A Gaussian Kernel-Based Spatiotemporal Fusion Model for Agricultural Remote Sensing Monitoring

Yonglin Shen, Guoling Shen, Han Zhai, Chao Yang, and Kunlun Qi

***Abstract*—Time series normalized difference vegetation index (NDVI) is the primary data for agricultural remote sensing monitoring. Due to the tradeoff between a single sensor's spatial and temporal resolutions and the impacts of cloud coverage, the time series NDVI data cannot serve well for precision agriculture. In this study, a Gaussian kernel-based spatiotemporal fusion model (GKSFM) was developed to fuse high-resolution NDVI (Landsat) and low-resolution NDVI (MODIS) to produce a daily NDVI product at a 30-m spatial resolution. Considering that the NDVI curve of crop in each growing season can be characterized by Gaussian function, GKSFM used the Gaussian kernel to fit the nonlinear relationship between the high-resolution NDVI and the low-resolution NDVI, to obtain a more reasonable temporal increment. The experimental results show that GKSFM outperformed the comparative models in different proportions of cropland/noncropland and different crop phenology. In addition, GKSFM was also applied for crop mapping of Mishan County by fusing the NDVI images during the crop growing season. This study demonstrates that the accuracy of the proposed method can be improved in the midseason of crops.**

***Index Terms*—Gaussian kernel, normalized difference vegetation index (NDVI), spatiotemporal fusion, time series.**

I. Introduction

TIME series normalized difference vegetation index (NDVI), which can effectively reflect the vegetation cover, crop growth [1], [2], and crop health status [3], [4], has been widely used for crop field extraction [5], [6], crop growth monitoring [7], and yield prediction [8]–[12]. NDVI is usually derived from optical multispectral remote sensing data. However, there is a prominent tradeoff between the spatial resolution and temporal resolution of a single sensor. For example, many satellite sensors are with a high temporal resolution of one day or several days, e.g., moderate resolution imaging spectroradiometer (MODIS). Its spatial resolution ranges from 250 m to several kilometers, which is rough and cannot meet the observation requirements of precision agricultural monitoring. On the contrary, sensors, such as Landsat TM/ETM+/OLI, Sentinel 2 MSI, have a high spatial resolution of 30 or 10 m, providing more spatial details, while the single sensor mentioned above has a revisit cycle of ten or more days. What is worse, the optical satellite images are inevitably influenced by the severe cloud coverage, which further degrades the temporal resolution and quality of the NDVI data. The previous study shows that around 35% of the Landsat and Sentinel 2A/B images are covered by the cloud [13]. As a result, NDVI with high spatial and temporal resolutions is often unavailable, which limits the implementations of large-scale and full-coverage agricultural monitoring. At this point, spatiotemporal fusion is one of the effective techniques to overcome these obstacles, by integrating the high spatial resolution and high temporal resolution of different sensors feasibly and at low cost [12], [14], [15].

In recent years, many spatiotemporal fusion methods have been developed under different assumptions and application purposes [16]–[21]. The spatial and temporal adaptive reflectance fusion model (STARFM) [22] was the first proposed and most widely used spatiotemporal fusion method [23]. It utilizes a sliding window to perform temporal, spatial, and spectral weighting operations on similar pixels in the neighborhood to avoid the uncertainty caused by single-pixel prediction. STARFM performs well in homogeneous regions with stable land cover during the period of prediction. To increase the accuracy of heterogeneous regions, the enhanced spatial and temporal adaptive reflectance fusion model (ESTARFM) was proposed by Zhu *et al.* [14] to enhance the prediction by adding a pair of images and adopting a linear conversion coefficient to characterize the relationship between the high-resolution and low-resolution images. That is, ESTARFM is based on two pairs of high-resolution and low-resolution images obtained on two dates to predict the target high-resolution image. It can reduce the system biases of different sensors and reserve more spatial details. Xie *et al.* [24] improved the STARFM in aid of the unmixing method, which performs better in the heterogeneous region but is still sensitive to the change of land cover [12]. Qing *et al.* [25] introduced the idea of nonlocal filtering to predict the target image more accurately and robustly, especially for both heterogeneous regions and temporal dynamic regions, although it is based on a linear assumption, which is not accurate over a long period. Zhu *et al.* [15] integrated the STARFM, the spatial interpolation and the unmixing method into a framework, which performs better in predicting abrupt land cover changes compared with other methods. However, the prediction accuracy largely depends on the degree of land cover changes between the two dates of the input images. Luo *et al.* [26] developed an efficient method named STAIR, which includes filling and fusion

Manuscript received October 30, 2020; revised January 19, 2021 and February 15, 2021; accepted March 9, 2021. Date of publication March 17, 2021; date of current version April 7, 2021. This work was supported in part by the State Key Laboratory of Resources and Environmental Information System under Grant 2018004 and in part by the National Natural Science Foundation of China under Grant 41701410 and Grant 42001313. *(Corresponding author: Kunlun Qi.)*

The authors are with the School of Geography and Information Engineering, China University of Geosciences, Wuhan 430074, China (e-mail: shenyl@cug.edu.cn; cugsgl@cug.edu.cn; zhaihan@cug.edu.cn; yangchao@cug.edu.cn; qikunlun@cug.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2021.3066055

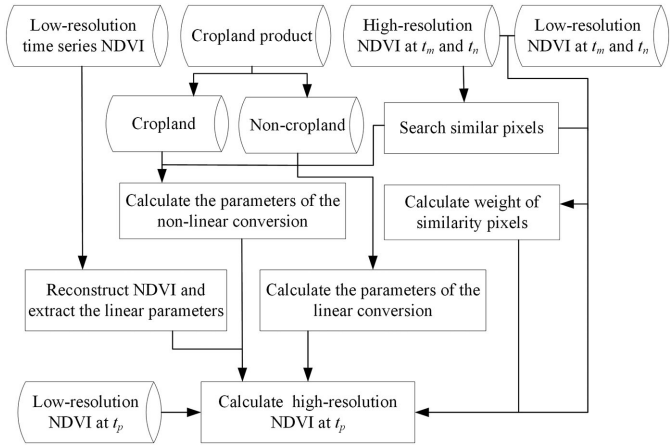

Fig. 1. Flowchart of the algorithm's implementation.

steps to generate a daily 30 m image. In addition to fused-based methods, time series-based methods, such as harmonic analysis of time series (HANTS), was used to generate a continuous time series NDVI based on high spatial resolution NDVI alone [7]. This kind of method can generate time series NDVI flexibly since they do not need information from other sensors. However, such methods are limited in areas with long periods of cloud coverage.

To sum up, most of the existing methods are based on the assumption that there is a linear relationship between the low-resolution and high-resolution image pairs [27], which ignores the nonlinear changes of NDVI as the crop progress. In this study, a Gaussian kernel-based spatiotemporal fusion model (GKSFM) was proposed to fuse high-resolution NDVI (Landsat) and low-resolution NDVI (MODIS) during the crop growing season to produce a daily NDVI product at a 30-m spatial resolution. This study verified the performance of the GKSFM for different phenological stages and different proportions of cropland/noncropland. The proposed method was also applied to produce a time series NDVI with a high spatiotemporal resolution for the crop mapping in Mishan County of Heilongjiang, China, to validate the effectiveness of the proposed method in practical application.

II. Gaussian Kernel-Based Spatiotemporal Fusion Model

There are four main steps for the implementation of GKSFM as follows (see Fig. 1).

1) Reconstruct the low-resolution time series NDVI and extract the linear parameters.
2) Search pixels that are similar to the central pixel in a local window by two high-resolution NDVI images acquired at different dates.
3) Calculate the weights and parameters of the nonlinear conversion of cropland pixels and the linear conversion of noncropland pixels.
4) Estimate the NDVI value of the center pixel on the prediction date for its cropland and noncropland.

A. Basic Principle

The NDVI of cropland changes gradually with crop growth. Existing literature and tools [28] reported the functional representations of the NDVI curve of the crop growing season, e.g., Gaussian function [29], asymmetric Gaussian function [30], double logistic method [31], etc. With the tradeoff between the number of parameters that each function needs to solve and the number of input image pairs, the Gaussian function is used to characterize the temporal changes of NDVI of crop growing season in this study, and a fusion model of GKSFM is proposed.

Due to the discrepancies of the sensor system, such as orbit pass, viewing angle, and spatial scale [32], the NDVI curves from different satellite sensors are inconsistent, mainly reflected in the inconsistency of mean and variance of Gaussian functions. Therefore, this study adopts the Gaussian function to represent two kinds of NDVI data with different spatial resolutions. The high-resolution and low-resolution NDVI data can be formulated by

$$F\ (i,j,t_k) = M_F\ \times g\,(i,j,t_k,\theta_F) + N_F \tag{1}$$

$$C\ (i,j,t_k) = M_C\ \times g\,(i,j,t_k,\theta_C) + N_C \tag{2}$$

where F represents the high-resolution NDVI data; C represents the low-resolution NDVI data; i and j are the coordinates of the image; t_k is the time; M_C, N_C, M_F, and N_F represent the linear parameters of the Gaussian function of the low-resolution NDVI data and the high-resolution NDVI data, respectively; $g(i,j,t_k,\theta_F)$ and $g(i,j,t_k,\theta_C)$ are the Gaussian kernel functions, with the mean and variance reflecting the corresponding crop phenological changes; $g(i,j,t_k,\theta_F)$ and $g(i,j,t_k,\theta_C)$ can be written as

$$g\ (x,y,t_k,\theta_F) = \exp\left\{-\left(\frac{t_k - a_F}{b_F}\right)^2\right\} \tag{3}$$

$$g\ (x,y,t_k,\theta_C) = \exp\left\{-\left(\frac{t_k - a_C}{b_C}\right)^2\right\} \tag{4}$$

where a_F and a_C are the mean of the Gaussian functions, and b_F and b_C are the variance of the Gaussian functions.

