ISPRS Journal of Photogrammetry and Remote Sensing 167 (2020) 214–229

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Generating spatiotemporally consistent fractional vegetation cover at different scales using spatiotemporal fusion and multiresolution tree methods

Bing Wang[a,b], Kun Jia[a,b,*], Xiangqin Wei[c], Mu Xia[a,b], Yunjun Yao[a,b], Xiaotong Zhang[a,b], Duanyang Liu[a,b], Guofeng Tao[a,b]

[a] State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China
[b] Beijing Engineering Research Center for Global Land Remote Sensing Products, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China
[c] Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China

ARTICLE INFO

Keywords:
Fractional vegetation cover
Landsat-7 ETM+
GLASS FVC
Spatiotemporal fusion
Multiresolution tree

ABSTRACT

Fractional vegetation cover (FVC) is considered one of the most important vegetation parameters and is relevant to characterizing vegetation status and ecosystem function. An FVC with a fine spatial resolution of 30 m is essential for monitoring vegetation change and regional studies, while an FVC with a coarse spatial resolution of hundreds to thousands of metres plays an important role in global change studies. However, high spatial resolution data usually have low temporal resolution and are often affected by cloud cover. The objective of this study is to propose a practical way to generate spatiotemporally consistent FVC products at Landsat and Moderate Resolution Imaging Spectroradiometer (MODIS) scales, which are 30 m and 250 m, respectively. The geostatistical neighbourhood similar pixel interpolator (GNSPI) was first used to fill in the missing values caused by unscanned gaps and clouds/shadows on Landsat-7 Enhanced Thematic Mapper Plus (ETM+) data and to generate spatially continuous Landsat reflectance. Then, the enhanced spatial and temporal adaptive reflectance fusion model (ESTARFM) was used to generate time series Landsat reflectance data with the same temporal resolution as that of Global LAnd Surface Satellite (GLASS) FVC generated from MODIS data. The high temporal resolution Landsat reflectance was preliminarily used to estimate FVC at the Landsat scale. Finally, MultiResolution Tree (MRT) was employed to fuse the Landsat FVC and GLASS FVC to generate spatiotemporally consistent FVC products at different scales. The results show that the missing Landsat-7 ETM+ data were filled well and spatial texture features were well preserved. The temporal resolutions of the Landsat and GLASS FVC products became consistent with an interval of one day at most. After MRT fusion, most of the root mean square error (RMSE) between the GLASS FVC and aggregated Landsat FVC dramatically decreased. The accuracy of the Landsat FVC validated by the ground-measured FVC improved after MRT fusion (before MRT: RMSE = 0.1031, R^2 = 0.9172, bias = −0.0697; after MRT: RMSE = 0.0958, R^2 = 0.9173, bias = −0.054). In addition, in the GNSPI-filled unscanned gaps and the ESTARFM-generated images, the Landsat FVC accuracy also improved slightly (before MRT: RMSE = 0.1065, R^2 = 0.9011, bias = −0.0644; after MRT: RMSE = 0.1022, R^2 = 0.9023, bias = −0.051). The accuracy of the GLASS FVC also improved (before MRT: RMSE = 0.0913, R^2 = 0.884, bias = −0.0504; after MRT: RMSE = 0.0673, R^2 = 0.9483, bias = −0.0444). Therefore, MRT could decrease the inconsistencies of different scales and reduce uncertainties in the FVC. In addition, MRT could fill in the missing data of the Landsat FVC directly, but there were a certain number of outliers in the fusion results, and the spatial transition was poor.

1. Introduction

Fractional vegetation cover (FVC) is defined as the proportion of the vertical projected area of green vegetation to the total statistical area, which is recognized as one of the most important parameters for describing terrestrial ecosystems and monitoring vegetation conditions (Gao et al., 2020; Wang et al., 2018). FVC has a significant influence on the exchange of carbon, water and energy at the land surface (Wang

* Corresponding author at: State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China.
E-mail address: jiakun@bnu.edu.cn (K. Jia).

https://doi.org/10.1016/j.isprsjprs.2020.07.006
Received 22 July 2019; Received in revised form 18 March 2020; Accepted 15 July 2020

et al., 2017b). FVC has also been applied to many land surface process simulations and global change studies, such as soil erosion monitoring, drought monitoring, desertification evaluation, hydrological simulation, agricultural monitoring (de Asis and Omasa, 2007; Jia et al., 2016; Matsui et al., 2005; Zhang et al., 2013), the Earth's energy balance and climate change (Gutman and Ignatov, 1998; Roujean and Lacaze, 2002). Therefore, accurate, timely and continuous FVC products at global and regional scales are of great significance.

Remote sensing is the most effective way to estimate FVC due to its advantages of extensive coverage and repeated observations (Liang et al., 2012). However, there is a contradiction between the temporal resolution and spatial resolution of remote sensing data; data with fine spatial resolution usually have a long revisit cycle, and vice versa (Bolton and Friedl, 2013). Currently, there are several large-scale FVC products with coarse spatial resolution, such as the Envisat Medium Resolution Imaging Spectrometer (MERIS) (Bacour et al., 2006), the Carbon Cycle and Change in Land Observational Products from an Ensemble of Satellites (CYCLOPES) (Baret et al., 2007), the Geoland-2/BioPar version 1 (GEOV1) (Baret et al., 2013) and the GLASS (Jia et al., 2015) FVC products. The spatial resolutions of these products range from a few hundred metres to several kilometres. High temporal resolutions within 10 days allow these FVC products to capture rapid changes in vegetation on the land surface. Furthermore, the spatial-temporal continuity of these products makes them capable of monitoring vegetation change over large areas and conducting global change analysis (Yang et al., 2018; Yu et al., 2018). However, coarse spatial resolution data are often inadequate for highly heterogeneous areas such as agricultural landscapes because they cannot capture the detailed spatial distribution and vegetation variation patterns, which are crucial for crop classification, crop growth monitoring and yield estimation (Azzari and Lobell, 2017; Doraiswamy et al., 2004). Therefore, high spatial resolution data are vital for regional studies, such as precision agriculture and ecosystem function monitoring (Lobell and Asner, 2003; Röder et al., 2008).

Landsat data have a fine spatial resolution of 30 m, providing sufficient details for land surface variation, which is an appropriate scale to reflect human activities (Gao et al., 2015). However, the 16-day revisit cycle and frequent cloud contamination make it difficult to directly utilize Landsat data to detect rapid vegetation change, such as monitoring the critical periods of crop growth (Bolton and Friedl, 2013). In fact, there are few cloud-free data in most areas, and some of them have to be discarded due to heavy cloud cover, which causes the discontinuity of land surface information in time series and space. Moreover, sensor failures (such as ETM+) make the application of Landsat data more challenging (Zhu et al., 2012a). Therefore, FVC estimated from high spatial resolution data is usually spatially and temporally discontinuous, which limits FVC application in land surface process simulations and ecosystem modelling research (Bian et al., 2017).

Many spatiotemporal fusion algorithms have been developed to weaken the influence of clouds and obtain high spatial and high temporal observations with Landsat satellites. The spatial and temporal adaptive reflectance fusion model (STARFM) takes advantage of the high temporal resolution of MODIS and high spatial resolution of ETM + to produce high spatial resolution and high frequency data (Feng et al., 2006). Zhu et al. (2010) improved the STARFM and proposed the ESTARFM algorithm by using two pairs of high- and low-resolution images, which produces a more accurate prediction of surface reflectance in heterogeneous landscapes and has been widely used (Dong et al., 2016; Fu et al., 2014; Yan et al., 2018). In regard to the unscanned gaps of Landsat-7 ETM+, several algorithms have been proposed to fill the missing data, such as the neighbourhood similar pixel interpolator (NSPI) (Chen et al., 2011) and GNSPI (Zhu et al., 2012b), which can recover the missing spectral information with good precision. Although such spatiotemporal algorithms can increase the spatial and temporal continuity while observing fine land surface variations compared with that of medium spatial resolution data, uncertainties will also be introduced to the reconstructed surface reflectance, which are caused by a series of assumptions that simplify the complex land surface conditions. GNSPI assumes that neighbouring similar pixels have similar temporal changing patterns (Zhu et al., 2012b), while ESTARFM assumes that reflectance changes linearly during a short period and that both the proportion and reflectance change rate of each endmember are stable (Zhu et al., 2010). Therefore, the accuracy of subsequent FVC estimation will be affected (Deng et al., 2019). In addition, different FVC estimation methods commonly lead to different results due to different algorithm mechanisms, such as empirical methods and pixel unmixing model and physical-based methods (Jia et al., 2017; Jiapaer et al., 2011; Zou et al., 2018). Moreover, FVC estimates derived from multiresource satellite data are usually spatially and temporally inconsistent, particularly across different spatial resolutions. Therefore, generating spatiotemporally consistent FVC products across different scales is of great significance and potential for related research.

To reduce the uncertainties and overcome the problem of data inconsistency across different spatial resolutions, an MRT method has been developed based on the assumption that a statistical model is autoregressive in its levels of resolution (Chou et al., 1994b). MRT has been used for a series of remote sensing issues, especially for estimating satellite-based variables using mass data because it can provide an optimal estimation with efficient computation. Parada and Liang (2004) assimilated different resolution near-surface soil moistures and demonstrated that MRT could recover relevant spatial features and significantly reduce the RMSE. He et al. (2014) fused three surface albedo products from Multiangle Imaging Spectroradiometer (MISR), MODIS and Landsat to generate consistent albedo data at different spatial resolutions. Shi et al. (2016) integrated advanced spaceborne thermal emission and reflection radiometer (ASTER) and GLASS broadband emissivity (BBE) products to obtain better BBE products. However, few studies have explored the potential of MRT for integrating high and medium spatial resolution vegetation parameters, including FVC. Considering the advantages of MRT in improving accuracy, reducing uncertainty and minimizing bias across different spatial resolutions, this method is explored to generate spatiotemporally consistent FVC products from Landsat FVC and GLASS FVC products in this study.

There are two objectives in this study: (1) reduce the spatial and temporal inconsistency between Landsat and GLASS FVC products and generate spatiotemporally consistent FVC products at 30 m and 250 m scales; and (2) improve the accuracy of FVC products, especially for the data reconstructed by the spatiotemporal fusion algorithm. To achieve these goals, the unscanned gaps of ETM+ were first filled using the GNSPI method, and then the ESTARFM was used to generate temporally consistent reflectance data between Landsat and MODIS, which was subsequently used to estimate FVC. Finally, MRT was investigated to produce spatially consistent FVC products across different spatial resolutions. In contrast, the missing Landsat FVC data were merged by GLASS FVC directly without filling in the unscanned gaps of Landsat reflectance data to test the ability of MRT to directly fill in missing values.

2. Method

The flow chart of generating spatiotemporally consistent FVC products at different scales is shown in Fig. 1. In the data pre-processing step, the function of mask (Fmask) algorithm (Zhu and Woodcock, 2012) was used to detect clouds and cloud shadows on Landsat-7 ETM + reflectance data, and then the cloud-contaminated pixels and unscanned gaps were filled by the GNSPI algorithm together. When data pre-processing was completed, the ESTARFM was first implemented to generate 30 m resolution reflectance with a temporal resolution of approximately 8 days by using Landsat-7 ETM+ reflectance data and

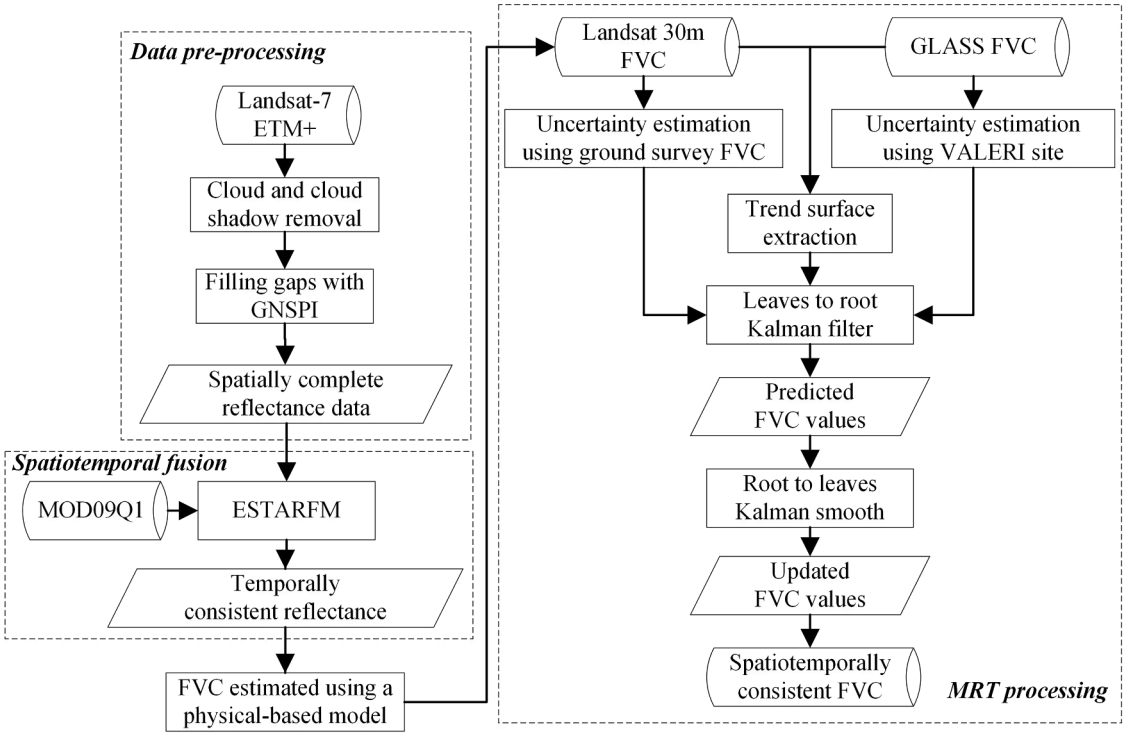

Fig. 1. Flow chart of the spatiotemporally consistent FVC generating method.

