**0.9240**	0.9004	0.9273	0.9039	**0.9401**
	Red	0.9122	0.9137	0.9182	**0.9484**	0.9169	0.9337	0.9399	**0.9516**
	NIR	0.9304	0.9323	0.9382	**0.9623**	0.9364	0.9433	0.9403	**0.9601**
	Mean	0.9030	0.9059	0.9055	**0.9375**	0.9140	0.9294	0.9211	**0.9440**
RMSE	Blue	0.0128	0.0126	0.0146	**0.0115**	0.0140	0.0126	0.0178	**0.0124**
	Green	0.0157	0.0157	**0.0144**	0.0148	0.0173	0.0149	0.0226	**0.0139**
	Red	0.0175	0.0179	0.0166	**0.0165**	0.0193	0.0170	0.0234	**0.0156**
	NIR	0.0243	0.0240	0.0239	**0.0221**	0.0268	0.0262	0.0285	**0.0252**
	Mean	0.0176	0.0176	0.0174	**0.0162**	0.0194	0.0177	0.0231	**0.0168**
MAD	Blue	0.0100	0.0098	0.0115	**0.0084**	0.0103	**0.0090**	0.0138	0.0094
	Green	0.0125	0.0126	**0.0107**	**0.0107**	0.0134	0.0113	0.0184	**0.0103**
	Red	0.0128	0.0128	0.0123	**0.0122**	0.0134	0.0123	0.0189	**0.0110**
	NIR	0.0185	0.0182	0.0183	**0.0168**	0.0209	0.0207	0.0229	**0.0198**
	Mean	0.0135	0.0134	0.0132	**0.0120**	0.0145	0.0133	0.0185	**0.0126**

Note: Bold values indicate the best accuracy by row.

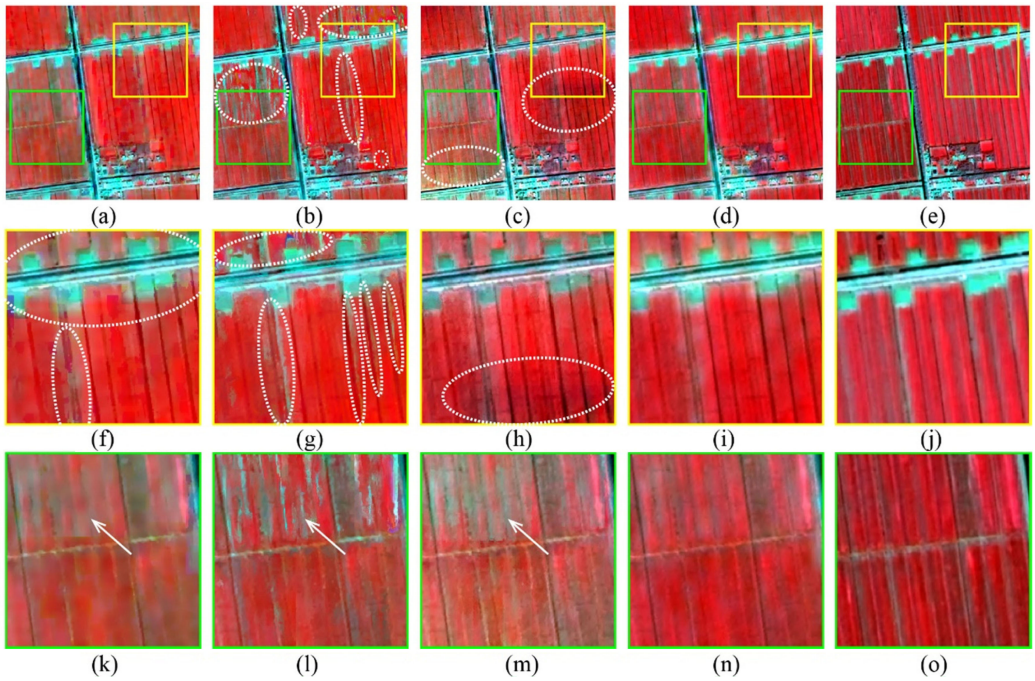

Fig. 6. Comparison of the predicted fine-resolution images with the four image fusion methods (a. STARFM, b. FSDAF, c. Fit-FC, and d. HISTIF) with the observation image (e) for Site 2 (NIR-red-green bands as RGB). The middle (f)–(j) and bottom (k)–(o) rows represent the subset images of areas outlined in (a)–(e) with yellow and green squares, respectively. The white dashed ellipses and arrows denote the outlying areas in the predicted images.

arrows in Fig. 5. Moreover, the spectral information was poorly estimated by Fit-FC. For example, the marked field in Fig. 5(k) was inaccurately predicted as shown in white patterns in GF(t_1). Generally, the predicted values of HISTIF were more strongly correlated to the actual GF-1 data than those of the other three methods (see Figs. 4, 5, 15, and 16).

The statistical assessment (see Table V) also demonstrates that the accuracy of HISTIF was generally higher than that of STARFM, FSDAF, and Fit-FC. For the real GF∼Landsat fusion, only the RMSE of Fit-FC in green band (0.0144) was smaller than that of HISTIF (0.0148) and HISTIF yielded the lowest mean RMSE (0.0162). For the real GF∼Sentinel fusion, although FSDAF achieved the lowest MAD for the blue band of the whole image (0.009), the MAD values of HISTIF were the lowest for green, red, and NIR bands. The accuracy of Fit-FC prediction was relatively worse than that of the other methods.

D. Test on Real Data on Small Farms (Site 2)

Fig. 6 presents the fine image at t_1 predicted by the four spatiotemporal fusion methods using real datasets on Site 2.

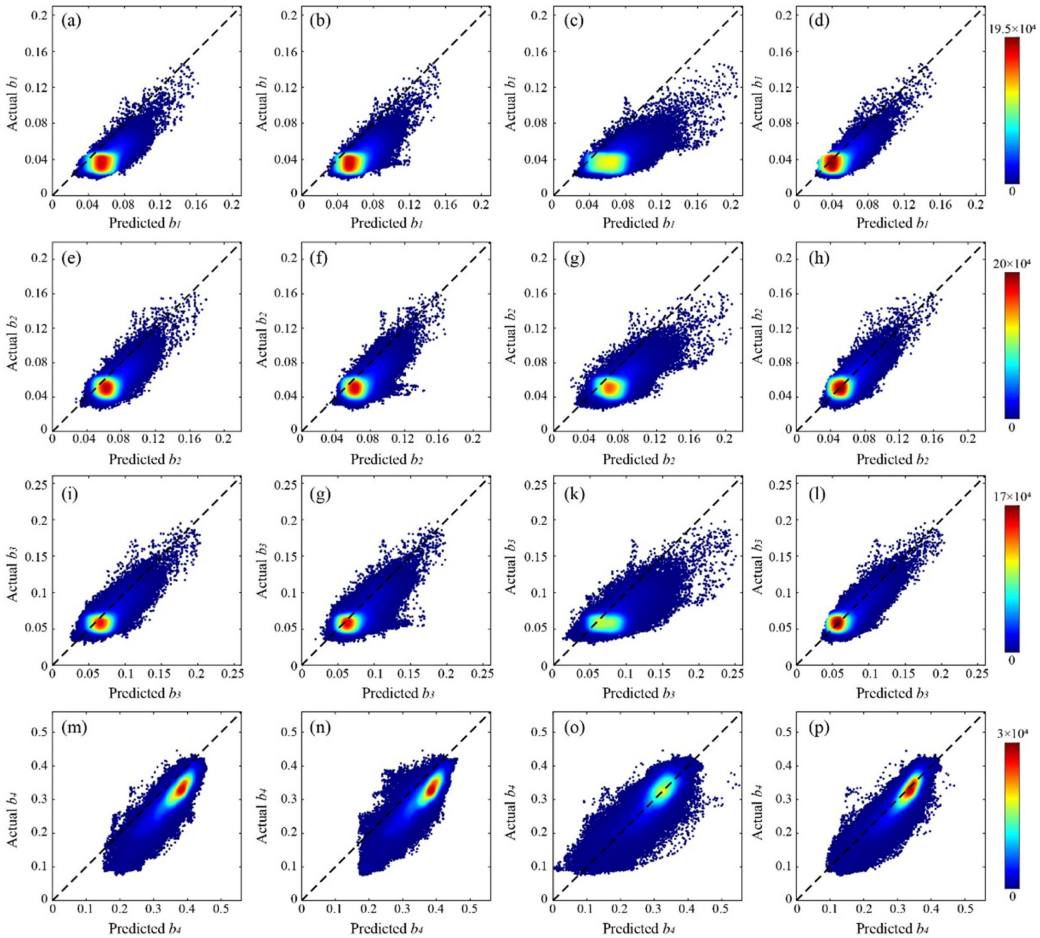

Fig. 7. Density plots of the actual and predicted values in four bands for the winter wheat site with the four spatiotemporal image fusion methods. Panels (a)–(d), (e)–h), (i)–(l), and (m)–(p) are the plots for blue, green, red, and NIR band, respectively. The four columns from left to right correspond to the image fusion methods STARFM, FSDAF, Fit-FC, and HISTIF, respectively.