For noncropland pixels, we assume that the NDVI is stable with approximately linear changes over the prediction date. The Gaussian kernel functions $g(x,y,t_k,\theta_F)$ and $g(x,y,t_k,\theta_C)$ can be regarded as the same constant because the intraseason changes in NDVI of noncropland pixels, such as buildings and water bodies, are relatively few. From (1) and (2), the relationship of noncropland pixels between the high-resolution and low-resolution NDVI data can be deduced by

$$F\ (i,j,t_k) = \frac{M_F}{M_C}\ \cdot C\,(i,j,t_k) - N_C \cdot \frac{M_F}{M_C} + N_F. \tag{5}$$

The relationship between the high-resolution and low-resolution NDVI data of noncropland pixels is linear. Therefore, we characterized the NDVI of cropland pixels by (1) and (2), and represented the NDVI of noncropland pixels by (5).

B. Spatiotemporal Fusion for Cropland

According to the growth stage of crops, the time variable t_k in (3) and (4) can be eliminated through logarithmic transformation

$$\sqrt{-\ln g\left(x,y,t_k,\theta_F\right)} = \frac{b_C}{b_F}\sqrt{-\ln g\left(x,y,t_k,\theta_C\right)} + \frac{a_C - a_F}{b_F}. \tag{6}$$

We set $\alpha = b_C/b_F$ and $\beta = (a_C - a_F)/b_F$. Then, (6) can be written as

$$\sqrt{-\ln g\left(x,y,t_k,\theta_F\right)} = \alpha\sqrt{-\ln g\left(x,y,t_k,\theta_C\right)} + \beta. \tag{7}$$

The nonlinear relationship of cropland pixels between the high-resolution and low-resolution NDVI data can be transformed into a linear relationship through logarithmic transformation. The coefficients α and β can be obtained by regressing two pairs of pixels at t_m and t_n from the high-resolution and the low-resolution NDVI data. To make the regression coefficient more robust, the sliding window is used to select similar pixels within the window for regression [14]. Information of similar pixels is then integrated into high-resolution NDVI calculation as described in

$$F\left(x_{\frac{\omega}{2}}, y_{\frac{\omega}{2}}, t_p\right) = F\left(x_{\frac{\omega}{2}}, y_{\frac{\omega}{2}}, t_k\right) + \sum_{i=1}^{N} W_i \cdot \left(F'(x_i, y_i, t_p)\right) \tag{8}$$

where N is the number of similar pixels; W_i is the weight of ith similar pixel; t_k (k – n, or n) is the time of the input image; t_p is the time of the predicted image; ω is the size of the window; and $(x_{\frac{\omega}{2}}, y_{\frac{\omega}{2}})$ is the center pixel of the window. $F'(x_i, y_i, t_p)$ in (8) can be formulated by

$$\begin{aligned} &F'\left(x_i, y_i, t_p\right) \\ &= M_F \cdot \exp\left\{-\left(\alpha \cdot \sqrt{-\ln\left(\frac{C\left(x_i, y_i, t_p\right) - N_c}{M_c}\right)} - \beta\right)^2\right\} \\ &+ M_F \cdot \exp\left\{-\left(\alpha \cdot \sqrt{-\ln\left(\frac{C\left(x_i, y_i, t_k\right) - N_c}{M_c}\right)} - \beta\right)^2\right\}. \end{aligned} \tag{9}$$

Equation (8) means that the high-resolution NDVI of the predicted date (t_p) equals the high-resolution NDVI obtained at one time (t_n or t_m) added to the weighted sum of all similar pixel changes within the window in the low-resoution NDVI. According to (8), the high-resolution NDVI at either t_n or t_m can be used as the NDVI at a base date to estimate the high-resolution NDVI at t_p, the results are marked as $F_m(x_{\frac{\omega}{2}}, y_{\frac{\omega}{2}}, t_p)$ and $F_n(x_{\frac{\omega}{2}}, y_{\frac{\omega}{2}}, t_p)$, respectively. In the heterogeneous region, the local land cover change may lead to significant uncertainty. A reliable predicted result could be obtained by combining the two predicted results by weighting, as in

$$\begin{aligned} F\left(x_{\frac{\omega}{2}}, y_{\frac{\omega}{2}}, t_p\right) &= T_m \cdot F_m\left(x_{\frac{\omega}{2}}, y_{\frac{\omega}{2}}, t_p\right) \\ &+ T_n \cdot F_n\left(x_{\frac{\omega}{2}}, y_{\frac{\omega}{2}}, t_p\right) \end{aligned} \tag{10}$$

where T_m and T_n are weights of the two predicted results. Equation (19), which follows, is used to calculate the time weight T_m and T_n.

C. Spatiotemporal Fusion for Noncropland

The low-resolution and high-resolution NDVI for noncropland pixels can be written as (5). We assume that the land covers and other conditions do not change at the time of t_m, t_n, and t_p, thus the relationship between the high-resolution and low-resolution NDVI is stable and linear. For convenience, we set $a = \frac{M_F}{M_C}$ and $b = N_F - N_C \cdot \frac{M_F}{M_C}$. Furthermore, the relationship of NDVI in the two phases can be written by

$$F\left(x,y,t_k\right) = a \cdot C\left(x,y,t_k\right) + b \tag{11}$$

$$F\left(x,y,t_p\right) = a \cdot C\left(x,y,t_p\right) + b \tag{12}$$

where t_k (t_m or t_n) is the date of the input image; t_p is the predicted date. Thus, (11) and (12) can be deduced to

$$F\left(x,y,t_p\right) = F\left(x,y,t_k\right) + a \cdot \left(C\left(x,y,t_p\right) - C\left(x,y,t_k\right)\right). \tag{13}$$

The coefficient a can be obtained by regressing a pixel at t_m and t_n. Theoretically, the high-resolution NDVI at t_p can be predicted by the high-resolution NDVI at t_m or t_n. However, only a single pair of pixels is used for fusion, which will cause part of the spatiotemporal details to be lost and great uncertainty. For noncropland pixels, the sliding window is used again to search neighboring pixels with similar NDVI values for robust regression. Then, (13) with neighboring regression can be written by

$$\begin{aligned} F\left(x_{\frac{\omega}{2}}, y_{\frac{\omega}{2}}, t_p\right) &= F\left(x_{\frac{\omega}{2}}, y_{\frac{\omega}{2}}, t_k\right) + \sum_{i=1}^{N} W_i \cdot a \cdot \\ &\left(C\left(x_i, y_i, t_p\right) - C\left(x_i, y_i, t_k\right)\right) \end{aligned} \tag{14}$$

where ω is the size of the sliding window; W_i is the weight of the ith pixel in the sliding window.

D. Weight and Linear Parameter Calculation

In order to compare the prediction effect of Gaussian, the same weighting method as ESTARFM was used. First, we need to select similar pixels in the sliding window. The discriminant criterion of similar pixels can be judged by

$$\left|F\left(x_i, y_i, t_k\right) - F\left(x_{\frac{\omega}{2}}, y_{\frac{\omega}{2}}, t_k\right)\right| \le \sigma \cdot 2/z \tag{15}$$

where σ is the variance of the whole image; z is the number of landcover types. When a pixel of t_m and t_n in the slide window meets this condition, it is marked as a similar pixel. The weight of the similar pixel determines its contribution to the predicted result, which is determined by the similarity between the NDVI of the similar pixel and the central pixel, and the distance between them. The weight W_i can be defined as

$$W_i = \left(1/D_i\right) / \sum_{i=1}^{N} \left(1/D_i\right) \tag{16}$$

$$D_i = \left(1 - R_i\right) \cdot d_i \tag{17}$$

TABLE I
INTRODUCTION OF MODIS AND LANDSAT-8 OLI DATA

Test site	Tile number of MODIS images	Tile number of Landsat images	Date of obtained Landsat images	Date of MODIS time series NDVI
Southern Junggar Basin	H24V04	Path: 143 Row: 29	1 Apr., 2017 (DOY: 91) 3 May., 2017 (DOY:123) 4 Jun., 2017 (DOY: 155) 20 Jun., 2017 (DOY: 171) 6 Jul., 2017 (DOY: 187) 22 Jul., 2017 (DOY: 203) 7 Aug., 2017 (DOY: 219) 23 Aug., 2017 (DOY: 235) 8 Sep., 2017 (DOY: 251)	DOY: 80~300
Central Jianghan Plain	H27V05	Path: 123 Row: 39	27 Apr., 2017 (DOY: 118) 17 Jul., 2017 (DOY: 198) 18 Aug., 2017 (DOY: 230)	DOY: 80~300
Mishan County	H27V04	Path: 114 Row: 28	9 Jun., 2017 (DOY: 160) 25 Jun., 2017 (DOY: 176) 13 Sep., 2017 (DOY: 256)	DOY: 60~300

DOY represents day of year.

$$d_i = 1 + \sqrt{\left(x_i - x_{w/2}\right)^2 + \left(y_i - y_{w/2}\right)^2} / (w/2) \quad (18)$$

where d_i is defined as the Euclidean distance between the similar pixel and the central pixel; R_i is the correlation coefficient between the low-resolution and high-resolution pixels for the ith similar pixel. The smaller d_i is, the larger R_i is, and the larger the weight is. Finally, the weight is normalized according to (16). The temporal weight can be calculated according to the change magnitude detected by the low-resolution NDVI between the date t_k ($k = m$ or n) and the prediction date t_p, (19) shown at bottom of this page.