MODIS Terra Surface Reflectance 8-Day L3 Global 250 m product (MOD09Q1). The MODIS data acquisition dates were one day later than those of Landsat, and this time difference could be ignored. Then, the temporal resolution between Landsat and MODIS reflectance became consistent. Next, the PROSPECT + SAIL (PROSAIL) canopy radiative transfer model (Jacquemoud et al., 2009) was used to generate the simulated reflectance of Landsat-7 ETM+ and the corresponding FVC using various combinations of input variables. The neural network was trained using these simulated data, and the Landsat FVC was estimated based on the trained neural network model (Jia et al., 2016). Finally, MRT was used to merge the Landsat FVC and GLASS FVC to make them consistent in spatial pattern.

2.1. ESTARFM

ESTARFM was first proposed by Zhu et al. (2010) in 2010, and to date, it is one of the most widely used spatial-temporal data fusion algorithms. The algorithm was employed to blend multisource satellite data and achieved satisfactory results, such as the fusion of Landsat-5/7/8 and MODIS, Landsat-8 and Visible Infrared Imaging Radiometer Suite (VIIRS), RapidEye and MODIS, and GaoFen-1 and MODIS (Dong et al., 2016; Huang et al., 2016; Ma et al., 2018; Tao et al., 2019; Tewes et al., 2015; Wang et al., 2017a), which demonstrated the good stability and applicability of the algorithm. Therefore, ESTARFM was also utilized in this study to blend the red and near-infrared (NIR) bands of the Landsat reflectance and MOD09Q1 data to improve the temporal resolution of Landsat.

ESTARFM requires two pairs of Landsat and MOD09Q1 images whose acquisition times (t_m and t_n) are earlier and later than the prediction time (t_p), respectively, to calculate a conversion coefficient v to represent the ratio of change in Landsat to change in MOD09Q1 data. If v and MOD09Q1 at the prediction time are known, the predicted Landsat image can be calculated by Eq. (1):

$$F(x, y, t_p) = F(x, y, t_0) + v(x, y) \times (C(x, y, t_p) - C(x, y, t_0)) \quad (1)$$

where F and C are the Landsat and MOD09Q1 reflectances, respectively; (x, y) is the location of a given pixel; and t_0 is the acquisition time of the base image. The conversion coefficient v is calculated by linear regression of the reflectance changes of all the similar Landsat and resampled MOD09Q1 pixels within the coverage of a MOD09Q1 pixel. However, Eq. (1) only uses the information of a single pixel. To take advantage of information from neighbouring pixels, a moving search window is used to select pixels similar to a given pixel based on the spectral similarity, and the information of similar pixels is integrated into Landsat reflectance. Thus, Eq. (1) can be modified as Eq. (2):

$$F(x_{w/2}, y_{w/2}, t_p) = F(x_{w/2}, y_{w/2}, t_0) + \sum_{i=1}^{N} W_i \times V_i \times (C(x_i, y_i, t_p) - C(x_i, y_i, t_0)) \quad (2)$$

where w is the search window size, N is the number of similar pixels, (x_i, y_i) is the location of the ith similar pixel, W_i is the weight of the ith similar pixel, and V_i is the conversion coefficient of the ith similar pixel. Considering that two Landsat images at t_m and t_n can be used as base images and generate two predicted Landsat images at t_p, the Landsat result is the weighted combination of the two prediction results using Eq. (3):

$$F(x_{w/2}, y_{w/2}, t_p) = T_m \times F_m(x_{w/2}, y_{w/2}, t_p) + T_n \times F_n(x_{w/2}, y_{w/2}, t_p) \quad (3)$$

where F_m and F_n are the Landsat images generated by base images at t_m and t_n, respectively, and T_m and T_n are the time weights of F_m and , respectively. The detailed introduction of the ESTARFM algorithms can be found in (Zhu et al., 2010).

2.2. MRT

The MRT algorithm was initially introduced by Chou et al. (1994a), and its kernel is Kalman filtering (Kalman, 1960). Different from conventional Kalman filtering, which updates model variables in the temporal domain, MRT applies Kalman filtering in the spatial domain. MRT is organized as a linear tree structure (Fig. 2) with the assumption that FVC at different spatial resolutions is autoregressive. Similar to the dynamic model of conventional Kalman filtering, MRT also needs an equation, named the state conversion model, to represent the transition

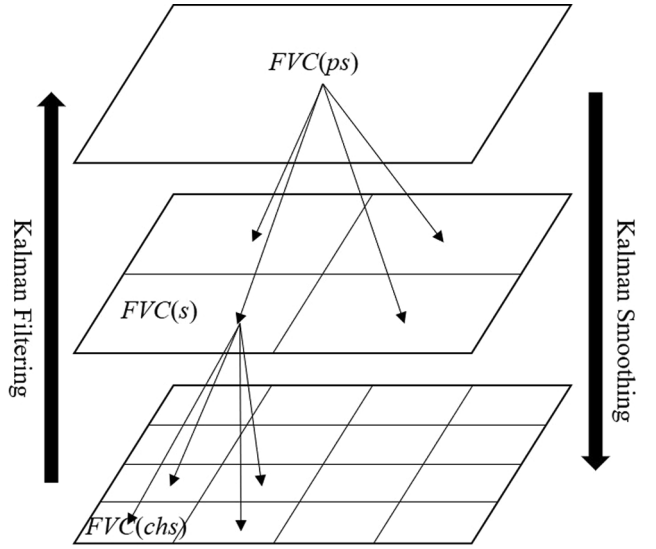

Fig. 2. Conceptual structure of MRT; node *FVC*(*ps*) is the parent node of node *FVC*(*s*), and *FVC*(*chs*) is the child node of *FVC*(*s*).

from coarse to fine spatial resolution FVC:

$$FVC(s) = A(s)FVC(ps) + w(s) \tag{4}$$

where *FVC*(*s*) and *FVC*(*ps*) are the state variables at scale *s* and its parent scale *ps*, respectively; *A*(*s*) is the state transition matrix from the parent scale *ps* to its child scale *s*, and is generally set to an identity matrix because the results of aggregation and decomposition between different scales in the same region are invariable; and *w*(*s*) is the white noise that represents the state transition error, which follows a Gaussian distribution $N(0, Q(s))$. Like Eq. (4), Eq. (5) is also a state conversion model that transfers the state variable from fine to coarse spatial resolution data:

$$FVC(ps) = F(s)FVC(s) + w'(s) \tag{5}$$

where *F*(*s*) is the state transition matrix from child scale (fine spatial resolution) to parent scale (coarse spatial resolution), which can be calculated by the following equation (Luettgen and Willsky, 1995):

$$F(s) = P(ps)A(s)P^{-1}(s) \tag{6}$$

where *P*(*ps*) and *P*(*s*) are the variances at the parent and children scales, respectively.

Except for the state conversion model, MRT also needs an equation, namely, an observation model, to link the observation variable and the state variable:

$$y(s) + H \bullet FVC(s) + v(s) \tag{7}$$

where *y*(*s*) is the satellite observation data, *v*(*s*) is the measurement error that follows a Gaussian distribution $N(0, R(s))$, and *H* is the observation matrix that maps the state variable into the measurement domain. Because the state variable and satellite data are the FVC of the same area at the same time, *H* is taken as an identity matrix.

Equations (5), (6) and (7) describe the basic concept of the dynamic process in MRT. Then, Kalman filtering is used to derive the optimal estimator by incorporating the state conversion model and observational data. From children to parents, when observations are available, the optimal estimator at scale *ps* is calculated:

$$\hat{F}VC(ps|ps) = \hat{F}VC(ps|s) + K(ps)(y(ps) - H \bullet \hat{F}VC(ps|s)) \tag{8}$$

where *K*(*ps*) is the Kalman gain and is derived from the following:

$$K(ps) = P(ps|s)HV^{-1}(ps) \tag{9}$$

where *V*(*ps*) is the innovation covariance, which is expressed as follows:

$$V(ps) = HP(ps|s)H^T + R(ps) \tag{10}$$

At the scale *ps*, the variance *p*(*ps*|*ps*) is updated:

$$P(ps|ps) = (I - K(ps)H)P(ps|s) \tag{11}$$

where *I* is an identity matrix. If observations are unavailable, the process evolves depending only on the state conversion model. Thus, the upward loop of Kalman filtering can be performed recursively, and observations are assimilated sequentially during this forward integration. Between every two adjacent scales of MRT, each child node provides an estimator to its parent node, and the optimal estimator $\hat{F}VC(ps|ps)$ is the weighted sum of the estimates of each child node. The weight of the estimate depends on the variance $P^*(s)$, which represents the uncertainty of an estimate; therefore, a node with larger uncertainty will contribute less to its parent node (Xu et al., 2019). To make full use of the information of the whole sequence of MRT at different resolutions, when the ascending filtering process reaches the root of the tree, Kalman smoothing is implemented to obtain accurate estimators $\hat{F}VC'(s)$. The details of Kalman smoothing can be found in Kannan et al. (2000).

MRT consists of two steps: “leaves-to-root” Kalman filtering and “root-to-leaves” Kalman smoothing. High-to-low resolution filtering is first implemented to provide finer information to the coarser resolution and to estimate the state variable from higher-resolution data. This step aims to fill the gaps at different resolutions. Low-to-high resolution smoothing is then applied to update the state variable with the assumption that the parent node provides the fundamental information for children nodes that it directly links to. After Kalman smoothing, the datasets at different spatial resolutions become smooth and consistent. More details of MRT can be found in Van de Vyver and Roulin (2009).

To implement MRT for different resolution satellite FVC products, there are four steps. (1) Evaluate and quantify the uncertainty of the two satellite FVC products. There is no Landsat FVC product available at the global scale. The Landsat FVC is usually generated by different methods and used for regional studies individually, which lacks systematic validation. Commonly, the Landsat FVC is validated directly by ground-measured FVC because the size of the ground sampling site can match its 30 m spatial resolution pixel. Therefore, the uncertainty of FVC derived from the processed Landsat-7 ETM+ reflectance data was evaluated based on ground FVC measurements. For the GLASS FVC, direct validation has been conducted in previous studies (Jia et al., 2015; Yang et al., 2016), and the validation accuracy of these studies was adopted in this study. (2) Considering the basic assumption that the data processed in MRT have a mean value of zero, it is necessary to extract the spatial trend surface and remove it from each of the original satellite products to ensure that the detrended FVC meets the requirements of the data fusion process. Many detrending methods have been proposed to extract spatial trend surfaces, such as spline fitting and kriging (Marcotte and David, 1988; Shaw and Lynn, 1972). However, these methods are usually time consuming and can hardly be applied in operational practice. He et al. used a simple moving spatial window and considered the mean value of the window as the trend (He et al., 2014). Xu et al. considered the mean value of the entire image as the trend (Xu et al., 2019). However, these mean value methods are empirical parameterization and fail to explain why the detrend result can be the background value and how to estimate the observational errors of the detrended data (Shi et al., 2016). In this study, the detrend method proposed by Shi et al., 2016 was used, which involves a simple interpolation FVC_d to obtain the background trend FVC_{trend} and the background error variance $P_k(s)$ (assuming that MRT has *k* scales) based on the variance of the two measurements (R_G, R_L).

$$FVC_d = (R_L y_G + R_G y_L)/(R_G + R_L) \tag{12}$$

$$FVC_{trend} = FVC_d + mean(y - FVC_d) \tag{13}$$

$$P_k(s) = (R_L^{-1} + R_G^{-1})^{-1} \tag{14}$$

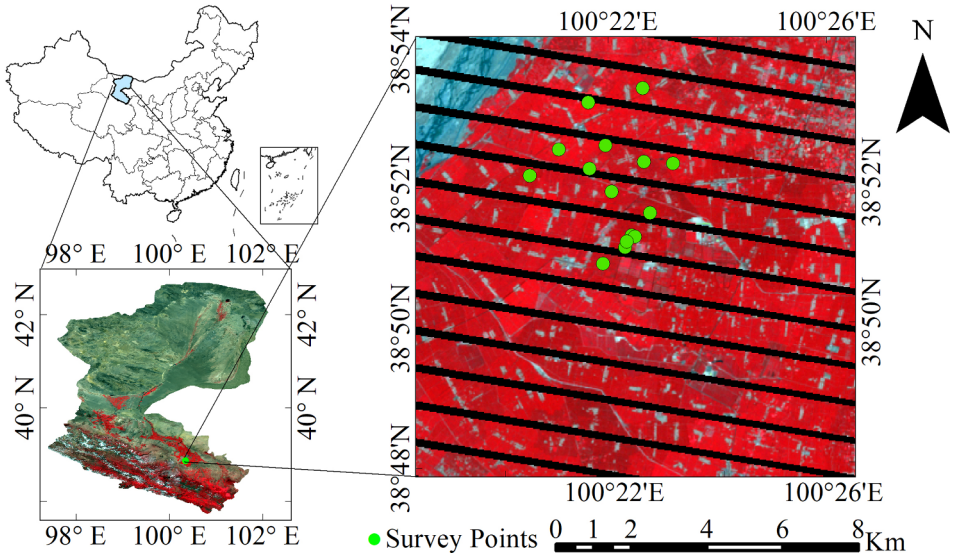

Fig. 3. Geographic location of the study area. The vector map at the upper left shows the location of the Hehei River Basin in China. The MODIS image at the lower left shows the entire Heihe River Basin and the geographic location of the study area. The standard false colour image from Landsat-7 ETM+ on the right shows the landscape of the study area, and the green points represent the locations of sampling sites in the study area.