It is apparent that the predicted image by HISTIF [see Fig. 6(d)] is closer to the actual observation from GF-1 [see Fig. 6(e)] than those predicted by STARFM, FSDAF, and Fit-FC [see Fig. 6(a)–(c)]. In the subset images [see Fig. 6(f)–(j)], there are bare soil, irrigation canals, and roads intersected around the fields. The image predicted by STARFM is blurry [see Fig. 6(k)]. At the junction of fields and other surface features, different objects seemed to be mixed and led to unclear boundaries. With respect to the FSDAF prediction, it has even more blurred boundaries [e.g., the areas marked with white dashed ellipses in Fig. 6(g)]. In contrast, the predicted images from Fit-FC [see Fig. 6(h)] and HISTIF [see Fig. 6(i)] present much clearer field boundaries. However, the color of Fit-FC prediction is visually darker than the original GF-1 image with some noticeable noises [see Fig. 6(h)]. Therefore, HISTIF has better performance in the distinction of between-class heterogeneity. As for the fusion performance in the area with within-field variability [see Figs. 6(k)–(o)], we can clearly identify that the image predicted by HISTIF contains more spatial and spectral details than the other three predictions. The STARFM prediction is the most blurred [see Fig. 6(k)]. Compared with the STARFM prediction, the FSDAF predicted image is clearer but with mottled features [see Fig. 6(l)]. Moreover, the spectral information is obviously misestimated by STARFM, FSDAF, and Fit-FC. For example, the fields are inaccurately predicted as a mix of red and cyan [the area denoted by arrows in Fig. 6(k)–(m)] as opposed to the red color in Fig. 6(n)–(o).

As shown in the density plots in Fig. 7, the values predicted are generally significantly closer to actual values for HISTIF than for the other methods. Its distribution is approximately a spindle along 1:1 line with the large red cluster, which represents the highest density of data points [see Fig. 6(d), (h), (l), and (p)]. Moreover, the range of red cluster exactly corresponds to the values of wheat pixels. However, STARFM, FSDAF, and Fit-FC overestimate the values for all bands especially in blue, green, and red bands. The data points of the Fit-FC prediction are the most discrete [see Figs. 6(c), (g), (k), and (o)].

The statistical assessment for each band on Site 2 is listed in Table VI. Generally, the mean RMSE for HISTIF is 0.0139 with a decrease by 47.5%, 49.1%, and 45.7% when compared with STARFM, FSDAF, and Fit-FC. Although the CC value of STARFM in NIR band is the highest compared with that of the other methods in the NIR band, it is only 0.058 larger than the CC value of HISTIF. The Fit-FC produced the lowest CC with a mean value around 0.6794. STARFM, FSDAF, and HISTIF yielded better accuracies and the mean CCs are 0.7721,

TABLE VI
STATISTICAL ASSESSMENT OF BANDWISE PREDICTIONS WITH THE FOUR SPATIOTEMPORAL IMAGE FUSION METHODS FOR THE REAL DATA ON SITE 2

		STARFM	FSDAF	Fit-FC	HISTIF
CC	Blue	0.6918	0.6841	0.5807	**0.7540**
	Green	0.7323	0.7120	0.6661	**0.7749**
	Red	0.7657	0.7447	0.6599	**0.8221**
	NIR	**0.8985**	0.8820	0.8108	0.8927
	Mean	0.7721	0.7557	0.6794	**0.8109**
RMSE	Blue	0.0211	0.0215	0.0278	**0.0105**
	Green	0.0155	0.0162	0.0195	**0.0081**
	Red	0.0136	0.0147	0.0218	**0.0102**
	NIR	0.0558	0.0567	0.0332	**0.0267**
	Mean	0.0265	0.0273	0.0256	**0.0139**
MAD	Blue	0.0193	0.0195	0.0244	**0.0082**
	Green	0.0131	0.0134	0.0163	**0.0064**
	Red	0.0110	0.0110	0.0171	**0.0080**
	NIR	0.0507	0.0516	0.0257	**0.0201**
	Mean	0.0235	0.0239	0.0209	**0.0107**

Note: Bold values indicate the best accuracy by row.

0.7557, and 0.8109, respectively. For all bands, the MAD values of HISTIF are the lowest.

V. DISCUSSION

A. Correction of the Cross-Scale Spatial Mismatch

To produce frequent fine images, many spatiotemporal fusion methods have been developed by blending temporally sparse fine images and frequent coarse images from different sensors. Because of the differences in system design and observation scale, the PSF and geometric discrepancies between two types of satellite data would influence the final fusion performance [15], [27], [30]. However, previous algorithms either ignored the two issues causing cross-scale spatial mismatch, or considered only one of them [27], [35].

To overcome this limitation, HISTIF considers both PSF effect and misregistration in the step of FCSM. In Table III, we can see that the shape of PSF is not a typical square or rectangle, which is in agreement with previous studies [35], [38], [45]. The difference in imaging direction between the two sensors was corrected by the rotation angle (θ). The geo-registration error was refined by shifting the coarse image in both East-West and North-South dimensions. The SMF of GF∼Sentinel fusion needed to be shifted eastward, although the shifty were 0 for all bands (see Table III). That suggested that there were still geo-registration errors even though the co-registration of image pairs had been performed during data preprocessing. Wang *et al.* [23] also suggested that additional geometric correction should be done especially for improving the performance of the spectral similarity-based fusion methods. The nonzero shift values in HISTIF indicated that the SMF should be moved along a distance and direction, rather than the image. Moreover, the optimal shifts relied strongly on the other parameters (i.e., FWHM_x, FWHM_y, and θ) and vice versa. Therefore, the parameters obtained from FCSM could not be separated from HISTIF and applied directly to images to be fused or other methods. The results demonstrated that we were able to correct the cross-scale spatial mismatch by combining the five parameters (FWHM_x, FWHM_y, shift_x, shift_y, θ) in HISTIF.

Those five parameters for the filter were optimized based on the image pair at t_0 and then applied to the coarse image at t_1, based on the assumption that the sensor difference was constant between t_0 and t_1. Although the optimal parameters at the two times were unequal (see Table III), the prediction accuracies of SMF at t_0 and t_1 only exhibited a marginal difference (ΔRMSE) of 0.00013 on average. This confirmed the stable performance of the proposed SMF across time, which means that the cross-scale spatial mismatch can be corrected at t_1 with the SMF derived from the image pair at t_0.

B. Characterization of Within-Field Variability

The farmland landscape often exhibits both between-class heterogeneity caused by the alternate distribution of different cover types and within-field variability caused by the difference in growth conditions within the same crop cover [46]. To improve spatial details, the two methods FSDAF and Fit-FC evaluated in this study adopt different solutions. FSDAF uses a thin plate spline interpolation to improve the image prediction accuracy, but it is only based on the spatial dependence of the coarse pixels [43]. The predicted images of FSDAF exhibited oversmoothness within fields, some flecks, and border effects (see Figs. 4–6). In agreement with previous studies [30], this research implies that FSDAF is not applicable to the areas where the within-class variability is small. However, we found that FSDAF could capture more spatial and spectral details within fields when the scale difference between coarse and fine images became smaller (e.g., the GF∼Sentinel fusion results in Figs. 4–6). Fit-FC uses spectrally similar pixels from the neighborhood and has the number of classes set up based on the size of objects to enhance spatial details [29]. Thus, it is expected to reveal greater within-field variability than STARFM and FSDAF. According to the visual assessment (see Figs. 4–6), the predicted images of Fit-FC showed more spatial details and clearer boundaries than those of STARFM and FSDAF. However, Fit-FC unexpectedly yielded poor estimation of some spectral information especially for the real GF∼Landsat fusion on Site 1 with negative values (see Fig. 13).