There are four unknown variables in (9), i.e., M_C, M_F, N_c, and N_F, which need additional information. Due to the high temporal resolution of the low-resolution NDVI, the least square fitting was used for the low-resolution time series NDVI reconstructed by maximum value composition (MVC) [33] and harmonic analysis [34] to eliminate noise according to (2), thus M_c and N_c can be obtained. According to (5), all stable noncropland pixels (e.g., buildings) at t_m and t_n are regressed. According to the regression coefficient, M_F and N_F can be estimated.

III. EXPERIMENTS

A. Test Sites

There are three test sites in China, i.e., the southern Junggar Basin [see Fig. 2(a)], central Jianghan plain [see Fig. 2(b)–(d)], and Mishan County [see Fig. 2(e)]. The corresponding data are listed in Table I. The first study area is mainly utilized to test the effectiveness of GKSFM for crops in different growth stages; the second study area is primarily for the validation in different proportions of cropland/noncropland; and the third study area is for crop mapping in practical applications. The details are stated in the following.

1) The first study area locates in the southern Junggar basin (44°14'29"N, 87°8'9"E), with an area of 30 × 30 km.

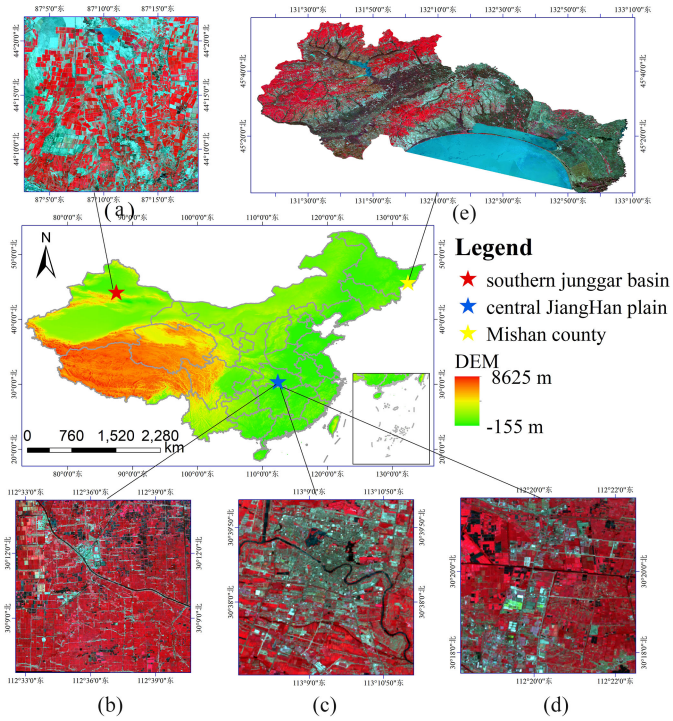

Fig. 2. Study areas. (a) Southern Junggar basin. (b)–(d) central Jianghan plain, which correspond to cropland-dominated area, noncropland-dominated area, and cropland/ noncropland equivalent area. (e) Mishan County.

$$T_k = \frac{1 \Big/ \left| \sum_{i=1}^{w} \sum_{j=1}^{w} C\left(x_i, y_j, t_k\right) - \sum_{i=1}^{w} \sum_{j=1}^{w} C\left(x_i, y_j, t_p\right) \right|}{\sum_{k=m,n} \left(1 \Big/ \left| \sum_{i=1}^{w} \sum_{j=1}^{w} C\left(x_i, y_j, t_k\right) - \sum_{i=1}^{w} \sum_{j=1}^{w} C\left(x_i, y_j, t_p\right) \right| \right)} \quad (19)$$

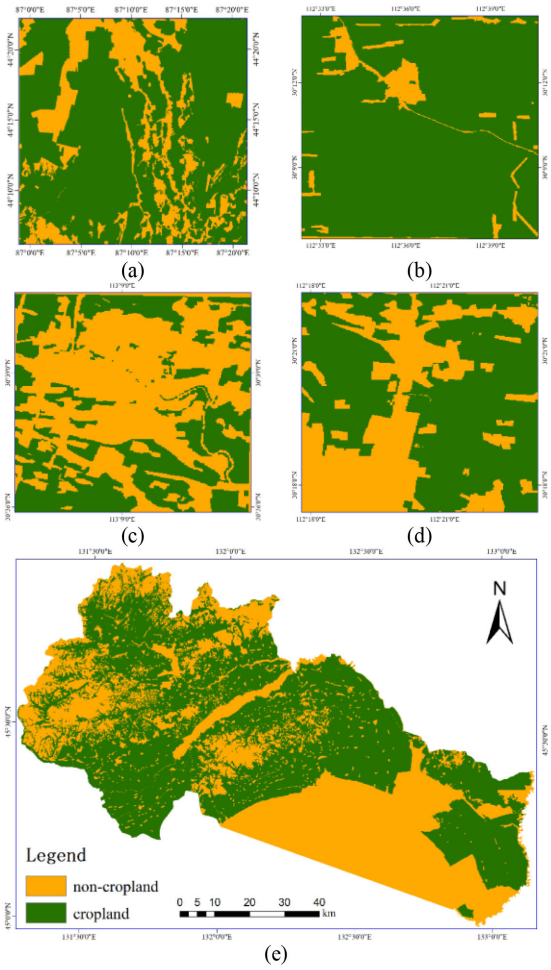

Fig. 3. Cropland maps of the study areas. (a) Junggar basin. (b)–(d) Central Jianghan plain, which correspond to cropland-dominated area, noncropland-dominated area, and cropland/ noncropland equivalent area. (e) Mishan County.

This region belongs to a temperate continental climate, with rare rainfall, large temperature difference between day and night, long and cold winter, and short and hot summer. This area is mainly consisted of croplands, with the major crop being cotton, which accounts for 76% of the total area, as shown in Fig. 3(a). Cotton is sown in middle April and harvested in October. It lasts about 185–200 days. This area also includes some other land covers, such as grassland and artificial surface.

2) The second study area locates in the center of Jianghan plain. We select three subregions that have different land cover characteristics, with each region having an area of 9 $\times$ 9 km. These three regions are the cropland-dominated area (30°40′45″N, 112°41′18″E), the noncropland-dominated area (30°38′16″N, 113°8′13″E), and the cropland/noncropland equivalent area (33°19′29″N, 112°24′40″E), as shown in Fig. 2(b)–(d); Fig. 3(b)–(d) is the corresponding cropland maps. The Jianghan plain is a plain alluvial deposit of the Yangtze River and the Han River, which is one of the major producing areas for grains and oilseed crops in China. It belongs to subtropical monsoon climate, with abundant rainfall, sufficient sunshine, and four distinct seasons. The major crop is rice. Its growing period is 130–140 days. Due to the cloud coverage, few optical remote sensing data are available for this area in 2017, as listed in Table I.

3) The third study area locates in Mishan County (45°31′47″N, 132°1′51″E) in the southeast of Heilongjiang. This area belongs to a temperate monsoon climate, with the seasons of different climate conditions. It has 257.4 km^2, with the cropland area being 143.3 km^2, as shown in Fig. 3(e). The three main crops in this study area are rice, soybeans, and corn. The crops are cultivated during a single growing season, which is mainly from April to October. The general sowing date of rice is in middle April, and the mature stage is in late September, with the average growth period being about 150–180 days. The general sowing date of soybeans is in late May, and the mature stage is in late September, with the average growth period lasting about 120–150 days. The general sowing date of corn is in middle May, and the mature stage is in middle September, with the average growth period lasting about 120–150 days. Crop calendar is derived from the Chinese National Meteorological Information Center [35]. During the crop growth period, only three Landsat 8 OLI images of different dates are available for this area in 2017, as shown in Table I.