Table 1
Data used in this study, including Landsat-7 ETM+ reflectance data, MOD09Q1, GLASS FVC data, ASTER FVC and field measured FVC data.

Data	Spatial resolution (m)	Temporal resolution (days)	Day of Year (DOY)
Landsat-7 ETM+ reflectance	30	16	144/176/192/208/224/240/256/272/
MOD09Q1	250	8	145/153/161/169/177/185/193/201/209/217/225/233/241/249/257/265/273
GLASS FVC	463.3	8	145/153/161/169/177/185/193/201/209/217/225/233/241/249/257/265/273
ASTER FVC	15	15	151/176/192/224/240
Field measured FVC	–	5 or 10	145/153/160/165/170/175/180/185/189/192/195/200/205/216/225/241

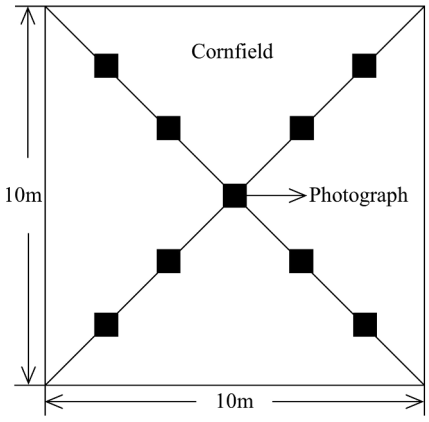

Fig. 4. Strategy of ground FVC measurement in one sampling site using digital cameras.

Based on the detrended data, the variance $Q(s)$ is calculated for the child nodes that are linked with the same parent node. To obtain the observation error $v(s)$ of the leaf nodes, $R(s)$ is derived from the standard deviation of the relative difference between the finest resolution data and the ground survey data. For the nodes that are not leaves, the observation error is calculated by the relative difference between the nodes and the aggregated values from their child nodes. (3) Kalman filtering and Kalman smoothing are implemented successively to obtain the optimal estimator at each scale. (4) The trended surface extracted previously is added back to the fusion data to obtain the final fusion result.

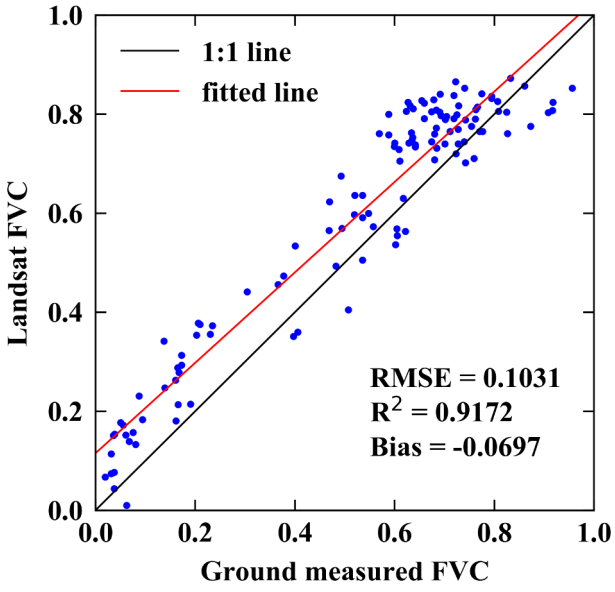

Fig. 5. Scatterplot of Landsat FVC validated by ground-measured FVC.

3. Study area and satellite data

3.1. Study area

The study area is located in an oasis of the upper reaches of the Heihe River Basin, Gansu Province, China (Fig. 3). The geographical location of the study area is between 100°18′5″E-100°27′15″E and 38°47′28″N-38°54′52″N. The annual average temperature and precipitation are approximately 7°C-10°C and 140 mm, respectively (Wang et al., 2016). The study area is a typical arid and semi-arid area and

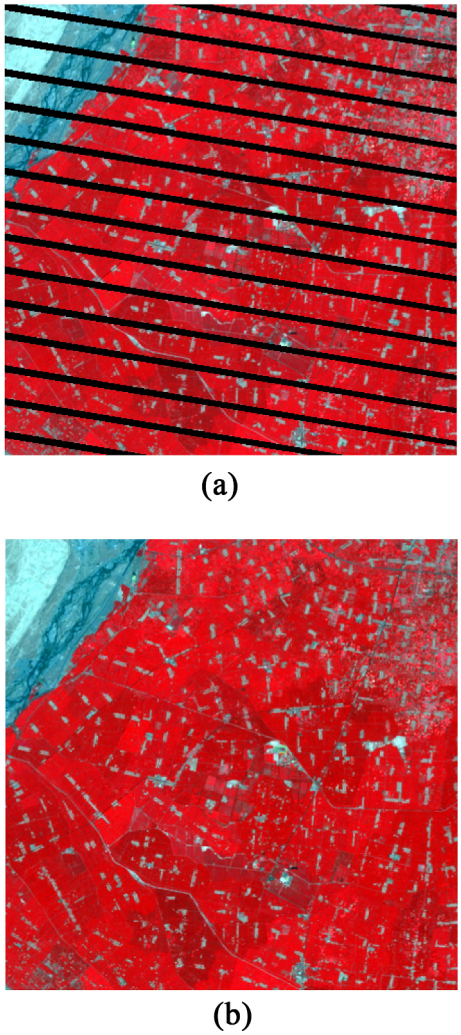

Fig. 6. Interpolated result of the Landsat-7 ETM+ surface reflectance image on DOY 176: (a) original image with unscanned gaps; and (b) interpolated image by GNSPI.

covers approximately 182.25 square kilometres. The landscape is dominated by farmland, and the main crop is corn. Roads crisscross the farmland and buildings are irregularly distributed, which increases surface heterogeneity. Gobi occupies an area northwest of the study area, where the FVC is generally zero. In addition, there are small amounts of grass, water and wetlands in the study area.

3.2. Field survey data

The field FVC measurements were conducted from May to September 2012, which covered the whole growing period of corn. The FVC measurement frequency was initially 5 days (before late July) and was then increased to 10 days, as listed in Table 1. The size of each sampling site was a 10 m square, where nine photographs were taken using digital cameras along two diagonals across the sampling site, as shown in Fig. 4. The FVC was quantitatively extracted from these photographs, and the average value of the nine photographs in one sampling site was considered the true ground FVC (Mu et al., 2015). Fifteen sampling sites were measured in corn fields, and their spatial position distributions were determined by the mean of surfaces with the non-homogeneity (MSN) method (Wang et al., 2009). The spatial distribution of sampling sites is shown in Fig. 3.

The FVC of one digital photograph is defined as the ratio of vegetation pixels to the total number of pixels. This study adopts the FVC extraction method from digital images proposed by Liu et al. (2012). First, twenty-five percent of the photograph edges were cropped to eliminate the influence of the geometric distortion caused by a large view angle. Then, all the digital images were transformed from RGB colour space to Commission Internationale de L'Eclairage (CIE) $L^*a^*b^*$, and Gaussian models were used to fit the green vegetation and background distributions of the a^* component. Finally, an optimal threshold was selected based on the fitted Gaussian models to segment the image into green vegetation and non-green background. The proportion of green vegetation was considered as FVC.

3.3. Satellite data and pre-processing

3.3.1. Landsat-7 ETM+ reflectance

The Landsat-7 ETM+ reflectance data were collected between May 23 and September 28, 2012, from EarthExplorer of the United States Geological Survey (USGS) (https://earthexplorer.usgs.gov/). There were 8 images with a spatial resolution of 30 m and temporal resolution of 16 days (Table 1) that covered the entire growth period of corn. The unscanned gaps and cloud contaminations led to some missing data in the image, and the GNSPI was used to fill the unscanned gaps and cloud contaminated areas. In this study, the red and NIR wavebands of Landsat-7 ETM+ reflectance data were used to estimate high spatial resolution FVC.

3.3.2. MOD09Q1 surface reflectance

The MOD09Q1 data were an 8-day composited MODIS surface reflectance product that provided the land surface reflectance of the red and NIR bands with a frequency of 8 days and a spatial resolution of 250 m in the sinusoidal projection. The highest spatial resolution of MODIS reflectance products is 250 m, which is beneficial for ESTARFM to blend Landsat and MOD09Q1 data to generate more accurate Landsat reflectance data. Therefore, the MOD09Q1 data from DOY 145 to 273 were collected (Table 1) as the coarse resolution reflectance data in this study. The pixels of MOD09Q1 data contaminated by snow/clouds were first identified using the time series cloud detection (TSCD) algorithm (Tang et al., 2013) and removed using a temporal-spatial filter, and then the missing values were filled using an optimum interpolation algorithm (Jia et al., 2015). Therefore, the preprocessed MOD09Q1 data used in this study were continuous spatially and temporally.

Furthermore, MOD09Q1 data should be converted to Universal Transverse Mercator (UTM) projection and resampled to 30 m spatial resolution for consistency with Landsat-7 ETM+ data because the ESTARFM algorithm requires that fine and coarse resolution data have the same projection and spatial resolution. The reprojection and resampling processes change the size, shape and position of a given pixel and distort the relative arrangement of grid cell matrices (Christman and Rogan, 2012), which introduces uncertainties into the processed results. Therefore, appropriate methods should be adopted to reduce the uncertainties caused by reprojection and resampling. The MODIS Reprojection Tool is widely used in the projection conversion of MODIS Level-2G, Level-3 and Level-4 products with high precision and efficient computation (Brooks et al., 2012; Dwyer and Schmidt, 2006; Weiss et al., 2014). The bilinear interpolation method considers the effect of four adjacent pixels on the interpolation point and usually achieves high interpolation accuracy with satisfactory computation efficiency. Therefore, the MODIS Reprojection Tool and bilinear interpolation method were used to reproject and resample MOD09Q1 data. These two methods minimized the uncertainties in reprojecting and resampling coarse-resolution MOD09Q1 data for integration with fine-resolution Landsat-7 ETM+ data in this study.