The combination of MMTC and FCSM in HISTIF could improve within-field heterogeneity and spectral accuracy. However, neither FCSM nor MMTC could be omitted in order to obtain the best performance. Our results show that HISTIF could benefit from the application of SFM in the step FCSM. If using the multiplicative factor without running FCSM, the predicted images would exhibit faint blocky artifacts (see Fig. 8) due to the pixel boundaries of the coarse image [20]. As a result, the spatial details within fields would decrease sharply and the between-class heterogeneity would be smoothed to lose clear boundaries (see Fig. 8). On the contrary, the use of SFM could enhance the predicted images because SFM considers the sensor PSF and performs deblurring. Kwan *et al.* [47] also found that the PSF as a convolution matrix could present the blur of hyperspectral

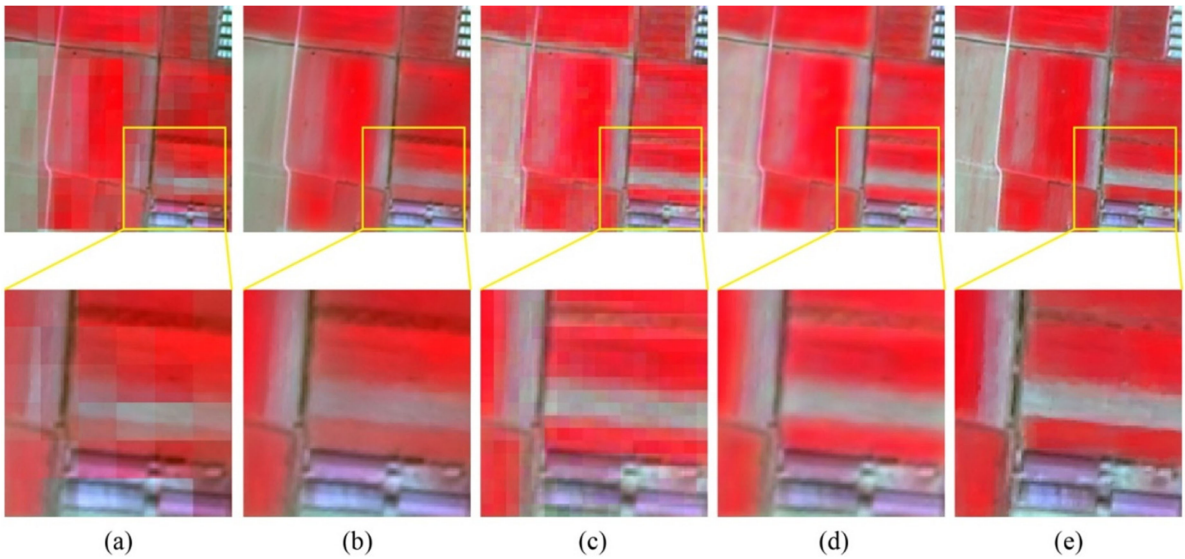

Fig. 8. Comparison of the predicted subset 1 images without and with using FCSM for the real GF∼Landsat fusion (a) and (b) and the real GF∼Sentinel fusion (c) and (d). (e) represents the original GF-1 subset image at t_1.

bands and enhance the super-resolution performance. Based on the output from FCSM, the implementation of MMTC would further improve the spectral reconstruction accuracy of HISTIF. Instead of class-based processing, the MMF of temporal changes is directly calculated per pixel between t_0 and t_1. As a result, even the minor within-field variability can be reconstructed and the boarder effect could be eliminated (see Figs. 4–6). Compared to linear operations used to quantify temporal changes in most pixel-based methods, MMF can avoid negative predictions of pixel values. For example, STARFM fills negative values as –9999 (see Fig. 16), and Fit-FC has no strategy to deal with this problem (see Fig. 13). Based on the ratio form of MMF (8), there are no negative predictions in HISTIF as long as the values from the input images are valid. However, a bias term needs to be considered in future work to improve the correction of differences between the images at two dates [13], [48].

C. Universality of HISTIF Across Sensors and Sites

In this study, we used both simulated and real datasets from different sensors (i.e., GF∼Landsat and GF∼Sentinel) to evaluate the fusion performance of HISTIF. The scale factor between Landsat-8 (30 m) and GF-1 (2 m) pixels is 15 and that between Sentinel-2 (10 m) and GF-1 (2 m) pixels is only 5 (see Table II). As the scale factor varied, the FCSM parameters in HISTIF were different for the two datasets. According to Table III, all the parameters of FCSM in HISTIF for GF∼Sentinel fusion were smaller than those for GF∼Landsat fusion. For example, the shifts of SMF in GF∼Sentinel fusion (2 to 3 GF-1 pixels of $shift_x$ and 0 of $shift_y$) were much smaller than those in GF∼Landsat fusion (10 to 15 and 4 to 5 GF-1 pixels of shift_x and shift_y, respectively) on Site 1. That suggested a smaller scale difference between coarse and fine images yielded a lower scale mismatch. Furthermore, it implied that FCSM could be adaptive to different datasets through the automatic determination of the five parameters. Without using FCSM, the applicability and accuracy of HISTIF would be greatly reduced, especially for the fusion with a large-scale difference between coarse and fine images.

In the relevant studies on FSDAF [43], Fit-FC [29], and other fusion methods [12], [49], the input coarse images were not real observations from satellites but simulated by spatial aggregation from the fine images. In those ideal situations, there was no sensor difference between coarse and fine images but only scale difference. Therefore, those studies lack evaluations of method performance in practical applications. In Fig. 9, we could see sharp decreases in the CC values from the real datasets for STARFM, FSDAF, and Fit-FC, and the RMSE and MAD values from the real datasets for those three methods increased dramatically, especially for Fit-FC. Regardless of simulated or real data, the performances of HISTIF were better than that of STARFM, FSDAF, and Fit-FC on both simulated and real data. In terms of different sensors, the CC, RMSE, and MAD values for HISTIF were similar in both real GF∼Sentinel fusion and GF∼Landsat fusion. However, performances of Fit-FC evaluated by RMSE and MAD based on the real GF∼Sentinel data were much worse than those in GF∼Landsat fusion. This may be due to overfitting, given that GF∼Sentinel data can provide more spatial details than GF∼Landsat data. In addition, common learning-based methods, such as SPSTFM [3] and one-pair learning [14], are not suitable for the dataset with a large-scale difference [43], [50]. Therefore, HISTIF can reconstruct within-field spatial details accurately and is stable across sensors.

We note that the parameters are different among bands and between sites (see Table III). Specifically, the SMF parameters are marginally different across bands (see Table VII). For example, the differences of the FWHM and shift are almost less than one pixel (i.e., 10 m for Sentinel-2 and 30 m for Landsat-8) in the three datasets (see Table VII). The slight difference might be caused by the spectral variation on each band. In recent test

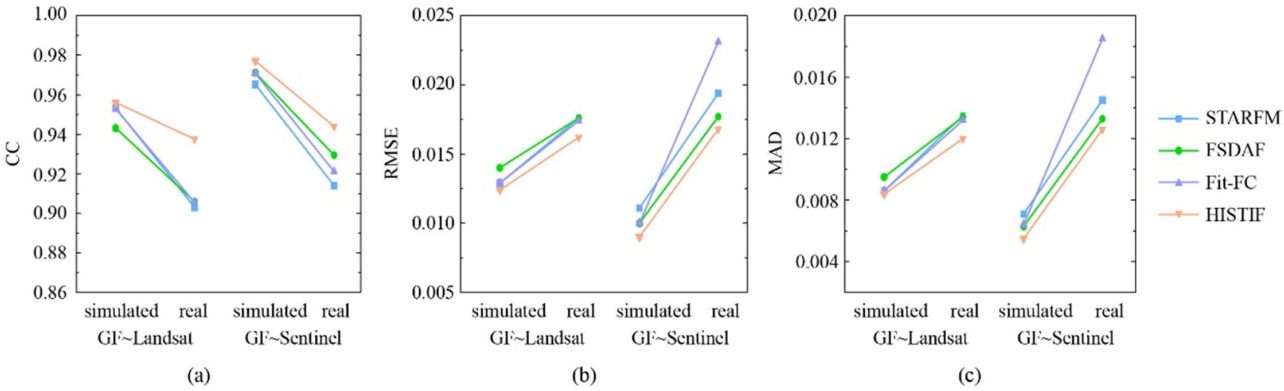

Fig. 9. Comparison of (a) the correlation coefficient (CC), (b) the root mean square error (RMSE), and (c) the mean absolute difference (MAD) between the simulated and real data.