B. Datasets and Data Preprocessing

In this study, three different kinds of data were used, including Landsat 8 OLI, MODIS surface reflectance products (MOD09GQ), and GlobeLand30. The Landsat 8 OLI image has nine bands, with a 16-day temporal frequency and a spatial resolution of 30 m for multispectral bands, which were obtained from the U.S. Geological Survey (USGS).[1] MOD09GQ is a daily reflectance product, which can be obtained from the National Aeronautics and Space Administration (NASA).[2] It has red and near-red bands, with a spatial resolution of 250 m. We selected the available Landsat OLI images of good quality under cloudless conditions (cloud cover less than 5%), and the MOD09GQ images from early April to the end of October as the data source for all study areas in 2017 (see Table I).

The MODIS images were reprojected and resampled to have the same coordinate system and spatial resolution with the Landsat 8 images. The NDVI was calculated from Landsat images and MODIS images. The Landsat NDVI was used as the high-resolution NDVI, and MODIS NDVI was used as the low-resolution NDVI. GlobeLand30 is the first global comprehensive high-resolution land cover dataset [36], with a spatial resolution of 30 m and ten different land cover types, including cultivated land, water, artificial surface, and so on, which can be downloaded from the website.[3] In this study, we took the GlobeLand30 data as the basis and determined the cultivated land as cropland area, with the other regions as noncropland areas.

[1] [Online]. Available: https://landsat.usgs.gov
[2] [Online]. Available: https://ladsweb.modaps.eosdis.nasa.gov
[3] [Online]. Available: http://www.globallandcover.com

TABLE II
FUSION ACCURACY (RMSE) OF DIFFERENT ALGORITHMS IN THE JUNGGAR BASIN

Prediction time (DOY)	Input time (DOY)	Method	Cropland	Non-Cropland	Overall
123	91 155	STARFM	0.1194	0.0902	0.1132
		ESTARFM	**0.0881**	0.0710	**0.0844**
		STAIR	0.0954	0.0755	0.0912
		GKSFM	0.0892	**0.0694**	0.0850
155	123 171	STARFM	0.1207	0.0983	0.1158
		ESTARFM	0.1170	0.0814	0.1096
		STAIR	0.1266	0.0954	0.1200
		GKSFM	**0.1090**	**0.0808**	**0.1085**
171	155 187	STARFM	0.1651	0.1163	0.1549
		ESTARFM	0.1367	**0.0823**	0.1260
		STAIR	0.1454	0.1062	0.1372
		GKSFM	**0.1288**	0.0824	**0.1195**
187	71 203	STARFM	0.1747	**0.1097**	0.1617
		ESTARFM	0.2573	0.1585	0.2377
		STAIR	0.2467	0.1630	0.2295
		GKSFM	**0.1641**	0.1641	**0.1340**
203	187 219	STARFM	0.2901	0.1598	0.2651
		ESTARFM	0.1711	**0.0974**	0.1569
		STAIR	0.3128	0.1867	0.2881
		GKSFM	**0.1649**	0.1018	**0.1523**
219	203 235	STARFM	0.0897	0.0614	0.0839
		ESTARFM	0.0919	**0.0528**	0.0843
		STAIR	0.1707	0.1215	0.1605
		GKSFM	**0.0824**	0.0550	**0.0768**
235	219 251	STARFM	0.0865	0.0619	0.0814
		ESTARFM	**0.0653**	0.0493	**0.0619**
		STAIR	0.1032	0.0899	0.1002
		GKSFM	0.0675	**0.0491**	0.0636

For convenience, we refer to the cropland area and noncropland area as cropland data.

C. Evaluation Criteria

1) Evaluation of Different Phenological Periods: NDVI images of Landsat and MODIS in Jungar basin, with a total of nine pairs, were selected to test GKSFM. Meanwhile, STAIR, STARFM, and ESTARFM were utilized as the comparative algorithms. Root-mean-square error (RMSE) was used as the evaluation indicator, which can be calculated as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(\mathrm{NDVI}_i - \widehat{\mathrm{NDVI}}_i\right)^2}{n}} \tag{20}$$

where NDVI_i represents the original Landsat NDVI; $\widehat{\mathrm{NDVI}}_i$ is the predicted NDVI value; and n is the total number of pixels.

The time series NDVI of Landsat and MODIS are shown in Table I, and the real images of the predicted date were used as the verification data, with a total of seven predicted images obtained (see Table II. Besides, because STARFM requires only a pair of high–low resolution NDVI images as input, for the sake of fairness, STARFM in this experiment also uses two pairs of images. The same weighting method as ESTARFM is adopted to obtain the predicted NDVI image.

TABLE III
FUSION ACCURACY (RMSE) OF DIFFERENT ALGORITHMS FOR DIFFERENT PROPORTIONS OF CROPLAND/NONCROPLAND IN CENTRAL JIANGHAN PLAIN

Proportion of Cropland/Non-cropland	Method	Cropland	Non-cropland	Overall
Cropland dominated	STARFM	0.1185	0.0975	0.1157
	ESTARFM	0.1149	**0.0879**	0.1121
	STAIR	0.1196	0.0917	0.1102
	GKSFM	**0.1073**	0.0886	**0.1053**
Non-cropland dominated	STARFM	0.1282	**0.0915**	0.1149
	ESTARFM	0.1239	0.0924	0.1164
	STAIR	0.1284	0.0939	0.1284
	GKSFM	**0.1154**	0.0916	**0.1097**
Cropland/non-cropland equivalent	STARFM	0.1152	0.0889	0.1106
	ESTARFM	0.1130	0.0865	0.1098
	STAIR	0.1158	0.0966	0.1134
	GKSFM	**0.1081**	**0.0864**	**0.1055**

2) Evaluation of Different Proportions of Cropland/Noncropland: NDVI images of Landsat and MODIS in central Jianghan plain, with a total of three pairs, were selected to test GKSFM. To verify the algorithm's

performance on different proportions of cropland/noncropland, three typical proportions with apparent differences in the central Jianghan plain were selected, which were the cropland-dominated region, noncropland-dominated region, and cropland/noncropland-equivalent proportion. Quantitative evaluation, regression analysis, and visual analysis were utilized to compare the algorithms' performance.

3) Evaluation by Crop Mapping: NDVI images of Landsat and MODIS in Mishan County, with a total of three pairs, were selected to fuse continuous time series NDVI for crop mapping. STARFM, ESTARFM, STAIR, and GKSFM were used for fusion to obtain continuous time series NDVI. Also, HANTS was used for generating continuous time series NDVI. Since there are only three Landsat NDVI, HANTS can only have one harmonic. These methods predicted the whole phenological period images of the time series NDVI with a 30-m resolution, from DOY 60 to DOY 300. The temporal resolution of the fused NDVI is daily. Thus, there are 240 days of NDVI in the growing season. Cloud noise interference in the original MODIS NDVI will be introduced into the fused result. Therefore, time series reconstruction on fused NDVI should be conducted. The MVC was adopted to eliminate low-value noise first. Then the Savitzky–Golay (S-G) filter was performed on it. To improve the crop mapping accuracy, fused NDVI was first masked by the existing cropland product [37]. Because most corn and soybeans were cultivated at the mountain's foot, the planting structure is scattered. Therefore, it is necessary to use the existing cropland product to carry out *a priori* constraint on the crop classification result. Many classifiers have been reported in previous literature [38]–[40]. In this study, the SVM classifier was utilized, with two-thirds of samples used for training, with the rest for testing. Meanwhile, multitemporal Landsat NDVI images were also used for crop classification. By comparing the classification accuracy, the reconstruction degree of the missing phenological information can be evaluated.

IV. Results and Discussions

A. Performance for Crop Phenology

By analyzing the fusion accuracy for cropland in Table II, the RMSEs of GKSFM in the four dates of DOY123, 155, 219, and 235 are close to that of ESTARFM. While in the midseason of crops, i.e., DOY 171, 187 and 203, RMSEs are significantly lower than that of ESTARFM, STARFM, and STAIR. That is, the performance of GKSFM is outstanding in the midseason of crops. For example, in DOY 203, the RMSE of GKSFM in cropland is 0.1649, which is the lowest among that of ESTARFM (RMSE is 0.1711), STARFM (RMSE is 0.2901), and STAIR (RMSE is 0.3128), indicating that the accuracy of GKSFM in cropland area is higher than that of the other three algorithms.

For noncropland, the RMSE of GKSFM is close to that of ESTARFM and is generally lower than that of STARFM and STAIR. RMSEs of all four algorithms increase first and then decrease. This is because the pixel of MODIS was usually mixed, which transmitted the phenological information of nearby vegetation to the predicted pixel and led to a certain deviation of the predicted NDVI from the real situation. For example, the RMSEs of GKSFM and ESTARFM for noncropland in DOY 187 are higher than that of DOY 123.

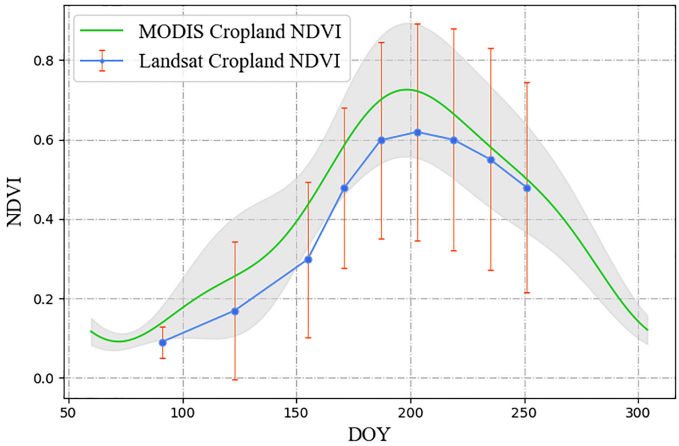

Fig. 4. NDVI curves of MODIS cropland and Landsat cropland. The grey shading denotes the NDVI standard deviation of MODIS cropland.