3.3.3. GLASS FVC

The GLASS FVC product was generated using MOD09A1 data and had the same preprocessing as MOD09Q1 data. The GLASS FVC was

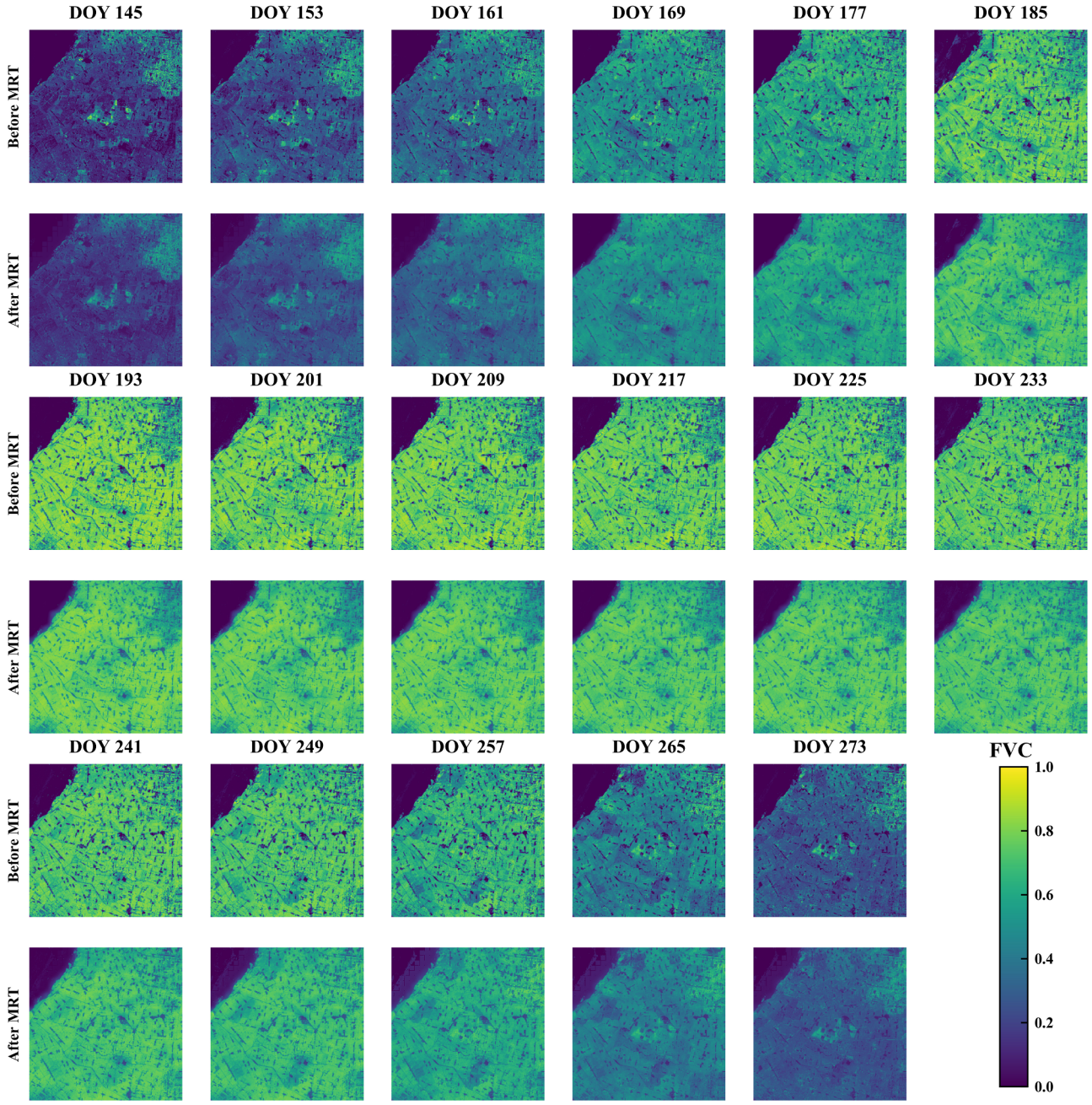

Fig. 7. Comparison of time series Landsat FVC from DOY 144 to DOY 272 before and after MRT. The images in the odd line are the original FVC, and the images in the even line are the FVC after MRT.

spatially completed with no missing data and temporally continuous with a revisit interval of 8 days. The validation results showed that the accuracy of the GLASS FVC was satisfactory (Jia et al., 2015; Liu et al., 2019). Corresponding to MOD09Q1 data, the GLASS FVC was also collected from DOY 145 to 273 (Table 1) with a sinusoidal projection and spatial resolution of 500 m (the real grid size of the nominal 500 m sinusoidal pixel size was 463.3 m). The GLASS FVC was also converted to UTM projection by the MODIS Reprojection Tool and resampled to 450 m using the bilinear interpolation method to match the 30 m pixel size of Landsat-7 ETM+ data by a factor of 15.

3.3.4. ASTER FVC

The ASTER L1B radiance data with 15 m spatial resolution on DOY 176, 192, 224 and 246 of 2012 were collected to generate high spatial resolution reference FVC to validate the GLASS FVC. The ASTER L1B data were atmospherically corrected by a second simulation of a satellite signal in the solar spectrum (6S) method (Vermote et al., 1997) to obtain surface reflectance, which was used to calculate the normalized difference vegetation index (NDVI). The following empirical transfer function was used to convert the NDVI into the FVC, and the coefficients a, b and k were determined by fitting the ASTER NDVI and ground-measured FVC (Mu et al., 2015).

$$FVC = (a \times NDVI + b)^k \quad (15)$$

4. Results

4.1. The fusion results of two satellite-derived FVC products

The uncertainties of Landsat and GLASS FVC products must be

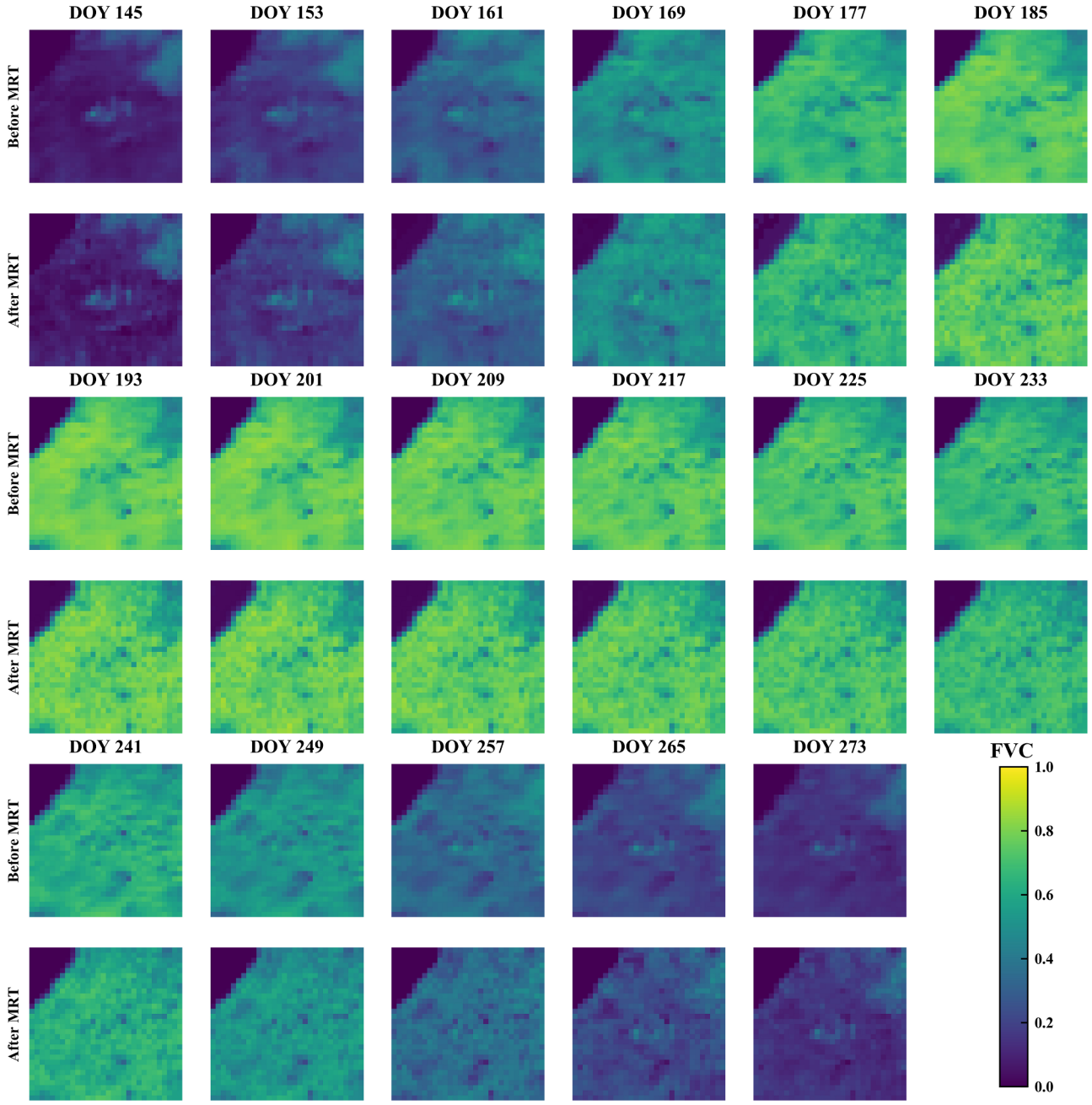

Fig. 8. Comparison of time series GLASS FVC from DOY 145 to DOY 273 before and after MRT. The images in the odd line are the original FVC, and the images in the even line are the FVC after MRT.

evaluated when merging two FVC products using the MRT method. The uncertainty of the GLASS FVC was validated directly based on the validation samples from the Validation of Land European Remote sensing Instruments (VALERI) site. The validation result indicated that the GLASS FVC product had an accuracy with an RMSE of 0.149 (Jia et al., 2019). Therefore, $R(s)$ for the GLASS FVC was assigned to 0.149. The uncertainty of the Landsat FVC was evaluated by ground FVC measurements. Ground measurements on 10 dates were selected out of the 15 survey dates to validate the Landsat FVC, including DOY 145, 153, 160, 170, 175, 185, 192, 200, 216 and 241, which were at most one day apart from the dates of the corresponding Landsat FVC. Fig. 5 shows the validation results of the Landsat FVC with an RMSE of 0.1031. Therefore, 0.1031 is assigned to the uncertainty of the Landsat FVC $R(s)$.

The unscanned gaps and cloudy areas in the original Landsat-7 ETM + reflectance images were interpolated by GNSPI first to reconstruct the complete images, and an example is shown in Fig. 6. In most of the filled images, the unscanned gaps are well filled, and the details of the spatial texture of the surface are recovered completely, as shown in Fig. 6. There is no residual gap effect on the images, and the spatial transition is good. All the filled images were checked carefully, and the GNSPI show poor performance in filling in the unscanned gaps on the DOY 161 image because of heavy clouds and cloud shadows. The predicted images on DOY 153 and 169 using ESTARFM would be affected by the remaining gaps on the DOY 161 image. Therefore, images on DOY 153, 161, and 169 were all generated by ESTARFM based on images of DOY 144 and 176.

Fig. 7 shows the results of the time series Landsat-scale FVC before and after MRT from DOY 144 to DOY 272. The revisit cycle of Landsat-7 ETM+ is 16 days, and the temporal resolution of GLASS FVC is 8 days. Because the image acquisition dates of these two satellite-derived FVC

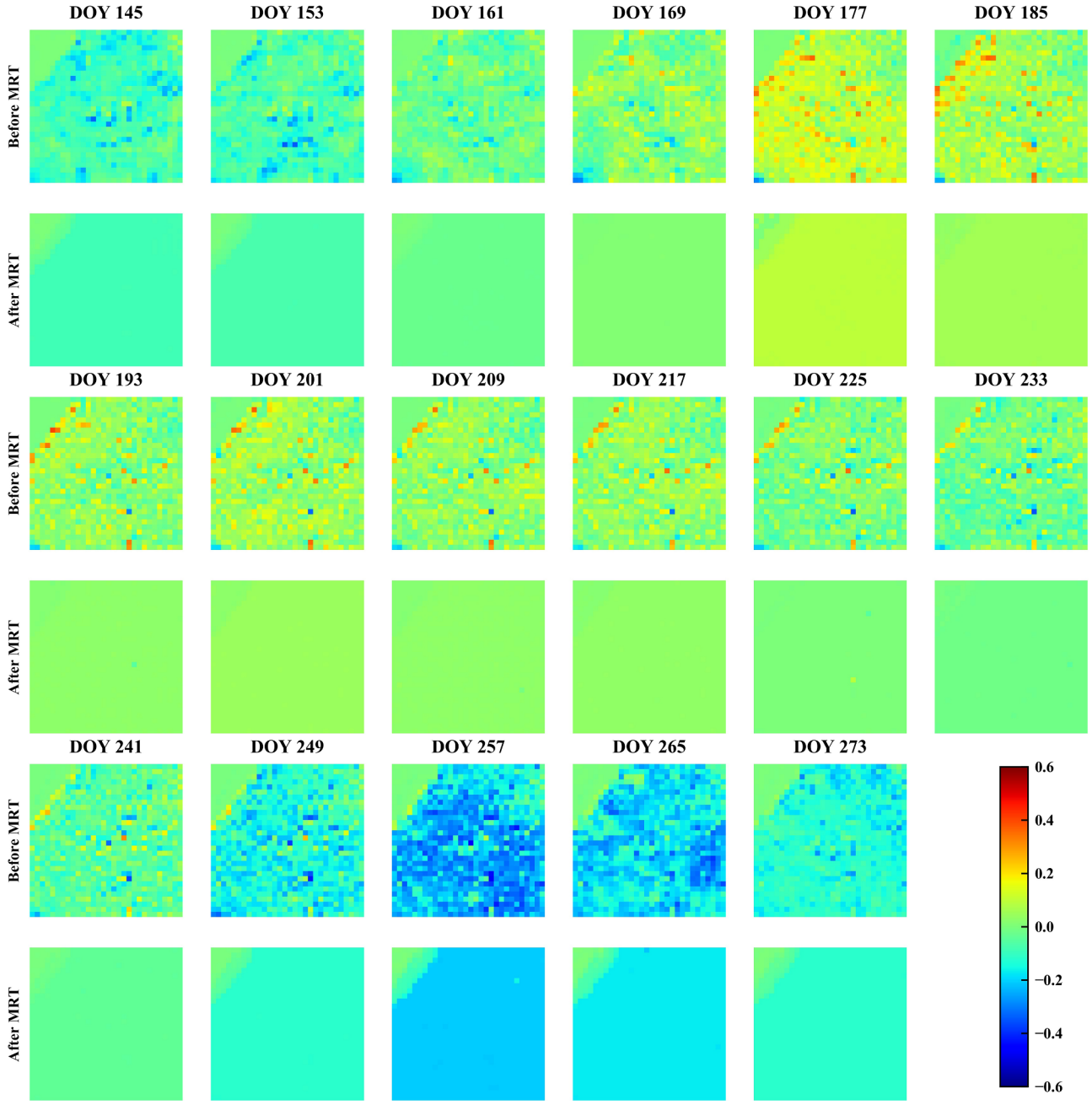

Fig. 9. The difference between aggregated Landsat and GLASS FVC before and after MRT.