TABLE VII
MINIMUM, MAXIMUM, AND DIFFERENCE OF THE PARAMETERS OF SMF

	FWHM		Shift (m)		Rotation angle (°)
	$FWHM_x$	$FWHM_y$	$shift_x$	$shift_y$	
Site 1 GF~Sentinel fusion					
Min	18	10	4 *E*	0	35
Max	26	18	6 *E*	0	39
Max-Min	**8**	**8**	**2 *E***	**0**	**4**
Site 1 GF~Landsat fusion					
Min	36	24	20 *E*	8 *N*	46
Max	74	56	30 *E*	10 *N*	59
Max-Min	**38**	**32**	**10 *E***	**2 *N***	**13**
Site 2 GF~Landsat fusion					
Min	20	24	28 *E*	26 *S*	49
Max	30	30	34 *E*	32 *S*	62
Max-Min	**10**	**6**	**6 *E***	**6 *S***	**13**

Note: *E*, *S*, and *N* represent shifts to the east, the south, and the north, respectively, and the bold values indicate the best accuracy by row.

over a large number of Sentinel-2 and Landsat-8 scenes globally, Yin *et al.* [51] suggested that the very accurate estimations of either the widths or shifts of PSF are not very important and we could transfer the parameters from one band to another. Although we also found that the cost function is fairly flat around the minimum, we still suggested optimizing the parameters of each band when applying HISTIF. In terms of different sites, the FWHM and rotation angles of SMF are similar but the shift appears more scene-dependent. Therefore, the effect of the pixel shift needs to be inferred on a scene by scene basis.

D. Advantages and Limitations of HISTIF and User Guide

For practical applications, the two major advantages of HISTIF are the minimal input requirement and high computational efficiency. In the growing season of either wheat or rice in recent five years, the number of available PMS/GF-1 high-resolution images (only one or two) was much smaller than that of available Landsat-8 and Sentinel-2 images (see Fig. 10). Therefore, the minimal input requirement of HISTIF could reduce the purchasing cost of high-resolution images and increase the practicability of spatiotemporal fusion. Moreover, consisting of only two steps with efficient processing, the HISTIF is advantageous in terms of computational efficiency. Some algorithms require trail-and-error experiments to define the optimal parameters (e.g., the size of moving window, the number of similar pixels, and the bands for determining the similar pixels). The fusion realization would fail if the parameters were inappropriate for the study area. Specifically, the computational efficiency of the STARFM-like methods would degrade dramatically when the size of moving window and the extent of study area increase. As for HISTIF, the SMF parameters in the first step were optimized automatically by employing the PSO algorithm. Due to the stability of SMF across time, the SMF could be calibrated only once over a set of image pairs before triggering the fusion algorithm. The second step is efficient since it only demands the time to traverse the entire image pixel by pixel. A comparison of the average runtimes between the four methods indicates that the single runtime was around 5 min for STARFM and HISTIF and tens of times longer for Fit-FC and FSDAF (see Fig. 11). Although the type of programming language has an impact on computational efficiency, the decisive factor for algorithm efficiency is the technical principle. If the time for parameterization is considered in practical use, the computational efficiency for Fit-FC and FSDAF will be even lower.

The new spatiotemporal fusion method was designed specifically for subfield-level crop monitoring and evaluated over crop sites with phenological changes between the two dates. Therefore, the land cover changes over the two dates were not considered. Although the use of MMF rather than an additive factor in HISTIF can eliminate negative prediction and yield better results in this study, HISTIF could be sensitive to sharp changes in spectral values (e.g., from small surface reflectance of farmlands to large values of greenhouses in the visible bands). Therefore, it is necessary to validate the application of HISTIF in the farmland area experiencing pronounced land cover changes and it is worthwhile to test the Index-then-Blend rather than Blend-then-Index blending strategy in the future [50], [52].

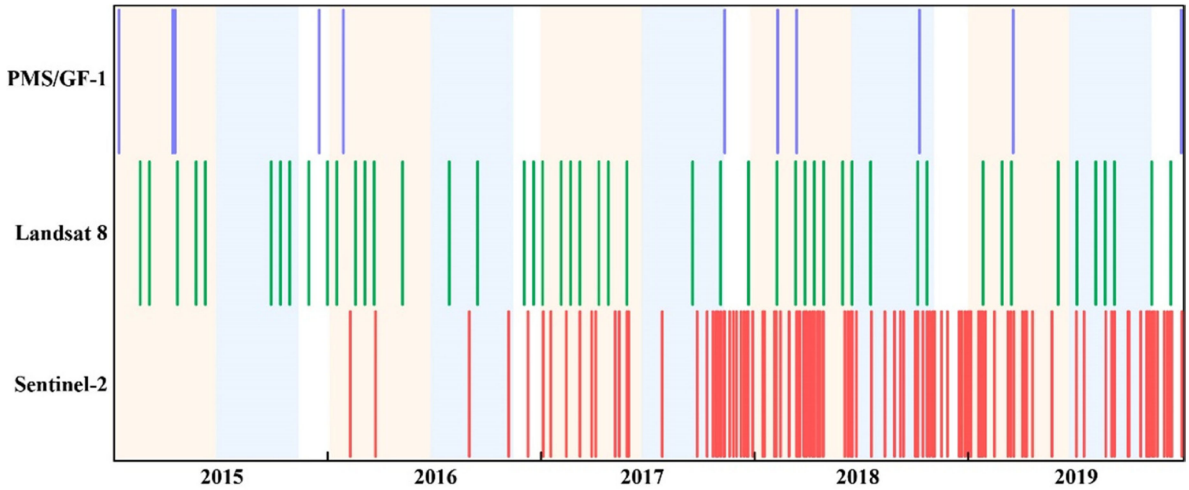

Fig. 10. Availability of PMS/GF-1, Landsat 8 and Sentinel-2 imagery for the study site in the recent five years. The growing season of winter wheat is from DOY (day of the year) 1 to 180, representing by a pale pink background. The growing season of paddy rice is from DOY 180 to 310 with a light blue background.

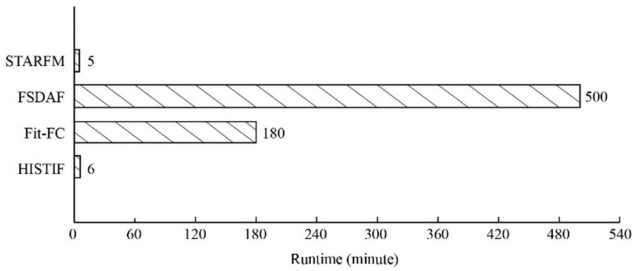

Fig. 11. Average runtimes of STARFM, FSDAF, Fit-FC and HISTIF for 3000×3000 pixels of one image pair with four bands on the same computer. Central processing unit (CPU): Intel(R) Xeon(R) CPU E5-2680 v2 @ 2.80 GHz 2.80 GHz (2 processors); installed memory (RAM): 128 GB; system type: 64-bit operating system, x64-based processor; operating system: Windows 10 Education. FSDAF was run in ENVI IDL 5.1, and both Fit-FC and HISTIF were run in MATLAB 2016a. STARFM was run in Ubuntu using C.

Since the step of MMTC operates pixel by pixel, the accuracy of the previous step FCSM is crucial. If the information of the image pair is unmatched (e.g., a clear fine image and cloudy coarse image), the SMF based on the image pair at t_0 would be unreliable and further affect the final output. Therefore, it is worth testing different ways of the MMF application over whole image in the future work. For example, Kwan *et al.* [13] found that predicted results would be more accurate if images were divided into patches. Accordingly, we might apply patch-based MMF in homogenous areas instead of using pixel-wise MMF to reduce overcompensation, while still applying pixel-wise MMF in heterogeneous areas. Moreover, the assumption that the temporal changes on different scales are consistent is crucial. If the temporal changes were considerably different between fine and coarse images, the fusion performance would be adversely affected. Therefore, it is necessary to check if temporal changes captured by image pairs are consistent before applying the HISTIF algorithm. Given the uncertainty in image processing, bandwidth differences, signal noise, and model errors, a residual might be considered in the future to correct for the small biases and eventually improve the fusion accuracy.