The overall fusion accuracy is a combination of both accuracies from cropland and noncropland areas. Because GKSFM has a significant improvement in cropland, and is close to ESTARFM in noncropland, the overall accuracy will be higher than the other linear methods, especially in the midseason of crops. Moreover, since cropland in this study area accounts for 76%, GKSFM has a significant improvement in overall fusion accuracy due to the accuracy improvement in cropland.

Fig. 4 shows the mean curve of time series NDVI of MODIS for cropland and the error bar of Landsat NDVI for cropland. The basic assumption of ESTARFM, STARFM, and STAIR is that there is a linear relationship between the high-resolution and low-resolution data. Because the change rate of cropland NDVI curve during the rising and falling phases is relatively small, the relationship between the high-resolution and low-resolution NDVI can be approximately regarded as linear (see Fig. 4). However, when the NDVI curve approaches the peak (midseason), the change rate of both high-resolution and low-resolution NDVI data is drastic, and the rate turns from positive to negative. In Fig. 4, the average NDVI change rates of Landsat cropland in the first three dates and the last three dates were relatively stable, which resulted in the performance of GKSFM being similar to that of ESTARFM, but did not affect the fusion accuracy of GKSFM for the whole growing seasons of crops. For example, RMSEs of GKSFM in DOY 123 and 235 for cropland are slightly different from that of ESTARFM, but the difference is no more than 0.003. Meanwhile, the accuracy of GKSFM in DOY 187 for cropland has a significant improvement, since the change rate of both high-resolution and low-resolution NDVI data is drastic during the midseason.

According to Fig. 4, DOY 187 is close to the peak of NDVI curve. To facilitate the analysis, we selected the local cropland area and marked it with a red box (see Fig. 5). There is a significant difference among Fig. 5(b)–(e), especially in the red circle, GKSFM [see Fig. 5(e)] performed better than the others. The absolute value of the difference between real NDVI and the

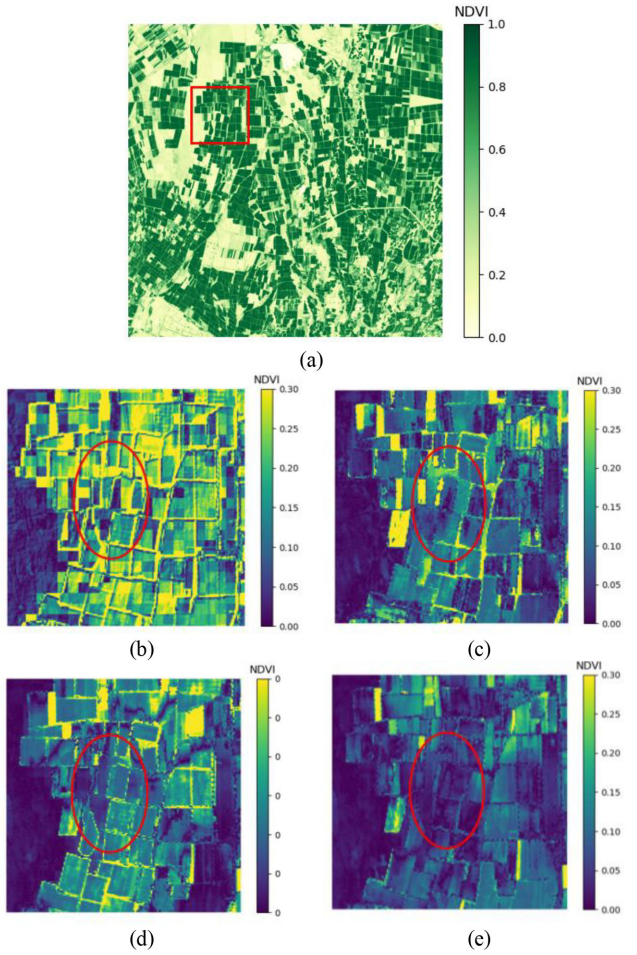

Fig. 5. Real NDVI and fused NDVI in DOY 187. (a) Real NDVI in DOY 187. (b)–(e) Corresponding local details of the absolute value of the difference between real NDVI and the predicted NDVI by STAIR, STARFM, ESTARFM, and GKSFM, respectively.

predicted NDVI by GKSFM tends to zero. The fused results near the peak of the NDVI curve are quite different from each other. Thus, the advantage of the nonlinear method (e.g., GKSFM) can be reflected. The time interval from DOY 91 to DOY 251 is the period from the planting to the harvesting of crops in this region. The NDVI curve of cropland pixels generally shows a trend of increasing first and then decreasing, which corresponds to a growth cycle of crops. Therefore, RMSEs of cropland pixels fused by ESTARFM, STARFM, and STAIR show an increasing first and then decreasing trend. Because the GKSFM algorithm assumes the nonlinear change of NDVI during the crop growing season and characterizes crop phenology more reasonably, the fusion accuracy for cropland is higher than that of ESTARFM, STARFM, and STAIR.

Because the theoretical basis of GKSFM for fusing noncropland is similar to ESTARFM, the fusion accuracy of GKSFM is close to ESTARFM and has been dramatically improved compared with STARFM. Since most of the MODIS pixels labeled as noncropland are mixed pixels, which contain information of many adjacent cropland pixels, the NDVI value of noncropland pixels will be higher than the expected value. Therefore, the noncropland pixels will transmit part phenological information, causing certain fluctuations of RMSE over time and certain errors.

B. Performance for Different Proportions of Cropland/Noncropland

To test the performance of GKSFM in different proportions of cropland/noncropland, the experiment selected three typical proportions with apparent differences in central Jianghan plain, which are cropland-dominant area, noncropland-dominant area, and cropland/noncropland equivalent area, respectively. This study used the three proportions of cropland/noncropland image pairs on DOY 118 and 230 as reference image pairs, and then, inputted MODIS NDVI on DOY 198 to predict Landsat NDVI on DOY 198. The absolute image of the difference between the predicted NDVI and the real NDVI is shown in Fig. 6.

In the cropland-dominated region, most of the land cover types are crops, such as rice. The predicted date is DOY 198, which is near the peak of the NDVI curve. In Fig. 7(a)–(d), the scatter plots of GKSFM fusion results and real image are more concentrated in the diagonal region, and R^2 (0.711) is also the highest, which indicates that the predicted result of GKSFM is closer to the real images. RMSEs of GKSFM for both cropland and the whole image are the lowest, as shown in Table III with values of 0.1073 and 0.1053, respectively. The prediction accuracy of GKSFM for noncropland pixels is very close to that of ESTARFM, only 0.8% lower.

In the noncropland-dominant region, most areas are noncropland pixels (buildings, roads, etc.), and cropland pixels only account for a few parts. As there is no phenological change for buildings and roads, the NDVI values of such types do not change much with time. Therefore, the fusion accuracies of all methods are high. Fig. 7(e)–(h) shows the scatter plots among the predicted results of STAIR, STARFM, ESTARFM, and GKSFM and the real image, in which the scatter plot of GKSFM is more concentrated in the diagonal area, and its R^2 (0.807) reaches the optimal level. GKSFM had the lowest RMSEs for cropland pixels and the whole image, as shown in Table III which are 0.1154 and 0.1097, respectively. Its RMSE (0.0916) for noncropland pixels is closed to that of STARFM (0.0915).

In the cropland/noncropland equivalent proportion, as shown in Fig. 6(k)–(o), the STARFM, ESTARFM, and GKSFM can achieve acceptable prediction results for buildings and roads, while the NDVI value is underestimated for cropland pixels. Because the cropland is too fragmented, MODIS cannot capture the subtle changes and is also affected by the adjacent buildings, roads, and other land covers, resulting in a low predicted value of some fragmented cropland. The scatter plots of Fig. 7(i)–(l) also show that NDVI is underestimated in all results, but the scatter plot of GKSFM is more concentrated in the diagonal region, and R^2 (0.797) is the highest of the four methods. In terms of RMSE, GKSFM is superior to the other three methods for both cropland pixels and noncropland pixels.