products are one day apart, the temporal resolution of fine spatial resolution FVC is 7 days or 9 days after ESTARFM, which means that temporal inconsistency is eliminated. The ESTARFM helps improve the temporal resolution of the Landsat FVC and does not violate the temporal change characteristics of Landsat FVC. The upper left corner of the study area is desert, where the FVC remains at a low value throughout the entire growth period of corn. In the agricultural area, FVC first increases and then stabilizes at a high level and finally decreases, which reflects the process of the growth, maturity and yellowing of corn. At the beginning (DOY 144 to DOY 161) and middle (DOY 192 to DOY 240) of the entire corn growth period, the value range of FVC is small and remains at low and high values, respectively. It is unnecessary to generate the FVC with high temporal resolution in the beginning and middle of corn growth periods because there is very little change in the FVC value. However, during the periods of rapid crop growth (DOY 161 to DOY 192) and yellowing (DOY 240 to DOY 272), the FVC varies sharply, and high temporal resolution is essential for monitoring crop changes. The original 16-day revisit cycle and frequent cloud contamination limits the application of Landsat data to a large extent, and it is very likely that the critical period of vegetation growth would be missed. After MRT, the Landsat-scale FVC becomes smoother than before, and the spatial texture details are well preserved. Despite the spatial heterogeneity of the study area, the Landsat-scale FVC images are also spatially continuous, and there is no abnormal value after MRT. The seasonal variation in Landsat FVC does not change and coincides with the original FVC.

Fig. 8 shows the time series GLASS FVC before and after MRT from DOY 145 to DOY 273. Whether before or after MRT, the seasonal variation in GLASS FVC is consistent with that of Landsat FVC. Generally, coarse spatial resolution images are smoother than fine spatial resolution images because the latter are more capable of capturing the spatial heterogeneity and spatial distribution patterns of the surface. This

Table 2
Statistical comparison of aggregated Landsat and GLASS FVC before and after MRT.

DOY	GLASS vs. Aggregated Landsat					
	Before MRT		After MRT		Variation	
	Bias	RMSE	Bias	RMSE	Bias	RMSE
144	−0.0824	0.1002	−0.0801	0.08200	−2.8182%	−18.1482%
153	−0.0454	0.0738	−0.0697	0.07121	−3.4585%	−23.5441%
160	0.0276	0.0730	−0.0263	0.02686	−2.7412%	−54.7116%
169	0.0213	0.0781	0.0103	0.01044	0.9097%	−84.5335%
176	0.0916	0.1166	0.0907	0.09189	−0.9381%	−21.1841%
185	0.0560	0.0967	0.0560	0.05618	−0.0193%	−42.0061%
192	0.0265	0.0769	0.0267	0.02705	0.8253%	−64.8093%
201	0.0436	0.0846	0.0436	0.04384	0.0799%	−48.1785%
208	0.0273	0.0747	0.0275	0.02774	0.7062%	−62.8879%
217	0.0300	0.0745	0.0301	0.03022	0.5841%	−59.4364%
224	0.0009	0.0701	0.0015	0.00398	64.4674%	−94.3286%
233	−0.0183	0.0716	−0.0174	0.01759	−4.7516%	−75.4463%
240	−0.0412	0.0817	−0.0394	0.04000	−4.4718%	−51.0434%
249	−0.1195	0.1436	−0.1149	0.11701	−3.8413%	−18.4944%
256	−0.2067	0.2295	−0.1992	0.20315	−3.6107%	−11.4980%
265	−0.1692	0.1877	−0.1637	0.16695	−3.2787%	−11.0548%
272	−0.1177	0.1287	−0.1140	0.11623	−3.1534%	−9.6766%

phenomenon can be found by comparing the Landsat FVC and GLASS FVC in Fig. 7 and Fig. 8. In the process of MRT, coarse and fine spatial resolution data interact with each other. In the Kalman filtering process, Landsat FVC provides GLASS FVC with finer spatial information for capturing more details of spatial variations in the land surface. Therefore, after MRT, the GLASS FVC becomes less smooth and more heterogeneous than before. In the Kalman smoothing process, GLASS FVC provides Landsat FVC with coarser information and makes it smoother, and the spatial texture of Landsat FVC is slightly weakened by MRT. Nevertheless, the fusion results after MRT are acceptable to reflect the spatial pattern at the 30 m scale. Compared with Landsat and GLASS FVC, high spatial and high temporal resolution FVC is capable of tracking the surface changes in regional areas with high spatial heterogeneity, and low spatial and high temporal resolution FVC is suitable for large areas with high spatial homogeneity.

4.2. Evaluation of spatial consistency and accuracy of the two scale FVC products before and after MRT

The main purpose of MRT is to eliminate the spatial inconsistency of FVC at different scales and reduce the uncertainties of both fine- and coarse-resolution FVC. To measure the degree of consistency between two satellite-derived FVC products, two statistical metrics, bias and RMSE, were selected to quantitatively describe the spatial consistency. The smaller the bias and RMSE are, the lower the spatial inconsistency. The Landsat FVC was aggregated to 450 m to match the pixel size of the GLASS FVC. The difference between the two FVC products was calculated by subtracting the aggregated Landsat FVC from the GLASS FVC. Fig. 9 shows the results of the differences between two FVC products before and after MRT. There are no obvious characteristics in the spatial distribution of differences before MRT. A certain number of outliers occur in the farmland area and result in large differences, which are mainly due to the strong heterogeneity of the surface. In the homogeneous Gobi area, the differences are relatively low. Images from most dates show large inconsistencies in the transition regions between Gobi and farmland, which are caused by mixed pixels of Gobi and farmland. After MRT, the outliers disappear, and the spatial inconsistencies between the two products are significantly reduced.

The statistical comparison between aggregated Landsat and GLASS FVC before and after MRT is shown in Table 2, and a negative number indicates that the inconsistency has declined. According to the statistical results, the variation in bias on DOY 169, 192, 201, 208, 217, and 224 has a slight increase, which remains almost unchanged. Except for these six days, the bias has a reduction on other days ranging from 0.0193% to 4.7516%. However, the bias levels before and after MRT are both relatively low. The RMSE of all days has a significant reduction ranging from 9.6766% to 94.3286%, and most of them are larger than 40%, which indicates that after MRT, the spatial inconsistencies of the Landsat and GLASS FVC products have decreased to a great extent. The improvement in accuracy demonstrates that MRT is capable of reducing the inconsistencies of satellite-derived FVC across different scales and generating spatially consistent FVC products.

Fig. 10 shows the histogram of the differences between GLASS FVC and aggregated Landsat FVC. After MRT, most histograms are more concentrated around zero, which indicates that the differences between the GLASS FVC and aggregated Landsat FVC are reduced again. However, on DOY 144, 153, 249, 256, 265 and 272, the distribution of histograms deviates from zero. Nevertheless, the distribution of histograms after MRT is more concentrated than that before MRT on these 6 days, which indicates that the inconsistencies between the two products have a certain degree of decline. This phenomenon can also be found in Table 2; the decrease in RMSE on these six days (−18.1482%, −23.5441%, −18.4944%, −11.4980%, −11.0548%, −9.6766%) is less than 24% and smaller than that on other days, which has a decrease in RMSE almost larger than 40%. Therefore, the fine-resolution and coarse-resolution FVC after MRT are spatially consistent, which is convenient for research at different scales. The problem of inconsistent results caused by the scale effect can be reduced to some extent.

The original Landsat-7 ETM+ reflectance was processed by GNSPI to generate spatially continuous high-resolution images, and ESTARFM was used to produce reflectance between two adjacent images in time with the purpose of improving the temporal resolution and making the frequency of the Landsat and GLASS FVC products consistent. These two algorithms might introduce uncertainties to the fused reflectance, which would affect the accuracy of FVC estimation. Therefore, MRT algorithms were used to reduce the uncertainties and improve the accuracy of FVC products. Fig. 11 (a) shows the scatterplots of the Landsat FVC validated by the ground-measured FVC of 10 selected dates including DOY 145, 153, 160, 170, 175, 185, 192, 200, 216 and 241. After MRT, the FVC estimation accuracy (RMSE = 0.0958, R^2 = 0.9173, Bias = −0.054) is better than that before MRT (RMSE = 0.1031, R^2 = 0.9172, Bias = −0.0697, Fig. 5). Fig. 9 (b) and (c) shows the scatterplots of ground-measured FVC located in GNSPI-filled gaps and ESTARFM-generated images before and after MRT, including DOY 153, 160, 170, 185, 200, 216 and part of points on DOY 160. Before MRT, the RMSE is 0.1056, R^2 is 0.9011, and the bias is −0.0644. The validation result shows that the FVC was slightly overestimated, which may be caused by the lack of shortwave-infrared (SWIR) information because SWIR is helpful for FVC estimation (Wang et al., 2018). After MRT, the RMSE is reduced to 0.1022, R^2 increases to 0.9023 and bias decreases to −0.051. The statistical metrics indicate that the accuracy of the Landsat FVC has been improved slightly. Otherwise, the phenomenon of overestimation has been improved in the median and high FVC values. The improvement of the Landsat FVC accuracy demonstrates that MRT is capable of reducing the uncertainty of fine spatial resolution FVC.

During the MRT process, the GLASS FVC was affected by the Landsat FVC, and the ASTER high-resolution FVC was used as reference data to evaluate the GLASS FVC before and after MRT. The ASTER FVC with a spatial resolution of 15 m was aggregated to 450 m to match the pixel size of the GLASS FVC. Fig. 12 shows the scatterplots of GLASS FVC validated by aggregated ASTER FVC before and after MRT. The scatter points contain DOY 152, 176, 192, 224 and 240. Before MRT, the scatter points are dispersed, and for the low FVC values of approximately 0.1, the FVC was overestimated slightly. After MRT, the accuracy is improved to some extent, and the scatter points are more concentrated around the 1:1 line than before MRT. Overall, MRT can reduce the uncertainty of coarse spatial resolution FVC.