Due to the lack of submeter imagery, the fine images used in this study are pan-sharpened PMS/GF-1 bands. Although pan-sharpened PMS/GF-1 data have been widely used as a real and reliable product in many applications [53]–[58], it is important to use the original high-resolution data to do the fusion. Our results indicate the great potential of HISTIF in the applications of fusing Landsat-8 or Sentinel-2 and WorldView-2 or other high-resolution imagery. These applications could be accomplished by automatically adjusting the parameters of SMF. According to previous studies and our experience, the maximum value of the FHWM could be set as three times larger as the resolution of coarse imagery and the shift generally would not exceed twice of the coarse resolution. We suggested setting FHWM in x and y directions to the same range. In terms of the rotation, its range is from 0° to 180°. When the ranges of FHWM_x and FHWM_y are set to the same, the range of the rotation could be set from 0° to 90° to avoid repeated operations. We increased the practicability of our method by implementing HISTIF in MATLAB, which provides user-friendly and concise interfaces.

VI. Conclusion

To improve the continuous subfield-level crop monitoring, this article presents a new spatiotemporal fusion algorithm HISTIF to blend temporally sparse high-resolution images and frequent medium-resolution images. HISTIF addresses the two challenging issues, spatial scale mismatch and within-field variability, with only two steps (FCSM and MMTC). It was tested using both simulated and real datasets of GF∼Landsat and GF∼Sentinel and was compared with the three state-of-the-art spatiotemporal fusion methods STARFM, FSDAF, and Fit-FC. The correction of spatial scale mismatch in HISTIF led to substantial reduction in PSF calibration errors and geo-registration

errors that affect all pixel-based fusion methods. The visual assessment demonstrated that STARFM performed the worst in reducing blocky artifacts and improving spatial details within fields, especially for the real data. The minor within-field variability could only be predicted by Fit-FC and HISTIF, but predicted images of Fit-FC were less accurate due to the spectral distortion.

Compared to the benchmark methods, HISTIF has advantages in reconstructing fine spatial details within fields and reducing fusion errors from spatial scale mismatch with minimal input data. Furthermore, HISTIF was designed to accommodate the temporal changes per pixel rather than per class such that its performance was independent from a land cover classification map as required by traditional methods. This method has great potential in the application to other spatiotemporal imagery from different satellite sensors with stable applicability and high computing efficiency. The generated time series or dense stacks of fine images allow better support to subfield-level crop monitoring, crop growth modeling, and yield forecasting for the precision agriculture sector.

APPENDIX

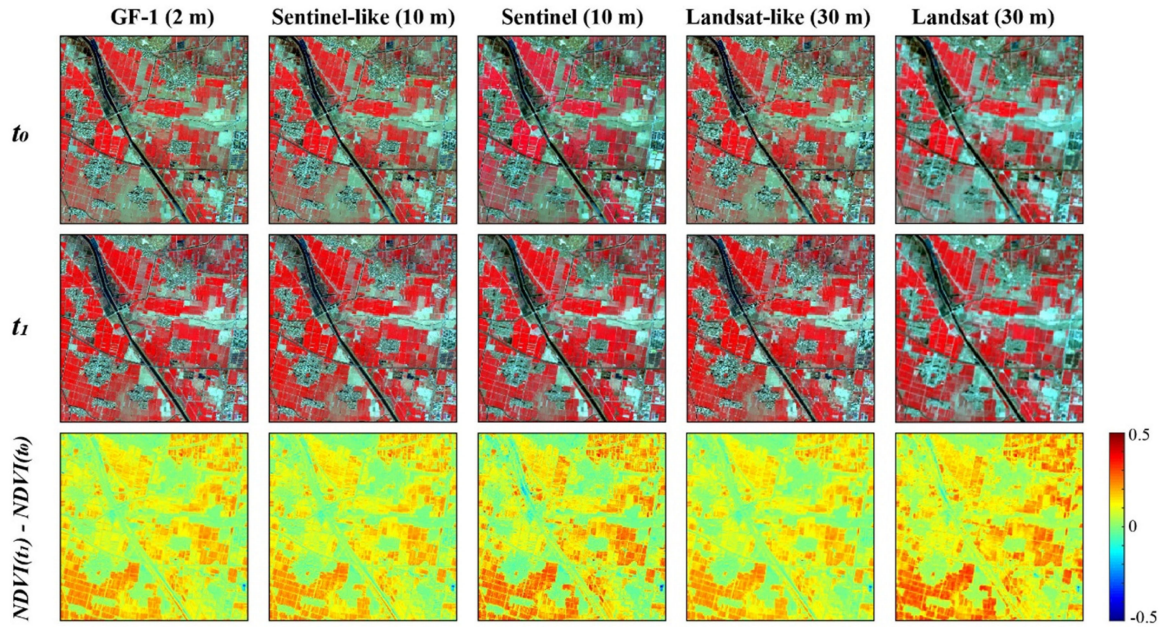

Fig. 12. Test whole images of Site 1 for t_0 and t_1 and the corresponding NDVI difference images at fine and coarse resolutions. The Sentinel-like and Landsat-like reflectance images were simulated from the corresponding GF-1 data, so t_0 and t_1 of them are the same as those of GF-1 on Site 1 in Table I. The NDVI difference images between the two dates were calculated by subtracting the NDVI at t_0 from the corresponding NDVI at t_1 for GF-1, Sentinel-like, Sentinel, Landsat-like, and Landsat dataset. All false-color images are displayed with NIR-red-green as RGB.

TABLE VIII
STATISTICAL COMPARISON OF THE TWO SITES IN FOUR BANDS FOR THE WHOLE IMAGES

		Site 1 GF~Sentinel		Site 1 GF~Landsat		Site 2 GF~Landsat	
		t_0	t_1	t_0	t_1	t_0	t_1
CC	Blue	0.7939	0.8877	0.6597	0.7467	0.3528	0.4319
	Green	0.7560	0.8743	0.6512	0.7185	0.4093	0.4328
	Red	0.8493	0.9086	0.7422	0.7767	0.5395	0.4557
	NIR	0.9368	0.9440	0.8437	0.8356	0.4249	0.5439
	Mean	0.8340	0.9037	0.7242	0.7694	0.4316	0.4661
RMSE	Blue	0.0247	0.0247	0.0347	0.0351	0.0244	0.0097
	Green	0.0294	0.0252	0.0306	0.0372	0.0132	0.0127
	Red	0.0357	0.0333	0.0393	0.0478	0.0290	0.0228
	NIR	0.0441	0.0384	0.0577	0.0645	0.0585	0.1102
	Mean	0.0335	0.0304	0.0406	0.0462	0.0313	0.0388
MAD	Blue	0.0226	0.0217	0.0320	0.0311	0.0225	0.0071
	Green	0.0268	0.0215	0.0268	0.0314	0.0105	0.0088
	Red	0.0326	0.0290	0.0347	0.0412	0.0260	0.0198
	NIR	0.0371	0.0290	0.0496	0.0544	0.0517	0.1003
	Mean	0.0298	0.0253	0.0358	0.0395	0.0277	0.0340

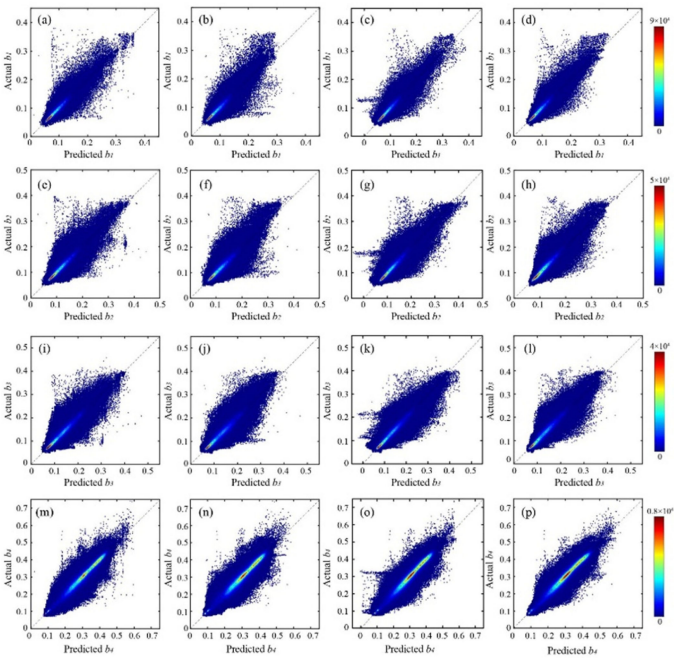

Fig. 13. Relationships of the actual and predicted values in four bands for the whole image with the four methods for the simulated GF∼Landsat fusion. Panels (a–d), (e–h), (i–l), and (m–p) are the plots for blue, green, red, and NIR band, respectively. The four columns from left to right correspond to the image fusion methods STARFM, FSDAF, Fit-FC, and HISTIF, respectively.