Based on the above quantitative and qualitative evaluations, it is shown that under the same weighted framework, the fusion

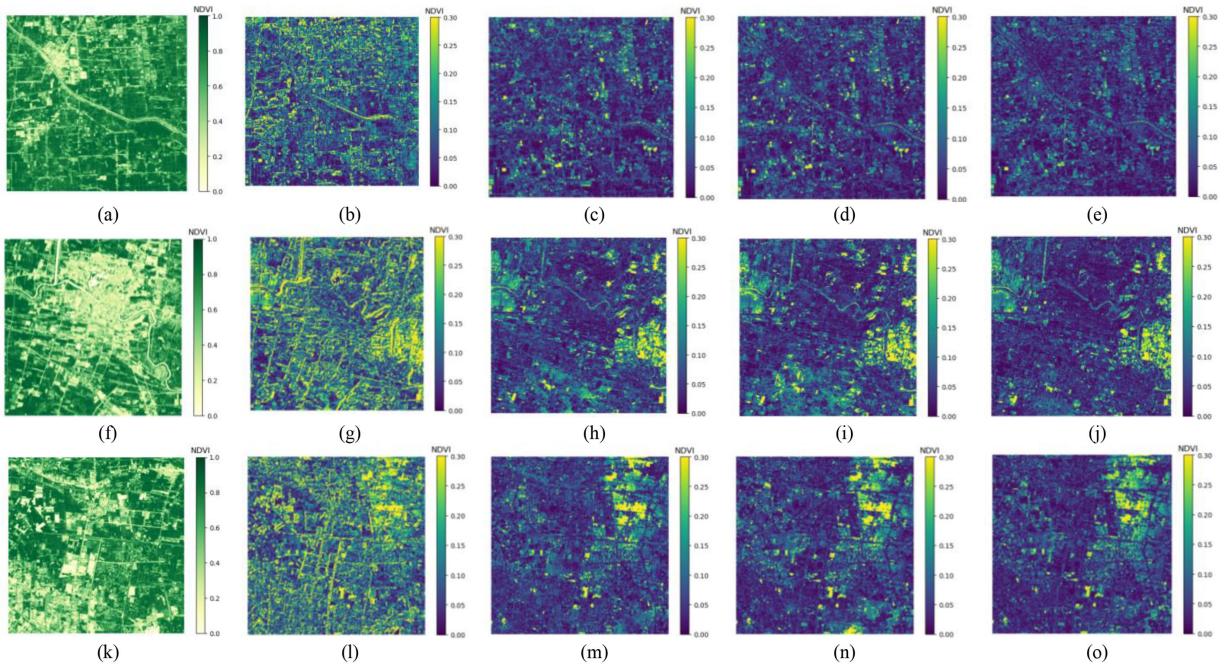

Fig. 6. Comparison of fusion results of different methods on different proportions of cropland/noncropland. (a)–(e) Real NDVI of the cropland-dominant area and the corresponding absolute NDVI of the difference between the real NDVI and fusion result of STAIR, STARFM, ESTARFM, and GKSFM, respectively. (f)–(j) Real NDVI of the noncropland-dominant area and the corresponding absolute NDVI of the difference between the real NDVI and fusion result of STAIR, STARFM, ESTARFM, and GKSFM respectively. (k)–(o) Real NDVI of cropland and noncropland equivalent area and the corresponding absolute NDVI of the difference between the real NDVI and fusion result of STAIR, STARFM, ESTARFM, and GKSFM, respectively.

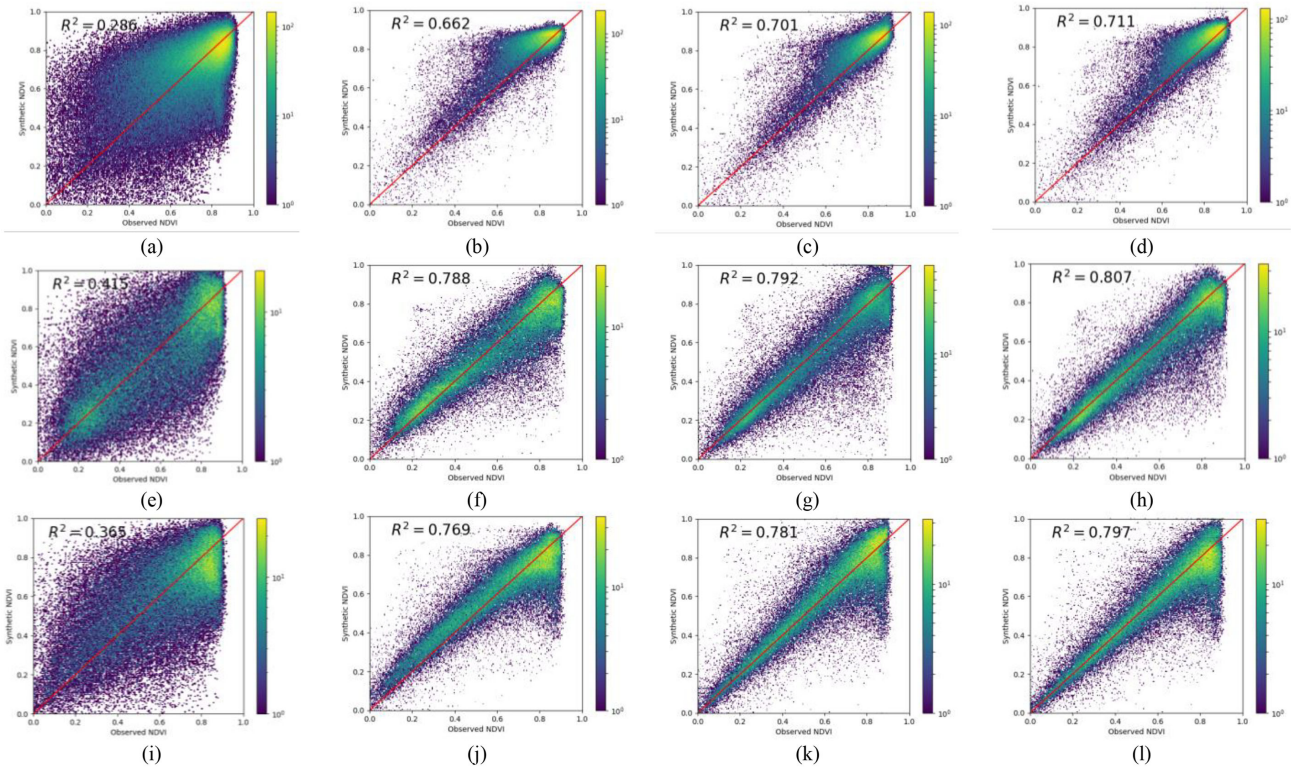

Fig. 7. Scatter plots of real and fused results in different proportions of cropland/noncropland. (a)–(d) STAIR, STARFM, ESTARFM, and GKSFM in the cropland-dominant area, respectively. (e)–(h) STAIR, STARFM, ESTARFM, and GKSFM in the noncropland-dominant area, respectively. (i)–(l) STAIR, STARFM, ESTARFM, and GKSFM in cropland and noncropland equivalent area, respectively.

accuracy of GKSFM optimized with the Gaussian kernel is higher than that of ESTARFM in cropland pixels, which proves the feasibility of the Gaussian kernel optimization method. In Fig. 7, GKSFM's scatter plot is more concentrated in the diagonal region, and R^2 reaches the highest among the four methods.

The accuracy of spatiotemporal fusion is different in different proportions of cropland/noncropland. Cropland-dominated region and noncropland-dominated region are relatively more homogeneous, the contribution of neighborhood similar pixels to the central pixels is consistent during the fusion. The cropland/noncropland equivalent proportions have considerable heterogeneity. A low-resolution pixel might contain building, water, cropland, etc. It is inevitable that information from other local land covers will be introduced into the central pixel during the fusion process, resulting in the underestimation on cropland NDVI. As shown in Fig. 7(i)–(l), the high-value area gathered some point below the diagonal. Therefore, it can be concluded that GKSFM performs better than STAIR, STARFM, and ESTARFM in both homogeneous and heterogeneous cropland.

C. Application for Crop Mapping

In this section, the characteristics of time series NDVI of different crops will be discussed, and the fused NDVI will be used to improve the accuracy of crop classification for crop mapping in Mishan County.

In Mishan County, rice, soybeans, and corn are all single-cropping crops. For rice, roughly DOY 110 to DOY 140 is the transplanting period. During this period, the surface is a mixture of rice and water so that NDVI will show a downward trend [41]. Then, with the growth of rice, NDVI will gradually increase. This is a unique feature of rice, at about DOY 225, NDVI of rice changes from rising to declining because rice gradually matures from milk stage. Meanwhile, rice husk and japonica gradually changed from green to yellow, leading to a decline in NDVI. For soybeans, the sowing stage is about DOY 120 to DOY 150, and then, NDVI gradually increased. At about DOY 225, NDVI is at the turning point from rising to falling, which is in late August, when soybeans begin to form pod and grain, green leaves gradually turn yellow, when NDVI is not obviously different from rice. For corn, the planted stage is similar to that of soybeans. NDVI turns from rising to declining trend at about DOY 240, and the dates of maturing and harvesting are later than that of soybeans and rice. For corn, the planted stage is similar to that of soybeans. Because leaves of corn are luxuriant, NDVI is relatively high in the corn growing season, and NDVI changes from rising to declining trends at about DOY 240, and the dates of maturing and harvesting are later than that of soybeans and rice.