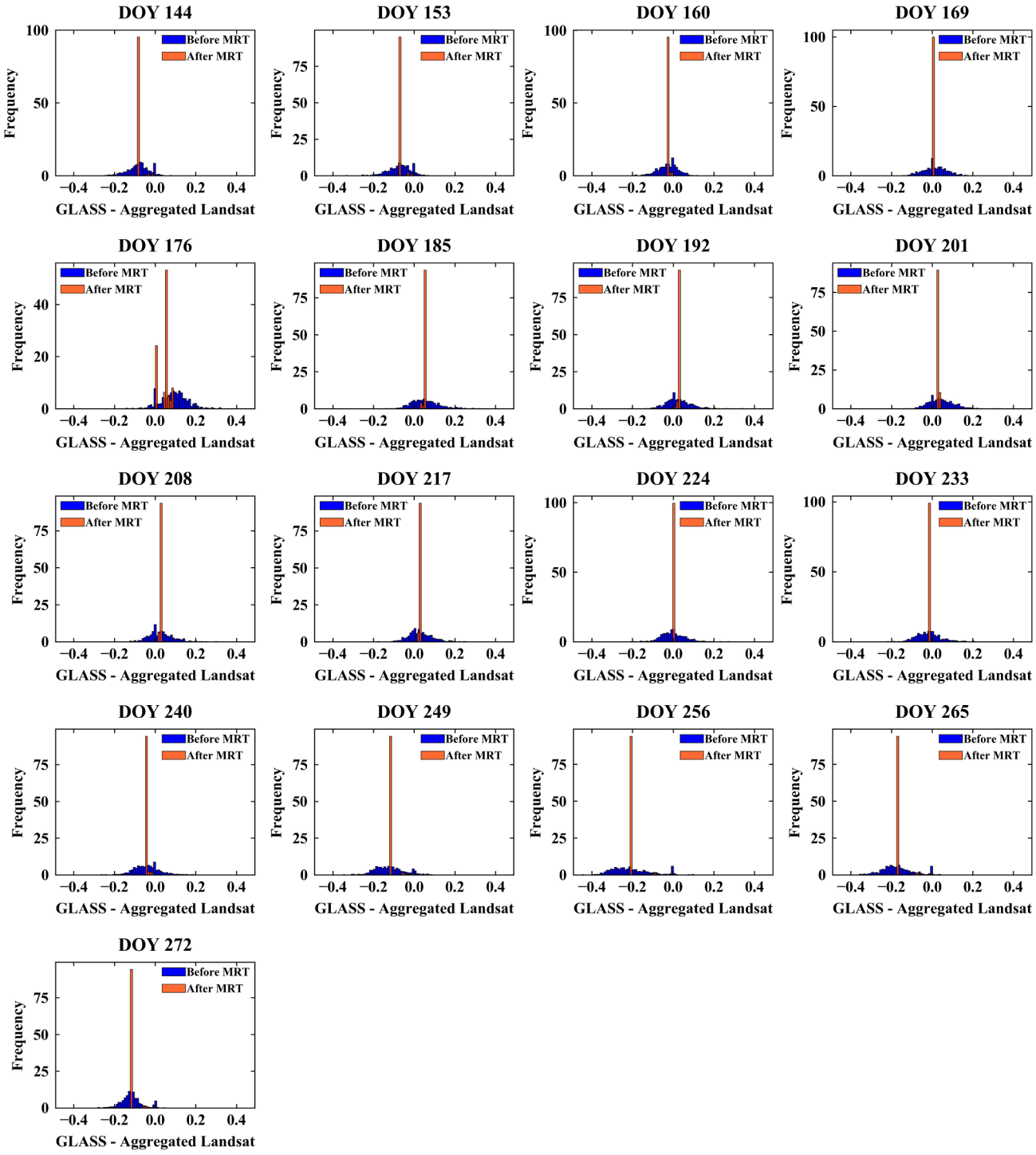

Fig. 10. Histograms of the differences between the GLASS FVC and aggregated Landsat FVC before and after MRT from DOY 144 to DOY 272. (DOY refers to the DOY of Landsat).

4.3. Fusing Landsat FVC directly using MRT without filling unscanned gaps

In this study, the missing Landsat reflectance data caused by unscanned gaps and clouds/shadows were filled by the GNSPI algorithm in advance, and the estimated Landsat FVC is spatially continuous. However, MRT is capable of filling the missing data of Landsat FVC directly because GLASS FVC is spatially continuous and can provide coarse spatial information for missing Landsat FVC. Therefore, the Landsat FVC was estimated using the reflectance image with unscanned gaps, and then the Landsat FVC gaps were filled by MRT directly to evaluate the ability of MRT to reconstruct the missing data of fine spatial resolution data. Fig. 13 shows the results of Landsat-7 ETM+ FVC with unscanned gaps directly interpolated by MRT using GLASS FVC. The unscanned gaps on Landsat FVC are completely filled with valid values, but a clear gap effect is still obvious in the fusion results. The filled areas have weak spatial texture, and spatial transitions are smooth. The direct filling results are inferior to the results by the proposed method in terms of their texture consistencies. The upper left corner of the image is the desert, and the FVC changes dramatically at the edge of the desert. It can be clearly seen from Fig. 13 that there are

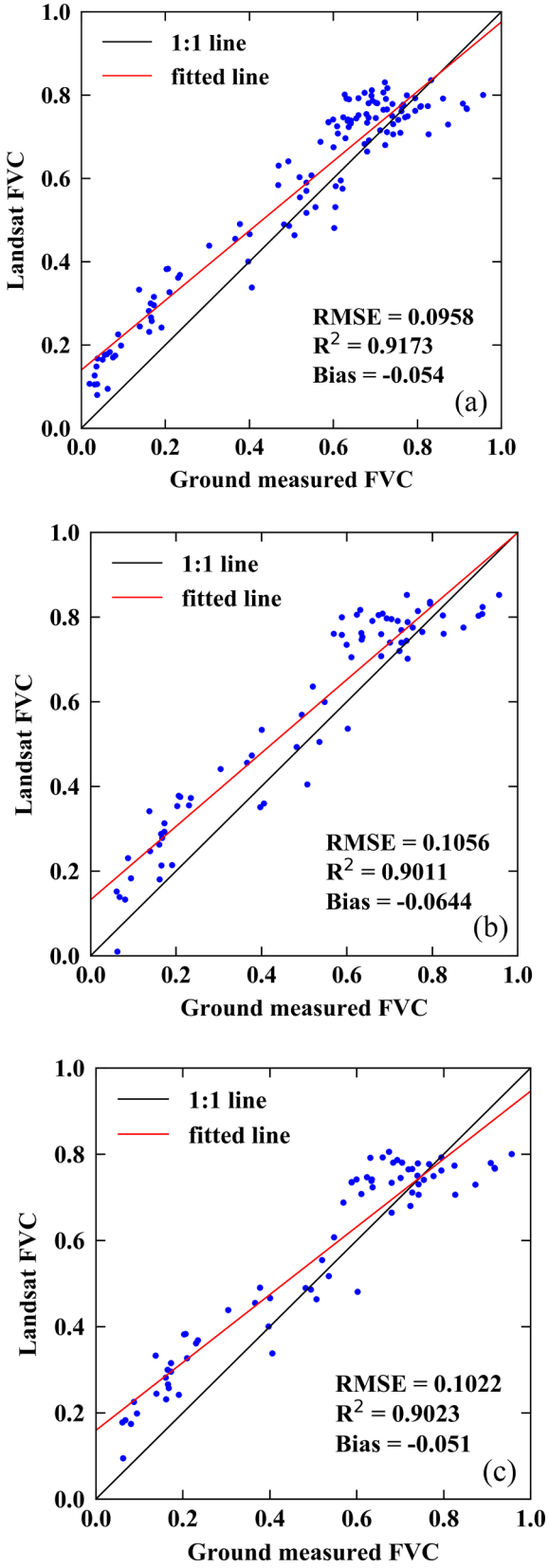

Fig. 11. Scatterplots of Landsat FVC validated by ground-measured FVC: (a) all field survey points of 10 selected dates after MRT; and (b) and (c) the survey points located in GNSPI-filled gaps and ESTARFM-generated images before and after MRT, respectively.

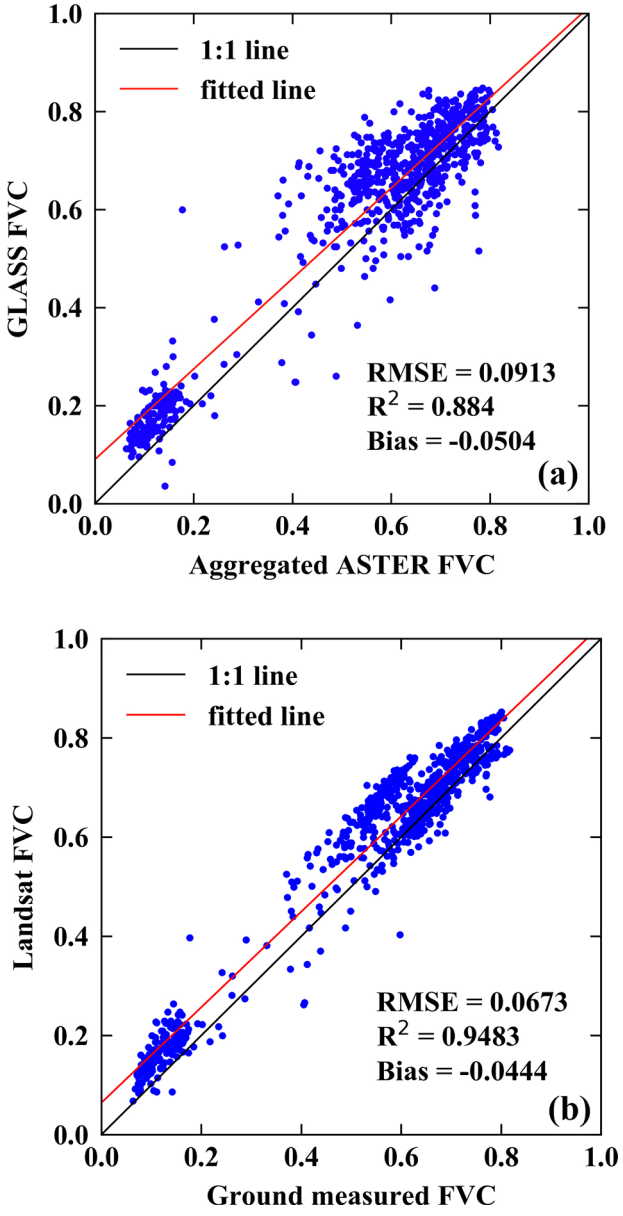

Fig. 12. Scatterplots of GLASS FVC validated by aggregated ASTER FVC before (a) and after (b) MRT.

some outliers in this area. The FVC values in the desert are supposed to be very low, but the unscanned gap areas of the edge of the desert are interpolated with higher values because the pixel of GLASS FVC with a relatively higher value covers across the edge of the desert and fills the unscanned gaps with high values, which are outliers on the side of the desert. Comparing the four fusion results, DOY 192 and 224 have relatively good spatial transitions and texture consistencies, especially in some homogeneous areas, but all the images are heterogeneous, and the fusion results are not satisfactory in the spatial pattern. The main reason that MRT cannot reasonably recover the spatial pattern is the lack of fine spatial resolution information. In heterogeneous areas, the GLASS FVC is comprehensive information due to the existence of mixed pixels. However, MRT could not decompose the GLASS FVC accurately without Landsat FVC information and therefore produced FVC values similar to dominant features in a GLASS FVC pixel.

Fig. 14 shows the scatterplots of FVC located inside the unscanned gaps of Landsat directly interpolated by MRT on DOY 176, 192, 224 and 240. The accuracy of FVC was validated by ASTER FVC. The RMSEs of these four days are between 0.14 and 0.2, and the R^2 values are

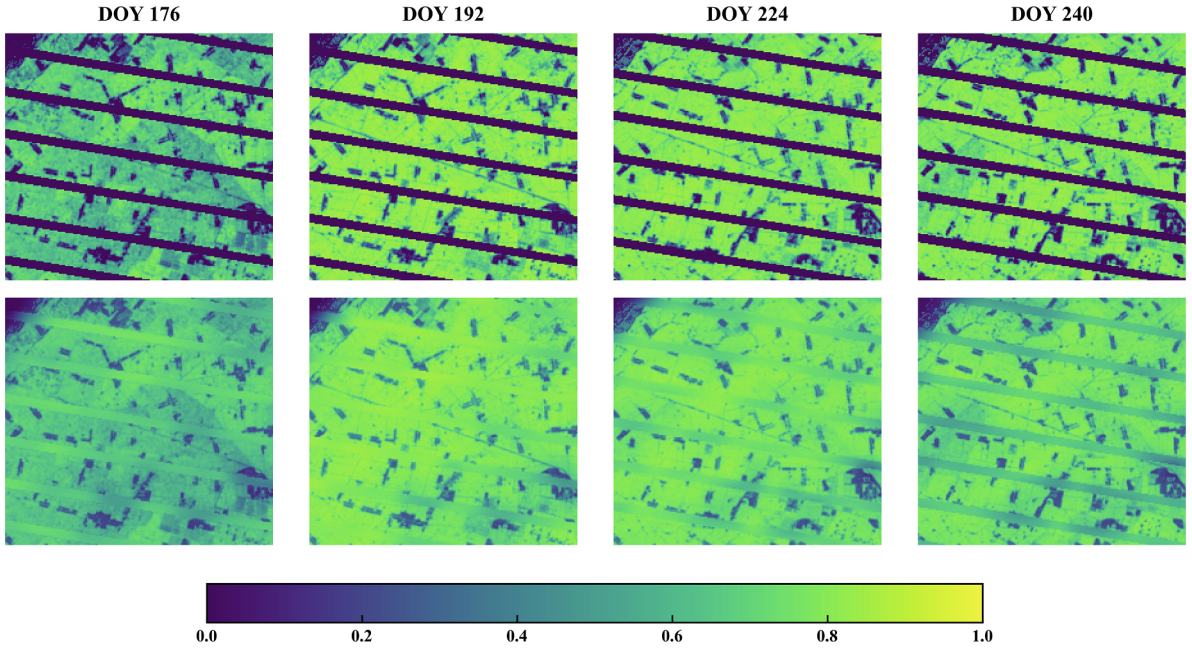

Fig. 13. Landsat-7 ETM+ FVC without filling unscanned gaps before (first row) and after (second row) MRT on DOY 176, 192, 224 and 240.

generally low. The bias of DOY 176 is above 0.1, relatively higher than that of the other three with low values under 0.05. The FVC on DOY 176 is overestimated overall, and most of the scatter points are above the 1:1 line. On DOY 192, 224 and 240, the high-density scatter points are located near the 1:1 line; thus, the bias of these three days is small. A certain percentage of scatter points are outliers in all four days, and most of them are overestimated. This phenomenon is caused by two factors: (1) as mentioned above, the edge of the desert has some overestimated outliers, which are circled with green ellipses in Fig. 14; (2) the reference FVC of other outliers circled by red ellipses in Fig. 14 has a large range, but the interpolated FVC has a certain small range. This is because the unscanned gap areas are covered not only by farmland but also by some buildings, roads and bare land, which have relatively low FVC. However, these areas are filled with GLASS FVC, the interpolated FVC cannot reflect the detailed spatial variation of the land surface, and the areas with low FVC are filled with high and similar FVC values, which causes the small changes in interpolated FVC. Overall, in terms of fusion results and accuracy validation, the performance of missing Landsat FVC directly filled by MRT using the GLASS FVC is unsatisfactory.