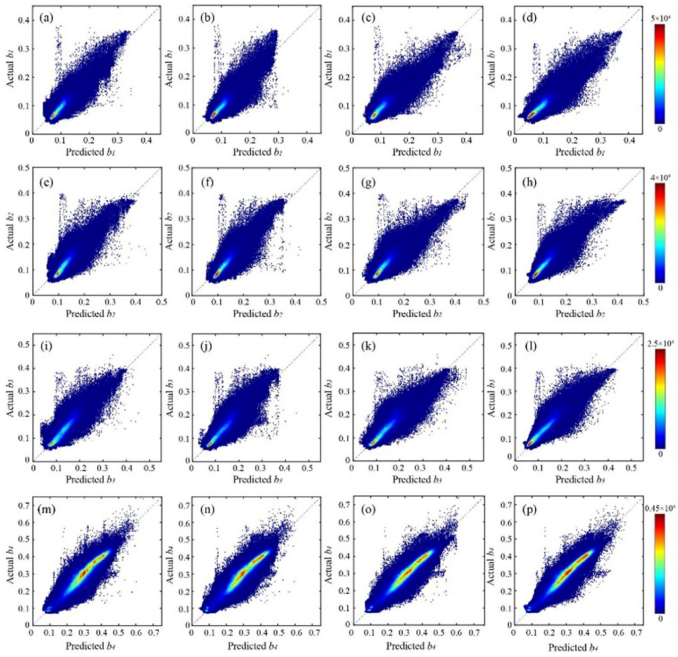

Fig. 15. Relationships of the actual and predicted values in four bands for the whole image with the four methods for the real GF∼Landsat fusion. Panels (a–d), (e–h), (i–l), and (m–p) are the plots for blue, green, red, and NIR band, respectively. The four columns from left to right correspond to the image fusion methods STARFM, FSDAF, Fit-FC, and HISTIF, respectively.

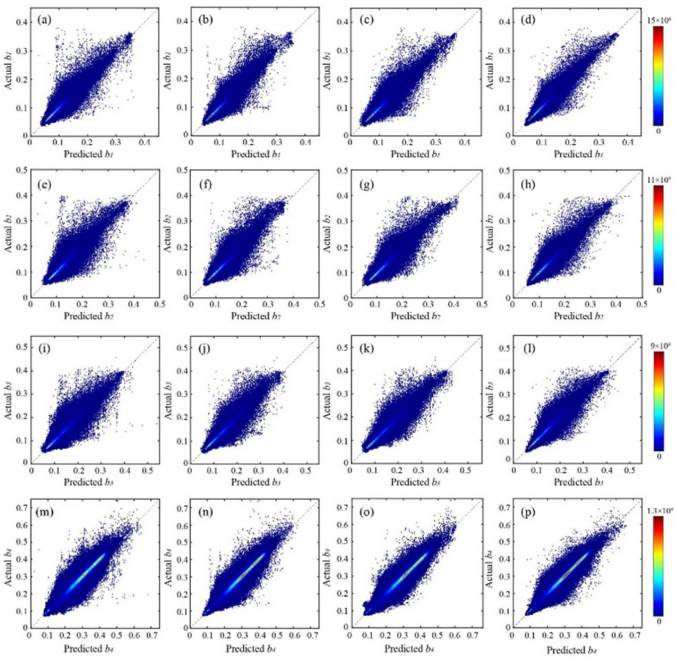

Fig. 14. Relationships of the actual and predicted values in four bands for the whole image with the four methods for the simulated GF∼Sentinel fusion. Panels (a–d), (e–h), (i–l), and (m–p) are the plots for blue, green, red, and NIR band, respectively. The four columns from left to right correspond to the image fusion methods STARFM, FSDAF, Fit-FC, and HISTIF, respectively.

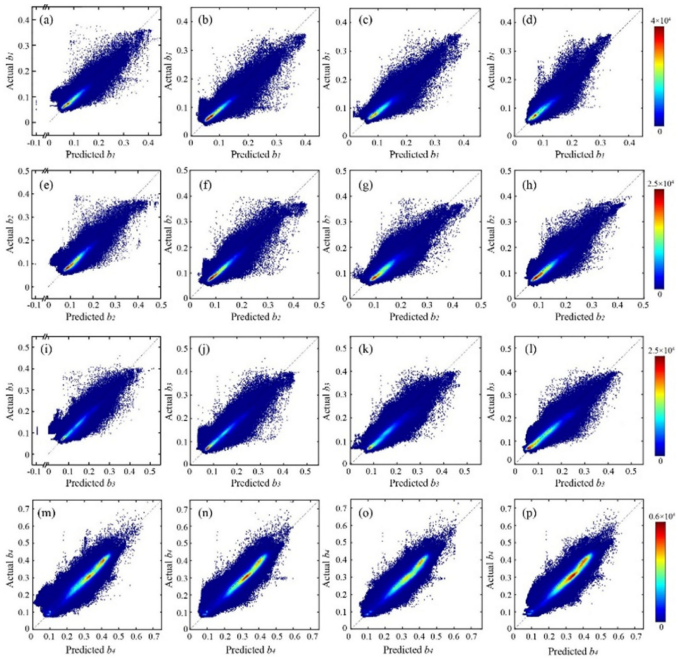

Fig. 16. Relationships of the actual and predicted values in four bands for the whole image with the four fusion methods for the real GF∼Sentinel fusion. Panels (a–d), (e–h), (i–l), and (m–p) are the plots for blue, green, red, and NIR band, respectively. The four columns from left to right correspond to the image fusion methods STARFM, FSDAF, Fit-FC, and HISTIF, respectively.

REFERENCES

[1] Y. Cai *et al.*, "A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach," *Remote Sens. Environ.*, vol. 210, pp. 35–47, 2018.

[2] Y. Yang *et al.*, "Field-scale mapping of evaporative stress indicators of crop yield: An application over Mead, NE, USA," *Remote Sens. Environ.*, vol. 210, pp. 387–402, 2018.

[3] B. Huang and H. Song, "Spatiotemporal reflectance fusion via sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3707–3716, Oct. 2012.

[4] M. Campos-Taberner *et al.*, "Multitemporal and multiresolution leaf area index retrieval for operational local rice crop monitoring," *Remote Sens. Environ.*, vol. 187, pp. 102–118, 2016.

[5] F. Gao *et al.*, "Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery," *Remote Sens. Environ.*, vol. 188, pp. 9–25, 2017.

[6] Y. Li, C. Huang, J. Hou, J. Gu, G. Zhu, and X. Li, "Mapping daily evapotranspiration based on spatiotemporal fusion of ASTER and MODIS images over irrigated agricultural areas in the Heihe River Basin, Northwest China," *Agricultural Forest Meteorol.*, vol. 244, pp. 82–97, 2017.

[7] C. Delloye, M. Weiss, and P. Defourny, "Retrieval of the canopy chlorophyll content from Sentinel-2 spectral bands to estimate nitrogen uptake in intensive winter wheat cropping systems," *Remote Sens. Environ.*, vol. 216, pp. 245–261, 2018.

[8] J. C. Jiménez-Muñoz, J. A. Sobrino, A. Gillespie, D. Sabol, and W. T. Gustafson, "Improved land surface emissivities over agricultural areas using ASTER NDVI," *Remote Sens. Environ.*, vol. 103, no. 4, pp. 474–487, 2006.

[9] J. McCarty, C. Neigh, M. Carroll, and M. Wooten, "Extracting smallholder cropped area in Tigray, Ethiopia with wall-to-wall sub-meter WorldView and moderate resolution Landsat 8 imagery," *Remote Sens. Environ.*, vol. 202, pp. 142–151, 2017.

[10] M. Burke and D. B. Lobell, "Satellite-based assessment of yield variation and its determinants in smallholder African systems," *Proc. Nat. Acad. Sci.*, vol. 114, no. 9, pp. 2189–2194, 2017.

[11] M. Jain *et al.*, "Mapping smallholder wheat yields and sowing dates using micro-satellite data," *Remote Sens.*, vol. 8, no. 10, 2016, Art. no. 860.