Fig. 8 shows the time series NDVI of rice, soybeans, and corn fused by different methods. We can evaluate and analyze the results from two perspectives. One is the curve consistency with high temporal profile of MODIS NDVI. Although MODIS pixels are inevitably affected by cloud coverage, the NDVI curve after time series reconstruction (e.g., MVC and S-G filter) can reflect the overall trend to a certain extent. Another is the

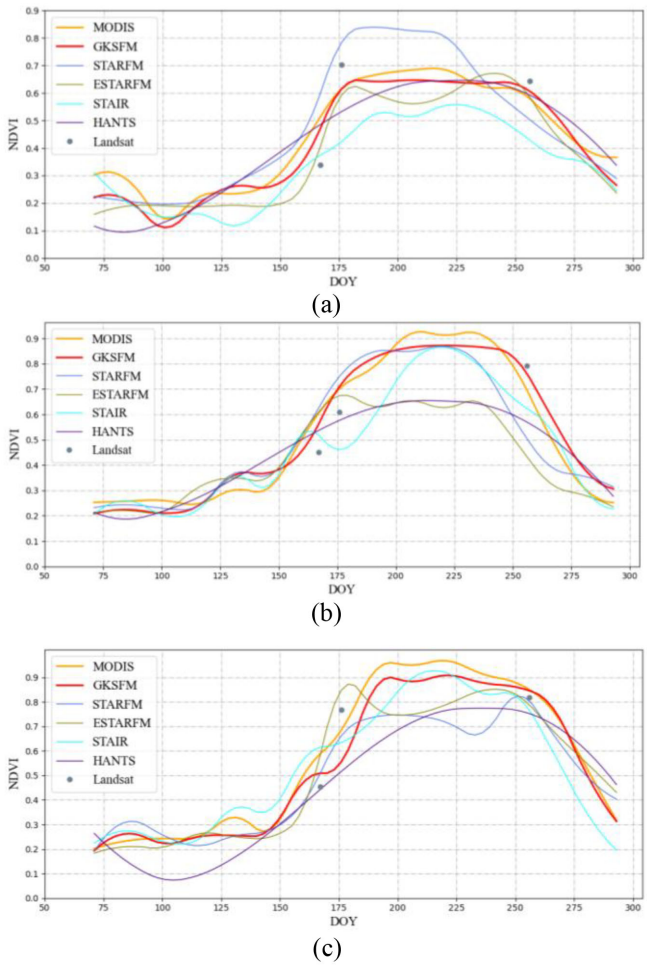

Fig. 8. 30 m Landsat NDVI, 250 m MODIS time series NDVI, and 30 m time series NDVI generated by STAIR, STARFM, ESTARFM, HANTS, and GKSFM of crop samples in Mishan. (a) Rice. (b) Soybeans. (c) Corn.

peak number. Because crops during the select time period are single-cropping, only one peak of the NDVI curve is reasonable. In Fig. 8(b) and (c), the result of HANTS is significantly underestimated because of the limited Landsat observations. The feature of peak of time series NDVI generated by ESTAFM and STAIR is not reasonable in Fig. 8, which has more than one peak. In the early and late growing seasons, the trend of NDVI generated by different methods is similar. However, there are significant differences in NDVI between different methods around midseason.

Table IV presents the classification results of the time series NDVI fused by different methods and the original Landsat NDVI, respectively. Overall accuracy (OA) of the NDVI fused by GKSFM is the highest 88.29%, and the Kappa coefficient is 0.7823. Compared with the original multitemporal Landsat NDVI and time series NDVI generated by other methods, the OA is improved from 6.09% to 23.28%. The temporal span of fused NDVI covers the whole crop growing season, which can fully reflect the growth characteristics of crops, such as the NDVI decrease first and then increase in the rice transplanting period. The time series NDVI of corn and rice has its own unique

TABLE IV
CLASSIFICATION ACCURACY OF DIFFERENT DATA IN MISHAN COUNTY

Method	Class	Ground Truth (%)			Producer Accuracy (%)	User Accuracy (%)	Overall Accuracy (%)	Kappa Coefficient
		Corn	Rice	Soybeans				
GKSFM	Corn	88.4	1.43	9.24	88.84	80.30	**88.29%**	**0.7823**
	Rice	7.60	95.26	19.98	95.26	90.59		
	Soybeans	3.56	3.31	70.78	70.78	87.33		
original Landsat NDVI	Corn	48.25	6.86	13.68	48.24	48.02	72.26	0.5225
	Rice	9.44	77.45	12.81	77.45	53.64		
	Soybeans	42.31	15.69	73.51	73.51	91.24		
ESTARFM	Corn	80.21	2.45	16.14	80.21	80.11	83.20	0.7315
	Rice	8.95	88.91	12.81	88.9	89.79		
	Soybeans	10.84	8.64	71.05	71.05	73.19		
STARFM	Corn	81.19	2.32	13.38	81.19	72.97	79.39	0.6805
	Rice	10.21	83.59	13.25	83.59	82.17		
	Soybeans	8.6	14.09	73.37	73.37	80.22		
STAIR	Corn	69.65	5.51	12.25	69.65	75.82	77.72	0.6453
	Rice	7.37	84.18	17.54	84.18	87.51		
	Soybeans	22.98	10.31	70.21	70.21	62.71		
HANTS	Corn	47.24	11.64	14.37	47.24	56.62	65.01	0.4669
	Rice	10.35	69.15	12.54	69.15	86.42		
	Soybeans	42.41	19.21	73.09	73.09	47.09		

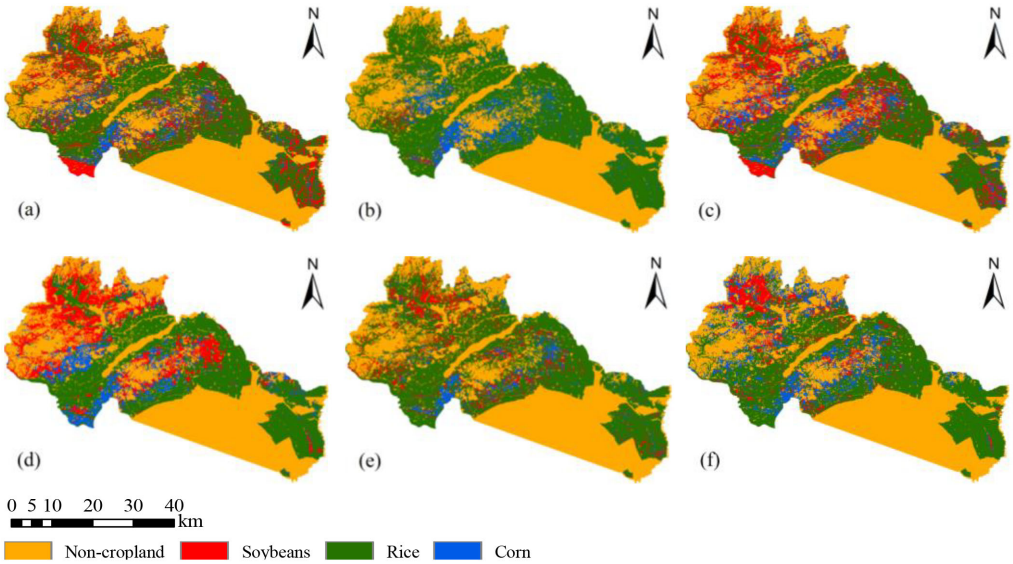

Fig. 9. Classification result. (a) Result of classifying with origin Landsat NDVI. (b)–(f) Result of classifying with NDVI fused by HANTS, STAIR, STARFM, ESTARFM, and GKSFM, respectively.

characteristics to achieve better classification results. However, due to some late sowing date of soybeans, the planted stage of soybeans is similar to that of the rice transplanting period, which is also in the low NDVI period. Therefore, the time series characteristics of soybeans and rice are relatively close, leading to the misclassification of some soybeans as rice.

According to the classification result in Fig. 9(f), most rice was planted in the middle and east of Mishan County with a large area, while corn and soybeans were mainly planted in the middle and west with a fragment planting structure. The main reason is that rice planting income is better than that of corn and soybeans, and rice planting requires much water resource from the perspective of comprehensive planting income. To reduce the cost, rice is basically planted in contiguous areas. However, in the fragment region, rice cannot be grown due to the limitation of water, soil environment, electric power, and other factors. Only corn, soybeans, and other crops can be planted.

The experiment results show that the time series NDVI fused by GKSFM guarantees the spatial resolution of NDVI, ensures the temporal variation of NDVI of crops, and restores the NDVI in crop progress stages. Classification using the time series NDVI fused by GKSFM can fully characterize the growth of crops, and achieve an OA improvement. Meanwhile, the classification result shows the cultivation structure of Mishan County,

where rice is planted in concentrated contiguous areas, soybeans and corn are grown in fragment fields due to the high profit of rice planting in the local area.