5. Discussion

Missing data caused by sensor misfunctions and cloud/shadow contamination as well as low revisit frequencies are two major reasons that limit the applications of Landsat images. It is necessary to improve the temporal resolution of Landsat and reconstruct missing values on the image to ensure that the critical periods or areas of interest can be captured. To address these issues, GNSPI was used to fill the missing values of Landsat data, and the spatiotemporal fusion algorithm ESTARFM was implemented to increase the temporal resolution of Landsat data to be consistent with MOD09Q1. After these two steps, the Landsat images are spatially continuous and have the same temporal resolution as MOD09Q1. However, uncertainties are also introduced to the fused reflectance, which affects the FVC estimates. MRT is used to reduce the uncertainties and eliminate the inconsistences between FVC at different scales. The results have demonstrated that MRT has the ability to reduce uncertainties, and the issue of overestimation of Landsat FVC at medium and high levels is addressed by MRT to some extent, but overestimation is still evident in the case of a low FVC level. This is mainly because the validation results of the two satellite-derived FVC show overestimation at the same time and MRT cannot reconcile the conflict that the data at fine and coarse resolution are both overestimated and generate reliable FVC values.

MRT has successfully reduced inconsistencies between FVC products at different scales according to statistical metrics and histograms. However, there are still some data with large inconsistencies on DOY 249, 256, 265 and 272 and some data with relatively larger inconsistencies on DOY 144 and 153. This is mainly because the Landsat FVC has a large deviation from the GLASS FVC. There may be two reasons leading to the differences between Landsat and GLASS FVC: (1) the estimation algorithms of the Landsat and GLASS FVC are different; and (2) from DOY 249 to 272, the corn has matured and continuous harvesting has begun, and the Landsat and GLASS FVC products may capture different states of corn due to their different temporal resolutions. As shown in Fig. 10, the histograms before MRT (blue) follow a Gaussian distribution with a mean difference between 0.1 and 0.2, which indicates that there are large inconsistencies between these two FVC products. A similar conclusion was also drawn by Xu et al. (2019) and Shi et al. (2016). From the statistical comparison of Table 2, most data have small variations in bias but large variations in RMSE. The bias is calculated as the average value of the difference between the GLASS FVC and aggregated Landsat FVC (Lindberg et al., 2012), and the differences follow a Gaussian distribution with a certain mean value (most of them are zero) around which the histograms after MRT are mainly distributed. Therefore, the biases of most days have small values and small variations. The RMSE reflects the error between two datasets, and a low RMSE indicates the ability of MRT to reduce uncertainties between different data. Because the Landsat FVC largely deviates from the GLASS FVC, the variation in the RMSE is small on DOY 144, 153, 249, 256, 265 and 272, which again suggests the limitation of MRT.

MRT can take advantage of data from other scales to recover the missing values of the image. In this study, the GLASS FVC is spatially continuous and directly used to fill in the unscanned gaps on the Landsat ETM+ FVC. The interpolation results using MRT are similar to those obtained by direct resampling from the GLASS FVC, but the fused value is the sum of the trend surface and a random process estimation (Shi et al., 2016). The interpolation results are better in homogeneous areas than those in heterogeneous areas. If the two datasets used for fusion have significant differences, the results will be worse. Although

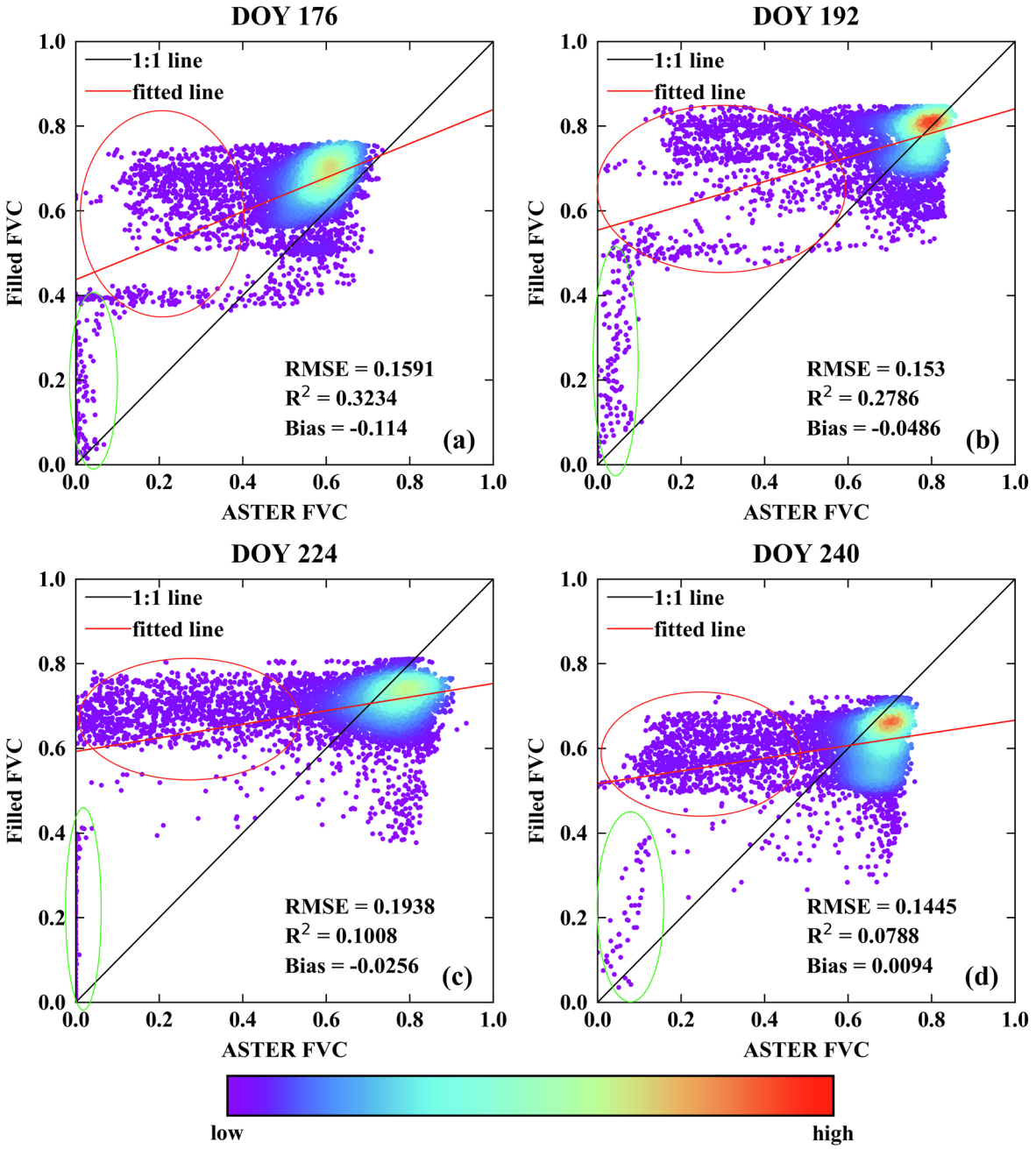

Fig. 14. Scatterplots of FVC directly interpolated by MRT using the GLASS FVC in the Landsat-7 ETM+ unscanned gaps validated by the ASTER FVC on DOY 176 (a), 192 (b), 224 (c) and 240 (d).

MRT can fill in the unscanned gaps of the Landsat ETM+ FVC completely, the algorithm is unable to recover the details of spatial texture, and the spatial transition is poor. There are a certain number of outliers in the fusion results. Therefore, in this study, the missing data caused by cloud/shadow or unscanned gaps were filled by GNSPI to generate reflectance images, which reflect the land cover more accurately, with good spatial transition and clear details of texture and to avoid the occurrence of outliers.

In this study, ESTARFM was used to improve the temporal resolution of Landsat reflectance images to 8 days. However, this temporal resolution may be insufficient for monitoring the critical period of vegetation growth, especially for faster-growing crops. The application of multisource data integration is a feasible way to further improve the temporal resolution of high spatial resolution data, such as Landsat (16 days), Sentinel-2 (5 days, pair of Sentinel-2 satellites) (Verrelst et al., 2012), GaoFen-1 (4 days) and GaoFen-6 (Gu and Tong, 2015).

6. Conclusions

This study proposed a feasible way to generate spatiotemporally consistent FVC products at different scales. The results indicate that MRT is capable of weakening the inconsistencies between Landsat and GLASS FVC products and reducing uncertainties to some extent. The details of the Landsat FVC spatial pattern were well preserved, and the heterogeneity of the GLASS FVC increased. MRT can also be used for directly interpolating Landsat FVC, but the results are poor, which indicates that the missing data in the high spatial resolution images must be processed beforehand to achieve good results. Further work will be focused on the application of multisource data to generate high spatial resolution data with higher temporal resolution and try to fuse other

vegetation parameters, such as leaf index area (LAI) and fractional absorbed photosynthetically active radiation (FAPAR).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was supported by the Second Tibetan Plateau Scientific Expedition and Research Program (No. 2019QZKK0405), the National Natural Science Foundation of China (No. 41671332), the National Key Research and Development Program of China (2016YFB0501404 and 2016YFA0600103), and the Tang Scholar Program (K. Jia is a Tang Scholar of Beijing Normal University). The authors would also like to thank Dr. X. Mu from Beijing Normal University for providing part of the field reference data.