[12] C. M. Gevaert and F. J. García-Haro, "A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion," *Remote Sens. Environ.*, vol. 156, pp. 34–44, 2015.

[13] C. Kwan, B. Budavari, F. Gao, and X. Zhu, "A hybrid color mapping approach to fusing MODIS and Landsat images for forward prediction," *Remote Sens.*, vol. 10, no. 4, 2018, Art. no. 520.

[14] H. Song and B. Huang, "Spatiotemporal satellite image fusion through one-pair image learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 1883–1896, Apr. 2012.

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

[16] J. Amorós-López *et al.*, "Multitemporal fusion of Landsat/TM and ENVISAT/MERIS for crop monitoring," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 23, pp. 132–141, 2013.

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

[18] M. Liu, X. Liu, L. Wu, X. Zou, T. Jiang, and B. Zhao, "A modified spatiotemporal fusion algorithm using phenological information for predicting reflectance of paddy rice in southern China," *Remote Sens.*, vol. 10, no. 5, 2018, Art. no. 772.

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

[20] D. P. Roy *et al.*, "Multi-temporal MODIS–Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data," *Remote Sens. Environ.*, vol. 112, no. 6, pp. 3112–3130, 2008.

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

[22] C. Liao, J. Wang, I. Pritchard, J. Liu, and J. Shang, "A spatio-temporal data fusion model for generating NDVI time series in heterogeneous regions," *Remote Sens.*, vol. 9, no. 11, 2017, Art. no. 1125.

[23] P. Wang, F. Gao, and J. G. Masek, "Operational data fusion framework for building frequent landsat-like imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7353–7365, Nov. 2014.

[24] Q. Wang, Y. Zhang, A. O. Onojeghuo, X. Zhu, and P. M. Atkinson, "Enhancing spatio-temporal fusion of modis and landsat data by incorporating 250 m modis data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 9, pp. 4116–4123, Sep. 2017.

[25] Q. Weng, P. Fu, and F. Gao, "Generating daily land surface temperature at Landsat resolution by fusing Landsat and MODIS data," *Remote Sens. Environ.*, vol. 145, pp. 55–67, 2014.

[26] B. Wu, B. Huang, K. Cao, and G. Zhuo, "Improving spatiotemporal reflectance fusion using image inpainting and steering kernel regression techniques," *Int. J. Remote Sens.*, vol. 38, no. 3, pp. 706–727, 2017.

[27] B. Zhukov, D. Oertel, F. Lanzl, and G. Reinhackel, "Unmixing-based multisensor multiresolution image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1212–1226, May 1999.

[28] R. Zurita-Milla, J. G. Clevers, and M. E. Schaepman, "Unmixing-based Landsat TM and MERIS FR data fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 453–457, Jul. 2008.

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

[30] Y. Zhao, B. Huang, and H. Song, "A robust adaptive spatial and temporal image fusion model for complex land surface changes," *Remote Sens. Environ.*, vol. 208, pp. 42–62, 2018.

[31] C. Pohl and J. L. Van Genderen, "Review article multisensor image fusion in remote sensing: Concepts, methods and applications," *Int. J. Remote Sens.*, vol. 19, no. 5, pp. 823–854, 1998.

[32] M. Li, G. Li, W. Cai, and X.-Y. Li, "A novel pixel-level and feature-level combined multisensor image fusion scheme," in *Proc. Int. Symp. Neural Netw.*, 2008, pp. 658–665.

[33] A. Cracknell, "Review article Synergy in remote sensing-what's in a pixel?" *Int. J. Remote Sens.*, vol. 19, no. 11, pp. 2025–2047, 1998.

[34] B. L. Markham, "The Landsat sensors' spatial responses," *IEEE Trans. Geosci. Remote Sens.*, vol. GE-23, no. 6, pp. 864–875, Nov. 1985.

[35] W. Li *et al.*, "Combining hectometric and decametric satellite observations to provide near real time decametric FAPAR product," *Remote Sens. Environ.*, vol. 200, pp. 250–262, 2017.

[36] R. A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*. New York, NY, USA: Elsevier, 2006.

[37] G. Duveiller, F. Baret, and P. Defourny, "Crop specific green area index retrieval from MODIS data at regional scale by controlling pixel-target adequacy," *Remote Sens. Environ.*, vol. 115, no. 10, pp. 2686–2701, 2011.

[38] M. Mira *et al.*, "The MODIS (collection V006) BRDF/albedo product MCD43D: Temporal course evaluated over agricultural landscape," *Remote Sens. Environ.*, vol. 170, pp. 216–228, 2015.

[39] M. R. Bonyadi and Z. Michalewicz, *Particle Swarm Optimization for Single Objective Continuous Space Problems: A Review*. Cambridge, MA, USA: MIT Press, 2017.

[40] J. Kennedy, "The particle swarm: social adaptation of knowledge," in *Proc. IEEE Int. Conf. Evol. Comput.*, 1997, pp. 303–308.

[41] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. Int. Conf. Neural Netw.*, 1995, vol. 4, pp. 1942–1948.

[42] R. Poli, "Analysis of the publications on the applications of particle swarm optimisation," *J. Artif. Evol. Appl.*, vol. 2008, 2008, Art. no. 685175.

[43] 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.

[44] B. Chen, B. Huang, and B. Xu, "A hierarchical spatiotemporal adaptive fusion model using one image pair," *Int. J. Dig. Earth*, vol. 10, no. 6, pp. 639–655, 2017.

[45] C. Huang, J. R. Townshend, S. Liang, S. N. Kalluri, and R. S. DeFries, "Impact of sensor's point spread function on land cover characterization: Assessment and deconvolution," *Remote Sens. Environ.*, vol. 80, no. 2, pp. 203–212, 2002.

[46] L. Wu *et al.*, "Scaling correction of remotely sensed leaf area index for farmland landscape pattern with multitype spatial heterogeneities using fractal dimension and contextural parameters," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1472–1481, May 2018.

[47] C. Kwan, J. H. Choi, S. H. Chan, J. Zhou, and B. Budavari, "A super-resolution and fusion approach to enhancing hyperspectral images," *Remote Sens.*, vol. 10, no. 9, 2018, Art. no. 1416.

[48] C. Kwan *et al.*, "Assessment of spatiotemporal fusion algorithms for planet and worldview images," *Sensors*, vol. 18, no. 4, 2018, Art. no. 1051.

[49] H. K. Zhang, B. Huang, M. Zhang, K. Cao, and L. Yu, "A generalization of spatial and temporal fusion methods for remotely sensed surface parameters," *Int. J. Remote Sens.*, vol. 36, no. 17, pp. 4411–4445, 2015.
[50] M. Liu *et al.*, "An improved flexible spatiotemporal data fusion (IFSDAF) method for producing high spatiotemporal resolution normalized difference vegetation index time series," *Remote Sens. Environ.*, vol. 227, pp. 74–89, 2019.
[51] F. Yin, P. E. Lewis, J. Gomez-Dansand, and Q. Wu, "A sensor-invariant atmospheric correction method: Application to Sentinel-2/MSI and Landsat 8/OLI," 21 Feb. 2019. [Online]. Available: http://eartharxiv.org/ps957
[52] X. Chen, M. Liu, X. Zhu, J. Chen, Y. Zhong, and X. Cao, "Blend-then-Index" or "Index-then-Blend": A theoretical analysis for generating high-resolution NDVI time series by STARFM," *Photogrammetric Eng. Remote Sens.*, vol. 84, no. 2, pp. 65–73, 2018.
[53] J. Jiang *et al.*, "Evaluation of three techniques for correcting the spatial scaling bias of leaf area index," *Remote Sens.*, vol. 10, no. 2, 2018, Art. no. 221.
[54] Q.-B. Zhou, Q.-Y. Yu, L. Jia, W.-B. Wu, and H.-J. Tang, "Perspective of Chinese GF-1 high-resolution satellite data in agricultural remote sensing monitoring," *J. Integrative Agriculture*, vol. 16, no. 2, pp. 242–251, 2017.
[55] W. Wan *et al.*, "A lake data set for the Tibetan Plateau from the 1960s, 2005, and 2014," *Sci. Data*, vol. 3, no. 1, pp. 1–13, 2016.
[56] Q. Xing *et al.*, "High-resolution satellite observations of a new hazard of golden tides caused by floating Sargassum in winter in the Yellow Sea," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1815–1819, Oct. 2017.
[57] L. Luo *et al.*, "Identifying linear traces of the Han Dynasty Great Wall in Dunhuang using Gaofen-1 satellite remote sensing imagery and the Hough transform," *Remote Sens.*, vol. 11, no. 22, 2019, Art. no. 2711.
[58] L. Luo *et al.*, "Auto-extraction of linear archaeological traces of Tuntian irrigation canals in Miran Site (China) from Gaofen-1 satellite imagery," *Remote Sens.*, vol. 10, no. 5, 2018, Art. no. 718.