V. Conclusion

In this study, GKSFM, which employs Gaussian kernel model to characterize the temporal variation of NDVI in agricultural area, was proposed to produce NDVI data with both high spatial resolution and high temporal resolution. Compared with the linear hypothesis of ESTARFM, GKSFM supposes the crop growing season as nonlinear change, which is closer to reality. The experiments verified the performance of GKSFM for different phenological stages and different proportions of cropland/noncropland, compared with STARFM, ESTARFM, and STAIR. GKSFM has an accuracy improvement in the cropland area, especially during the midseason of crops. And the accuracy of GKSFM in noncropland pixels is comparable to that of ESTARFM. These results show that GKSFM has better fusion accuracy for time series NDVI of crops, which is of great significance for predicting the missing NDVI. GKSFM was also adopted to fuse the time series NDVI in Mishan County for crop mapping. The classification accuracy measured by OA is 88.29%, which is improved from 6.09% to 23.28% for the original Landsat NDVI and time series NDVI generated by other methods. It demonstrates that the fused time series NDVI can make up the critical missing information. More efforts are needed in the future for the improvement of the computational efficiency and for integrating other high-resolution sensors (e.g., Sentinel 2), for agricultural remote sensing monitoring.

References

[1] Z. Pan *et al.*, "Mapping crop phenology using NDVI time-series derived from HJ-1 A/B data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 34, pp. 188–197, 2015.

[2] P. Defourny *et al.*, "Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world," *Remote Sens. Environ.*, vol. 221, pp. 551–568, 2019.

[3] R. Pu, P. Gong, Y. Tian, X. Miao, R. I. Carruthers, and G. L. Anderson, "Using classification and NDVI differencing methods for monitoring sparse vegetation coverage: A case study of saltcedar in Nevada, USA," *Int. J. Remote Sens.*, vol. 29, no. 14, pp. 3987–4011, 2008.

[4] R. Suzuki, T. Nomaki, and T. Yasunari, "Spatial distribution and its seasonality of satellite-derived vegetation index (NDVI) and climate in Siberia," *Int. J. Climatol.*, vol. 21, no. 11, pp. 1321–1335, 2001.

[5] L. Yan and D. P. Roy, "Automated crop field extraction from multi-temporal web enabled Landsat data," *Remote Sens. Environ.*, vol. 144, pp. 42–64, 2014.

[6] L. Yan and D. P. Roy, "Conterminous United States crop field size quantification from multi-temporal Landsat data," *Remote Sens. Environ.*, vol. 172, pp. 67–86, 2016.

[7] Y. T. Solano-Correa, F. Bovolo, L. Bruzzone, and D. Fernández-Prieto, "Automatic derivation of cropland phenological parameters by adaptive non-parametric regression of Sentinel-2 NDVI time series," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 1946–1949.

[8] L. Busetto, M. Meroni, and R. Colombo, "Combining medium and coarse spatial resolution satellite data to improve the estimation of sub-pixel NDVI time series," *Remote Sens. Environ.*, vol. 112, no. 1, pp. 118–131, 2008.

[9] S. Bhandari, S. Phinn, and T. Gill, "Preparing Landsat image time series (LITS) for monitoring changes in vegetation phenology in Queensland, Australia," *Remote Sens.*, vol. 4, no. 6, pp. 1856–1886, 2012.

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

[11] E. M. Olexa and R. L. Lawrence, "Performance and effects of land cover type on synthetic surface reflectance data and NDVI estimates for assessment and monitoring of semi-arid rangeland," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 30, pp. 30–41, 2014.

[12] 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, pp. 1125, Nov. 2017.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[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] B. Huang and H. Zhang, "Spatio-temporal reflectance fusion via unmixing: Accounting for both phenological and land-cover changes," *Int. J. Remote Sens.*, vol. 35, no. 16, pp. 6213–6233, 2014.

[19] J. Xue, Y. Leung, and T. Fung, "A Bayesian data fusion approach to spatio-temporal fusion of remotely sensed images," *Remote Sens.*, vol. 9, no. 12, 2017, Art. no. 1310.

[20] 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.

[21] Z. Tan, P. Yue, L. Di, and J. Tang, "Deriving high spatiotemporal remote sensing images using deep convolutional network," *Remote Sens.*, vol. 10, no. 7, 2018, Art. no. 1066.

[22] 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.

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

[24] D. Xie *et al.*, "An Improved STARFM with help of an unmixing-based method to generate high spatial and temporal resolution remote sensing data in complex heterogeneous regions," *Sensors*, vol. 16, no. 2, 2016, Art. no. 207.

[25] Q. Cheng, H. Liu, H. Shen, P. Wu, and L. Zhang, "A spatial and temporal nonlocal filter-based data fusion method," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4476–4488, Aug. 2017.

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

[27] J. Yang *et al.*, "High spatial and temporal resolution vegetation index generation based on multi-source data fusion," *J. Remote Sens.*, vol. 23, no. 5, pp. 935–943, 2019.

[28] L. Eklundh and P. Jönsson, "TIMESAT for processing time-series data from satellite sensors for land surface monitoring," in *Multitemporal Remote Sensing: Methods and Applications*, Y. Ban, Ed. Cham, Switzerland: Springer, 2016, pp. 177–194.

[29] N. Cong *et al.*, "Spring vegetation green-up date in China inferred from SPOT NDVI data: A multiple model analysis," *Agricultural Forest Meteorol.*, vol. 165, pp. 104–113, 2012.

[30] P. Jonsson and L. Eklundh, "Seasonality extraction by function fitting to time-series of satellite sensor data," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 8, pp. 1824–1832, Aug 2002.

[31] N. Vorobiova and A. Chernov, "Curve fitting of MODIS NDVI time series in the task of early crops identification by satellite images," *Procedia Eng.*, vol. 201, pp. 184–195, 2017.

[32] Q. Wang, Y. Tang, X. Tong, and P. M. Atkinson, "Virtual image pair-based spatio-temporal fusion," *Remote Sens. Environ.*, vol. 249, 2020, Art. no. 112009.
[33] B. N. Holben, "Characteristics of maximum-value composite images from temporal AVHRR data," *Int. J. Remote Sens.*, vol. 7, no. 11, pp. 1417–1434, 1986.
[34] J. Zhou, L. Jia, and M. Menenti, "Reconstruction of global MODIS NDVI time series: Performance of harmonic analysis of time series (HANTS)," *Remote Sens. Environ.*, vol. 163, pp. 217–228, 2015.
[35] N. You and J. Dong, "Examining earliest identifiable timing of crops using all available Sentinel 1/2 imagery and Google Earth Engine," *ISPRS J. Photogramm. Remote Sens.*, vol. 161, pp. 109–123, 2020.
[36] J. Chen and J. Chen, "GlobeLand30: Operational global land cover mapping and big-data analysis," *Sci. China Earth Sci.*, vol. 61, no. 10, pp. 1533–1534, 2018.
[37] X. Liu *et al.*, "Large-scale crop mapping from multisource remote sensing images in Google Earth Engine," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 414–427, 2020.
[38] Y. T. Solano-Correa, F. Bovolo, and L. Bruzzone, "A semi-supervised crop-type classification based on Sentinel-2 NDVI satellite image time series and phenological parameters," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 457–460.
[39] M. Belgiu and O. Csillik, "Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis," *Remote Sens. Environ.*, vol. 204, pp. 509–523, 2018.
[40] Y. T. Solano-Correa, F. Bovolo, L. Bruzzone, and D. Fernández-Prieto, "Spatio-temporal evolution of crop fields in Sentinel-2 satellite image time series," in *Proc. 9th Int. Workshop Anal. Multitemporal Remote Sens. Images*, 2017, pp. 1–4.
[41] J. Dong *et al.*, "Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine," *Remote Sens Environ*, vol. 185, pp. 142–154, 2016.

Yonglin Shen received the B.Sc. degree in geography information system from the Wuhan University of Technology, Wuhan, China, in 2006, the M.Sc. degree in cartography and geography information system from Nanjing Normal University, Nanjing, China, in 2009, and the Ph.D. degree in cartography and geography information system Beijing Normal University, Beijing, China, in 2013.

During 2011–2013, he was a Joint Ph.D. Student with the Center for Spatial Information Science and Systems, George Mason University, Fairfax, VA, USA. He is currently an Associate Professor with the China University of Geosciences, Wuhan, China. His research interests include remote sensing applications in agriculture and atmospheric environment.

Guoling Shen received the B.Sc. degree, in 2019, from the School of Geography and Information Engineering, China University of Geosciences, Wuhan, China, where he is currently working toward the M.Sc. degree in remote sensing of environment and natural resources.

Han Zhai received the B.Sc. degree from the School of Surveying and Mapping, Shandong University of Science and Technology, Qingdao, China, in 2014, and Ph.D. degree from the State Key Laboratory of Information Engineering in Surverying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China, in 2019.

He is an Associate Professor at the China University of Geosciences, Wuhan, China. His research interests include remote sensing image processing, remote sensing information extraction, and sparse representation.

Chao Yang received the Ph.D. degree from Wuhan University, Wuhan, China in 2012.

He is currently an Associate Professor with the China University of Geosciences, Wuhan, China. His research interests include data mining, text analytics, and social media analytics for human geography studies.

Kunlun Qi received the B.S. degree from Henan University, Kaifeng, China, in 2009, and the M.S. and Ph.D. degrees from Wuhan University, Wuhan, China, in 2011 and 2016, respectively.

He is currently an Associate Professor with the China University of Geosciences, Wuhan, China. His research interests include deep learning, remote sensing image analysis, and geographic information system.