References

Azzari, G., Lobell, D.B., 2017. Landsat-based classification in the cloud: An opportunity for a paradigm shift in land cover monitoring. Remote Sens. Environ. 202, 64–74.
Bacour, C., Baret, F., Béal, D., Weiss, M., Pavageau, K., 2006. Neural network estimation of LAI, fAPAR, fCover and LAI×Cab, from top of canopy MERIS reflectance data: Principles and validation. Remote Sens. Environ. 105, 313–325.
Baret, F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., Berthelot, B., Niño, F., Weiss, M., Samain, O., Roujean, J.L., Leroy, M., 2007. LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION: Part 1: Principles of the algorithm. Remote Sens. Environ. 110, 275–286.
Baret, F., Weiss, M., Lacaze, R., Camacho, F., Makhmara, H., Pacholcyzk, P., Smets, B., 2013. GEOV1: LAI and FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part1: Principles of development and production. Remote Sens. Environ. 137, 299–309.
Bian, J., Li, A., Zhang, Z., Zhao, W., Lei, G., Yin, G., Jin, H., Tan, J., Huang, C., 2017. Monitoring fractional green vegetation cover dynamics over a seasonally inundated alpine wetland using dense time series HJ-1A/B constellation images and an adaptive endmember selection LSMM model. Remote Sens. Environ. 197, 98–114.
Bolton, D.K., Friedl, M.A., 2013. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. Agric. For. Meteorol. 173, 74–84.
Brooks, E.B., Thomas, V.A., Wynne, R.H., Coulston, J.W., 2012. Fitting the multitemporal curve: a Fourier series approach to the missing data problem in remote sensing analysis. IEEE Trans. Geosci. Remote Sens. 50, 3340–3353.
Chen, J., Zhu, X., Vogelmann, J.E., Gao, F., Jin, S., 2011. A simple and effective method for filling gaps in Landsat ETM+ SLC-off images. Remote Sens. Environ. 115, 1053–1064.
Chou, K.C., Willsky, A.S., Benveniste, A., 1994a. Multiscale recursive estimation, data fusion, and regularization. IEEE Trans. Autom. Control 39, 464–478.
Chou, K.C., Willsky, A.S., Nikoukhah, R., 1994b. Multiscale systems, Kalman filters, and Riccati equations. IEEE Trans. Autom. Control 39, 479–492.
Christman, Z.J., Rogan, J., 2012. Error propagation in raster data integration. Photogramm. Eng. Remote Sens. 78, 617–624.
de Asis, A.M., Omasa, K., 2007. Estimation of vegetation parameter for modeling soil erosion using linear Spectral Mixture Analysis of Landsat ETM data. ISPRS J. Photogramm. Remote Sens. 62, 309–324.
Deng, Z., Zhu, X., He, Q., Tang, L., 2019. Land use/land cover classification using time series Landsat 8 images in a heavily urbanized area. Adv. Space Res. 63, 2144–2154.
Dong, T., Liu, J., Qian, B., Zhao, T., Jing, Q., Geng, X., Wang, J., Huffman, T., Shang, J., 2016. Estimating winter wheat biomass by assimilating leaf area index derived from fusion of Landsat-8 and MODIS data. Int. J. Appl. Earth Obs. Geoinf. 49, 63–74.
Doraiswamy, P.C., Hatfield, J.L., Jackson, T.J., Akhmedov, B., Prueger, J., Stern, A., 2004. Crop condition and yield simulations using Landsat and MODIS. Remote Sens. Environ. 92, 548–559.
Dwyer, J., Schmidt, G., 2006. The MODIS reprojection tool. In: Earth Science Satellite Remote Sensing. Springer, Berlin, Heidelberg, pp. 162–177.
Feng, G., Masek, J., Schwaller, M., Hall, F., 2006. On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance. IEEE Trans. Geosci. Remote Sens. 44, 2207–2218.
Fu, D., Chen, B., Zhang, H., Wang, J., Black, T.A., Amiro, B.D., Bohrer, G., Bolstad, P., Coulter, R., Rahman, A.F., Dunn, A., McCaughey, J.H., Meyers, T., Verma, S., 2014. Estimating landscape net ecosystem exchange at high spatial–temporal resolution based on Landsat data, an improved upscaling model framework, and eddy covariance flux measurements. Remote Sens. Environ. 141, 90–104.
Gao, F., Hilker, T., Zhu, X., Anderson, M., Masek, J., Wang, P., Yang, Y., 2015. Fusing landsat and MODIS data for vegetation monitoring. IEEE Geosci. Remote Sens. Mag. 3, 47–60.
Gao, L., Wang, X., Johnson, B.A., Tian, Q., Wang, Y., Verrelst, J., Mu, X., Gu, X., 2020. Remote sensing algorithms for estimation of fractional vegetation cover using pure vegetation index values: A review. ISPRS J. Photogramm. Remote Sens. 159, 364–377.
Gu, X., Tong, X., 2015. Overview of China earth observation satellite programs [Space Agencies]. IEEE Geosci. Remote Sens. Mag. 3, 113–129.
Gutman, G., Ignatov, A., 1998. The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models. Int. J. Remote Sens. 19, 1533–1543.
He, T., Liang, S., Wang, D., Shuai, Y., Yu, Y., 2014. Fusion of satellite land surface albedo products across scales using a multiresolution tree method in the North Central United States. IEEE Trans. Geosci. Remote Sens. 52, 3428–3439.
Huang, C., Chen, Y., Zhang, S., Li, L., Shi, K., Liu, R., 2016. Surface water mapping from Suomi NPP-VIIRS imagery at 30 m resolution via blending with Landsat data. Remote Sens. 8, 631.
Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., François, C., Ustin, S.L., 2009. PROSPECT+SAIL models: A review of use for vegetation characterization. Remote Sens. Environ. 113, S56–S66.
Jia, K., Li, Y., Liang, S., Wei, X., Yao, Y., 2017. Combining estimation of green vegetation fraction in an arid region from Landsat 7 ETM+ data. Remote Sens. 9, 1121.
Jia, K., Liang, S., Gu, X., Baret, F., Wei, X., Wang, X., Yao, Y., Yang, L., Li, Y., 2016. Fractional vegetation cover estimation algorithm for Chinese GF-1 wide field view data. Remote Sens. Environ. 177, 184–191.
Jia, K., Liang, S., Liu, S., Li, Y., Xiao, Z., Yao, Y., Jiang, B., Zhao, X., Wang, X., Xu, S., Cui, J., 2015. Global land surface fractional vegetation cover estimation using general regression neural networks from MODIS surface reflectance. IEEE Trans. Geosci. Remote Sens. 53, 4787–4796.
Jia, K., Yang, L., Liang, S., Xiao, Z., Zhao, X., Yao, Y., Zhang, X., Jiang, B., Liu, D., 2019. Long-term global land surface satellite (GLASS) fractional vegetation cover product derived from MODIS and AVHRR data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12, 508–518.
Jiapaer, G., Chen, X., Bao, A., 2011. A comparison of methods for estimating fractional vegetation cover in arid regions. Agric. For. Meteorol. 151, 1698–1710.
Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. J. Basic Eng. 82, 35–45.
Kannan, A., Ostendorf, M., Karl, W.C., Castanon, D.A., Fish, R.K., 2000. ML parameter estimation of a multiscale stochastic process using the EM algorithm. IEEE Trans. Signal Process. 48, 1836–1840.
Liang, S., Li, X., Wang, J., 2012. Advanced Remote Sensing: Terrestrial Information Extraction and Applications. Academic Press, Cambridge, MA, USA.
Lindberg, E., Olofsson, K., Holmgren, J., Olsson, H., 2012. Estimation of 3D vegetation structure from waveform and discrete return airborne laser scanning data. Remote Sens. Environ. 118, 151–161.
Liu, D., Jia, K., Wei, X., Xia, M., Zhang, X., Yao, Y., Zhang, X., Wang, B., 2019. Spatiotemporal comparison and validation of three global-scale fractional vegetation cover products. Remote Sens. 11, 2524.
Liu, Y., Mu, X., Wang, H., Yan, G., 2012. A novel method for extracting green fractional vegetation cover from digital images. J. Veg. Sci. 23, 406–418.
Lobell, D.B., Asner, G.P., 2003. Comparison of Earth Observing-1 ALI and Landsat ETM+ for crop identification and yield prediction in Mexico. IEEE Trans. Geosci. Remote Sens. 41, 1277–1282.
Luettgen, M.R., Willsky, A.S., 1995. Likelihood calculation for a class of multiscale stochastic models, with application to texture discrimination. IEEE Trans. Image Process. 4, 194–207.
Ma, Y., Liu, S., Song, L., Xu, Z., Liu, Y., Xu, T., Zhu, Z., 2018. Estimation of daily evapotranspiration and irrigation water efficiency at a Landsat-like scale for an arid irrigation area using multi-source remote sensing data. Remote Sens. Environ. 216, 715–734.
Marcotte, D., David, M., 1988. Trend surface analysis as a special case of IRF-k kriging. Math. Geol. 20, 821–824.
Matsui, T., Lakshmi, V., Small, E.E., 2005. The effects of satellite-derived vegetation cover variability on simulated land–atmosphere interactions in the NAMS. J. Clim. 18, 21–40.
Mu, X., Huang, S., Ren, H., Yan, G., Song, W., Ruan, G., 2015. Validating GEOV1 fractional vegetation cover derived from coarse-resolution remote sensing images over croplands. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8, 439–446.
Parada, L., Liang, X., 2004. Optimal multiscale Kalman filter for assimilation of near-surface soil moisture into land surface models. J. Geophys. Res.: Atmosph. 109, D24109.
Röder, A., Udelhoven, T., Hill, J., del Barrio, G., Tsiourlis, G., 2008. Trend analysis of Landsat-TM and -ETM+ imagery to monitor grazing impact in a rangeland ecosystem in Northern Greece. Remote Sens. Environ. 112, 2863–2875.
Roujean, J.-L., Lacaze, R., 2002. Global mapping of vegetation parameters from POLDER multiangular measurements for studies of surface-atmosphere interactions: A pragmatic method and its validation. J. Geophys. Res. Atmosph. 107 ACL 6-1-ACL 6–14.
Shaw, E.M., Lynn, P.P., 1972. Areal rainfall evaluation using two surface filting techniques. Hydrol. Sci. Bull. 17, 419–433.
Shi, L., Liang, S., Cheng, J., Zhang, Q., 2016. Integrating ASTER and GLASS broadband emissivity products using a multi-resolution Kalman filter. Int. J. Digital Earth 9, 1098–1116.
Tang, H., Yu, K., Hagolle, O., Jiang, K., Geng, X., Zhao, Y., 2013. A cloud detection method based on a time series of MODIS surface reflectance images. Int. J. Digital Earth 6, 157–171.
Tao, G., Jia, K., Zhao, X., Wei, X., Xie, X., Zhang, X., Wang, B., Yao, Y., Zhang, X., 2019. Generating high spatio-temporal resolution fractional vegetation cover by fusing GF-1 WFV and MODIS data. Remote Sens. 11, 2324.
Tewes, A., Thonfeld, F., Schmidt, M., Oomen, J.R., Zhu, X., Dubovyk, O., Menz, G., Schellberg, J., 2015. Using RapidEye and MODIS data fusion to monitor vegetation

dynamics in semi-arid rangelands in South Africa. Remote Sens. 7, 6510–6534.
Van de Vyver, H., Roulin, E., 2009. Scale-recursive estimation for merging precipitation data from radar and microwave cross-track scanners. J. Geophys. Res. Atmosph. 114, D8.
Vermote, E.F., Tanre, D., Deuze, J.L., Herman, M., Morcette, J., 1997. Second simulation of the satellite signal in the solar spectrum, 6S: an overview. IEEE Trans. Geosci. Remote Sens. 35, 675–686.
Verrelst, J., Romijn, E., Kooistra, L., 2012. Mapping vegetation density in a heterogeneous river floodplain ecosystem using pointable CHRIS/PROBA data. Remote Sens. 4, 2866–2889.
Wang, B., Jia, K., Liang, S., Xie, X., Wei, X., Zhao, X., Yao, Y., Zhang, X., 2018. Assessment of sentinel-2 MSI spectral band reflectances for estimating fractional vegetation cover. Remote Sens. 10, 1927.
Wang, C., Fan, Q., Li, Q., SooHoo, W.M., Lu, L., 2017a. Energy crop mapping with enhanced TM/MODIS time series in the BCAP agricultural lands. ISPRS J. Photogramm. Remote Sens. 124, 133–143.
Wang, J., Christakos, G., Hu, M., 2009. Modeling spatial means of surfaces with stratified nonhomogeneity. IEEE Trans. Geosci. Remote Sens. 47, 4167–4174.
Wang, X., Jia, K., Liang, S., Li, Q., Wei, X., Yao, Y., Zhang, X., Tu, Y., 2017b. Estimating fractional vegetation cover from Landsat-7 ETM+ reflectance data based on a coupled radiative transfer and crop growth model. IEEE Trans. Geosci. Remote Sens. 55, 5539–5546.
Wang, X., Jia, K., Liang, S., Zhang, Y., 2016. Fractional vegetation cover estimation method through dynamic Bayesian network combining radiative transfer model and crop growth model. IEEE Trans. Geosci. Remote Sens. 54, 7442–7450.
Weiss, D.J., Atkinson, P.M., Bhatt, S., Mappin, B., Hay, S.I., Gething, P.W., 2014. An effective approach for gap-filling continental scale remotely sensed time-series. ISPRS J. Photogramm. Remote Sens. 98, 106–118.
Xu, J., Yao, Y., Liang, S., Liu, S., Fisher, J.B., Jia, K., Zhang, X., Lin, Y., Zhang, L., Chen, X., 2019. Merging the MODIS and Landsat terrestrial latent heat flux products using the multiresolution tree method. IEEE Trans. Geosci. Remote Sens. 57, 2811–2823.
Yan, Y., Liu, X., Ou, J., Li, X., Wen, Y., 2018. Assimilating multi-source remotely sensed data into a light use efficiency model for net primary productivity estimation. Int. J. Appl. Earth Obs. Geoinf. 72, 11–25.
Yang, L., Jia, K., Liang, S., Liu, J., Wang, X., 2016. Comparison of four machine learning methods for generating the GLASS fractional vegetation cover product from MODIS data. Remote Sens. 8, 682.
Yang, L., Jia, K., Liang, S., Liu, M., Wei, X., Yao, Y., Zhang, X., Liu, D., 2018. Spatio-temporal analysis and uncertainty of fractional vegetation cover change over Northern China during 2001–2012 based on multiple vegetation data sets. Remote Sens. 10, 549.
Yu, T., Sun, R., Xiao, Z., Zhang, Q., Liu, G., Cui, T., Wang, J., 2018. Estimation of global vegetation productivity from global LAnd surface satellite data. Remote Sens. 10, 327.
Zhang, X., Liao, C., Li, J., Sun, Q., 2013. Fractional vegetation cover estimation in arid and semi-arid environments using HJ-1 satellite hyperspectral data. Int. J. Appl. Earth Obs. Geoinf. 21, 506–512.
Zhu, X., Chen, J., Gao, F., Chen, X., Masek, J.G., 2010. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. Remote Sens. Environ. 114, 2610–2623.
Zhu, X., Gao, F., Liu, D., Chen, J., 2012a. A modified neighborhood similar pixel interpolator approach for removing thick clouds in Landsat images. IEEE Geosci. Remote Sens. Lett. 9, 521–525.
Zhu, X., Liu, D., Chen, J., 2012b. A new geostatistical approach for filling gaps in Landsat ETM+ SLC-off images. Remote Sens. Environ. 124, 49–60.
Zhu, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in Landsat imagery. Remote Sens. Environ. 118, 83–94.
Zou, J., Lan, J., Shao, Y., 2018. A hierarchical sparsity unmixing method to address endmember variability in hyperspectral image. Remote Sens. 10, 738.