Jiale Jiang received the B.S. degree in geographic information system, and the Ph.D. degree in surveying and mapping, in 2012 and 2017, respectively, both from China University of Geosciences, Beijing, China.

Since July 2017, she has been a Postdoctoral Fellow with the National Engineering & Technology Center for Information Agriculture (NETCIA) and College of Agriculture, Nanjing Agricultural University. From 2015 to 2016, she was a Visiting Ph.D. Student with the Department of Geography, University College London, London, U.K. She has authored and coauthored 16 papers and 1 patent and the investigator of a postdoctoral project. Her research interests include scale transformation, data fusion, and crop growth monitoring based on optical remote sensing at multiscale satellites and unmanned aerial vehicles.

Qiaofeng Zhang received the B.S. degree in remote sensing from Nanjing University of Information Science & Technology, Nanjing, China, in 2018. Since September 2018, she has been working toward the M.S. degree in agricultural informatics from Nanjing Agricultural University, Nanjing.

Her research interests include image processing, multispectral remote sensing, and machine learning.

Xia Yao (Senior Member, IEEE) received the B.S. degree in agronomy, the M.S. degree in agricultural ecology, and the Ph.D. degree in crop informatics, all from Nanjing Agricultural University, Nanjing, China, in 2001, 2004, and 2009, respectively.

Since December 2015, she has been a Professor with the National Engineering & Technology Center for Information Agriculture (NETCIA) and College of Agriculture, Nanjing Agricultural University. From 2016 to 2017, she was a Visiting Scholar with the Department of Geography, University of Hawaii, Honolulu, HI, USA. She has authored and coauthored more than 60 papers, 3 monographs, and 12 patents. She supervised 12 M.S. and Ph.D. students, and investigated over 10 projects as a PI. Her research activities are built on the sophistical multi-scale platforms for acquiring timely remotely sensed data over crop fields. Her research interests focus on crop hyperspectral remote sensing, crop growth/biotic/abiotic stress/senescence monitoring based on SIF, active remote sensing (LiDAR), unmanned aerial vehicles (UAVs), and high throughput crop phenotyping.

Dr. Yao won "the Young and Middle-aged Academic Leader in Blue and Indigo Engineering" in Jiangsu province, in 2016 and dour national and province level awards. Also, she is the Reviewer of *Remote Sensing of Environment*, *Field Crop Research*, *ISPRS Journal of Photogrammetry* and *Remote Sensing*, *International Journal of Applied Earth Observation and Geoinformation*, etc. She has been serving on the Editorial Board of Remote Sensing since 2018. She was appointed as the Dean of Smart Agriculture Department since December 2019.

Yongchao Tian received the B.S. degree in agronomy, the M.S. degree in crop cultivation and farming system, and the Ph.D. degree in crop informatics, all from Nanjing Agricultural University (NAU), Nanjing, China, in 2000, 2003, and 2008, respectively.

Since December 2014, he has been a Professor with the National Engineering & Technology Center for Information Agriculture (NETCIA) and College of Agriculture, NAU. Currently, he is the executive Deputy Director with the Institute of Smart Agriculture, NAU. He has authored more than 90 papers in journals indexed by the Web of Science, and holds 26 patents. He is the Principal Investigator of several national level projects funded by the Ministry of Science and Technology and National Natural Science Foundation of China. He works mainly on crop growth monitoring and precision farming management.

Dr. Tian was the recipient of two national second-class awards for Science and Technology Advancement Technology Advancement, and won "Thousand Talents Program" in Jiangxi province, in 2018, "the Young and Middle-aged Academic Leader in Qinglan Project" in Jiangsu province, in 2014.

Yan Zhu received the Ph.D. degree in crop cultivation and farming from Nanjing Agricultural University, Nanjing, China, in 2003.

From 2012 to 2013, she was a Visiting Scientist with the Department of Agricultural and Biological Engineering, University of Florida, Gainesville, FL, USA. She is currently a Professor of information agronomy and Dean of the College of Agriculture, Nanjing Agricultural University. She is also heading the National Engineering & Technology Center for Information Agriculture (NETCIA), MOE Engineering Research Center of Smart Agriculture, MARA Key Laboratory of Crop System Analysis and Decision Making, and Jiangsu Key Laboratory for Information Agriculture. Her research has been broadly focused on smart farming theory and technology and has made significant achievements in digital simulation of crop productivity formation process, and spectral monitoring and quantitative diagnosis of crop growth indicators. She has authored more than 170 papers on Web of Science indexed journals including Nature Climate Change and Global Change Biology, and three book chapters. She holds eight national invention patents and 16 computer software copyrights. Her H-index is 29.

Dr. Zhu has won the Second-class National Award for Progress in Science and Technology twice, and the First-class Ministerial or Provincial Award for Progress in Science and Technology twice, as the second recipient. She has also won the First-class Jiangsu Provincial Agricultural Technology Extension Award as the leading recipient. She was a recipient of the Youth Science and Technology Award of China and the Young Women Scientist of China and the National Science Fund for Distinguished Young Scholars by the National Natural Science Foundation of China (2017). She is a member of the Steering Council for the Agricultural Model Intercomparison and Improvement Project (AgMIP). She has served as a Guest Editor of Special Issues in *European Journal of Agronomy and Remote Sensing*.

Weixing Cao received the Ph.D. degree in crop science from Oregon State University, Corvallis, OR, USA, in 1989.

From 1989 to 1994, he was a Postdoctoral Research Associate and then an Assistant Scientist with the College of Agricultural and Life Sciences, University of Wisconsin, Madison, WI, USA. Since 1994, he has been a Professor with the College of Agriculture, Nanjing Agricultural University, Nanjing, China. He is currently the Director of Institute of Smart Agriculture and Honorary Director of the National Engineering & Technology Center for Information Agriculture (NETCIA), Nanjing Agricultural University, the Co-convener of the Discipline Evaluation Group in Crop Science under the Degree Committee of the State Council, Head of the National Consortium of the Professional Degree in Agricultural Engineering and Information Technology. His main research interests include crop ecology, information agronomy and smart agriculture. He has accomplished outstanding achievements in the specific areas of crop system simulation and design, growth monitoring and diagnosis, and precision crop management. His academic outputs include more than 200 publications on internationally peer-reviewed journals, eight monographs and textbooks, and 41 national invention patents. In the past 25 years, he has advised a total of 156 Master's and Ph.D. students.

Prof. Cao has won the Second-class National Award for Progress in Science and Technology three times, and the First-class Ministerial or Provincial Award for Progress in Science and Technology five times. He was a recipient of the National Science Fund for Distinguished Young Scholars by the National Natural Science Foundation of China, and the Achievement Award by Crop Science Society of China.

Tao Cheng (Senior Member, IEEE) received the B.S. degree in geographic information system from Lanzhou University, Lanzhou, China, in 2003, the M. E. degree in photogrammetry and remote sensing from Peking University, Beijing, China, in 2006, and the Ph.D. degree in earth & atmospheric sciences from University of Alberta, Edmonton, AB, Canada, in 2010.

From January 2011 to December 2013, he was a Postdoctoral Scholar with the Department of Land, Air and Water Resources, University of California, Davis, CA, USA. Since December 2013, he has been a Professor with the National Engineering & Technology Center for Information Agriculture (NETCIA) and College of Agriculture, Nanjing Agricultural University, Nanjing, China. He has authored more than 50 papers on international and national journals and one book chapter, and holds two patents. He is the principal investigator of several national level projects funded by the National Key R&D Program and the National Natural Science Foundation of China. His main research interests include crop mapping and monitoring, crop phenotyping, imaging and nonimaging spectroscopy of vegetation, quantitative remote sensing, machine learning, and image analysis.

Dr. Cheng has served as a Guest Editor of a Special Issue in *Remote Sensing* and has been serving on the Editorial Board of *ISPRS International Journal of Geo-Information* since 2015. He is currently the Chair of IEEE GRSS Nanjing Chapter. He was a recipient of the Youth Science & Technology Award from the Crop Science Society of China, in 2019.